GLOBAL WATCH MISSION REPORT

# Grids, distributed computing and applications – a mission to CERN, Switzerland

JUNE 2004

dti

The DTI drives our ambition of 'prosperity for all' by working to create the best environment for business success in the UK. We help people and companies become more productive by promoting enterprise, innovation and creativity.

We champion UK business at home and abroad. We invest heavily in world-class science and technology. We protect the rights of working people and consumers. And we stand up for fair and open markets in the UK, Europe and the world.

**Global Watch Missions**

The UK government Department of Trade and Industry (DTI) Global Watch Service provides funds to assist small groups of technical experts from UK companies and academia to visit other countries for short, fact finding missions.

Global Watch Missions serve a number of related purposes, such as establishing contacts with overseas organisations for purposes of collaboration; benchmarking the current status of UK industry against developments overseas; identifying key developments in a particular field, new areas of progress or potentially disruptive technologies; or studying how a specific industry has organised itself for efficient operation or how governments, planners or decision makers have supported or promoted a particular area of industry or technology within their own country.

**Disclaimer**

This report represents the findings of a mission organised by Qi3 with the support of DTI. Views expressed represent those of individual members of the mission team and should not be taken as representing the views of any other member of the team, their employers, Qi3 or the DTI.

Although every effort has been made to ensure the accuracy and objective viewpoint of this report, and information is provided in good faith, no liability can be accepted for its accuracy or for any use to which it might be put. Comments attributed to organisations visited during this mission were those expressed by personnel interviewed and should not be taken as those of the organisation as a whole.

Whilst every effort has been made to ensure that the information provided in this report is accurate and up to date, DTI accepts no responsibility whatsoever in relation to this information. DTI shall not be liable for any loss of profits or contracts or any direct, indirect, special or consequential loss or damages whether in contract, tort or otherwise, arising out of or in connection with your use of this information. This disclaimer shall apply to the maximum extent permissible by law.

# Grids, distributed computing and applications

## – a mission to CERN, Switzerland

REPORT OF A DTI GLOBAL WATCH MISSION
JUNE 2004

Report prepared by
Nathan Hill
Peter Rice
Ted Ridgway Watt
with contributions from other mission delegates

# CONTENTS

# 1 EXECUTIVE SUMMARY

CERN, the European particle physics laboratory in Geneva, is one of the world's most prestigious research centres, funded by the UK and 19 other European countries. CERN is currently leading the way in computing and network technologies related to the construction by 2007 of the 'Large Hadron Collider' (LHC) which will be the world's largest scientific instrument. This DTI Global Watch Mission visited CERN in June 2004 to explore the potential for UK industry to benefit from the UK's investment in CERN's information technology (IT).

CERN has a large budget for the construction of the LHC accelerator. The mission also considers the IT related to the four large-scale international 'experiments' which are being constructed along the path of the LHC. These experiments will generate unprecedented volumes of data with the aim of a fundamental new understanding of physics.


*Figure 1.1* *CERN's LHC is one of the largest civil engineering projects in the world*

The data volumes are so large that CERN is leading the way in developing and adopting 'Grid computing', in which very large numbers of computers and data storage are coupled together on a global scale to distribute and analyse the results. Even at this scale, over 99% of the data created by the experiments must be immediately processed and discarded by advanced 'filtering' and 'trigger' systems which can identify known and expected events at rates of several million per second.

The mission concentrated on three distinct areas of IT:

- Distributed computing
- Scientific computing
- Administrative computing

### Distributed computing

CERN's Grid computing technology is already in use. The core computing centre in Geneva must store and process the data from the physics experiments as fast as it is generated. Massive tape and disk storage arrays are needed for this, together with dedicated quality assurance to ensure reliability and data authenticity.

Data are transferred over very high speed networks to national computing centres, including the Rutherford Appleton Laboratory (RAL) in the UK. The mission team also visited RAL for a tour of their national computing centre. The national centres provide processing and data access for further national and institutional centres where teams of physicists can use and create scientific software for research.

**Scientific computing**

The mission concentrated on one experiment, ALICE, as an example of the data processing required at CERN. The other experiments have similar needs. At the level of the front-end electronics, ALICE generates data at a rate of 100 gigabytes per second (GB/s). A comparison with another scientific field may help to appreciate the scale: the human genome project generated about 100 GB of data over 5 years.

CERN's data volume is so great that each experiment has a built-in trigger system which ignores 'expected' events, reducing the output to some 0.5 GB/s. This data must then be both saved on permanent tape storage at 1.25 GB/s and transferred over international networks as fast as it is generated.

Grid computing is the only technology available to handle such high rates of data processing and transmission globally. In addition to its core physics needs, CERN is playing a leading role in a number of Grid computing projects in Europe.

**Administrative computing**

As a major international research organisation, with 8,000 people on site and many more collaborators and suppliers worldwide, CERN has a sophisticated internal computing administration, including novel approaches to its own administrative computing.

The mission examined the needs of openness and security for CERN and its research community. CERN is able to process more than 1 million e-mails per day, with over 14,000 computer accounts, filtering unwanted messages with a sophisticated system tailored to the needs of individual users. CERN also provides and manages videoconferencing facilities from 1,500 conferences per year, from small meetings to major scientific conferences. CERN's web servers handle over 1 million transactions per day. Scientific users manage their own websites, with facilities provided by CERN's IT department to create, monitor and manage each site.

Particular interest was generated by the technology CERN uses for electronic document management. All CERN's administrative procedures use a customised system which has already become an example of technology transfer (TT) to industry.

**Opportunities for UK industry**

Although the business of CERN is high-energy physics (HEP), its scale, and scope of its collaboration have driven technology developments which in some cases are world leading. For data-centric industrial sectors – for example telecommunications, bioinformatics and finance – the nascent Grid technology is a precursor to the ubiquitous computing resources which will be expected and required by these sectors. Furthermore, the enabling e-business technologies have been developed to meet the demands of a dispersed organisation and have applicability to similarly complex industries or cooperative networks.

The mission identified three ways for UK industry to benefit from partnering with CERN:

- Technology transfer (TT): Although CERN's software projects are generally released under open source (OS) licensing, as CERN holds the full intellectual property (IP), the opportunity exists for industry to obtain exclusive rights to develop and market various products. Specific opportunities for TT include:
  - *EDH*
  - *CRISTAL*
  - *CDSWare*
  - *ROOT*
  - *POOL*
  - *CASTOR*
  - *GEANT4*

- Collaborative partnerships: CERN collaborates successfully with a number of large companies through its 'openlab' programme, and has many other small collaborations involving staff at CERN with visiting fellowships to work and train on areas of mutual interest.
  - *There are significant opportunities for collaborative partnerships through the openlab established by the IT department. In particular, the mission found merit in the establishment of a project for network and Grid security. A workshop on this topic is to be held in London in October 2004.*
  - *Companies may establish partnerships by supporting CERN Fellowships.*

- Procurement: CERN has very large purchasing needs. The organisation's procurement regulations allow UK industry to competitively bid for the supply of products and services to the core facilities, and also to the large-scale physics experiments.

## Conclusions

The mission received tremendous cooperation from our hosts at CERN, and delegates were deeply impressed by the expertise and quality of the IT department. The mission team were able to identify a number of areas where UK industry could improve its exploitation of CERN's IP, and take advantage of closer collaboration with CERN staff and scientists.

***Figure 1.2*** *Mission participants in one of the LHC experiment chambers*

# 2 INTRODUCTION

## 2.1 Background

The mission to CERN on distributed IT applications was sponsored by the UK Department of Trade and Industry's (DTI) Global Watch Service.

CERN is one of the world's most prestigious research centres. It has always been at the forefront of computing and networking technology. The World Wide Web was invented at CERN in 1989, and the laboratory has continued to invest much time and effort in developing the technologies that will allow it to run the largest scientific instrument on the planet (the LHC), gather and analyse the data from its experiments, allow its worldwide community of scientists to keep in touch, to share their data, and to use their distributed computing resources with maximum efficiency.

The UK government has committed considerable resources to e-science over the past three years, mainly through the Research Councils. The UK Particle Physics and Astronomy Research Council (PPARC) has led the way in Grid deployment through its GridPP and AstroGrid projects. GridPP has been strongly integrated into the delivery of a robust Grid capable of handling the data flows and processing requirements of the LHC, the new particle accelerator experiment that is currently under construction at CERN and due to commence experimental operation in 2007. Grids are now in operation at CERN and its associated research institutes, and this advance into the era of real Grid deployment has led to increased awareness and interest of Grids in industry. The time was thus seen as opportune to hold such a mission.

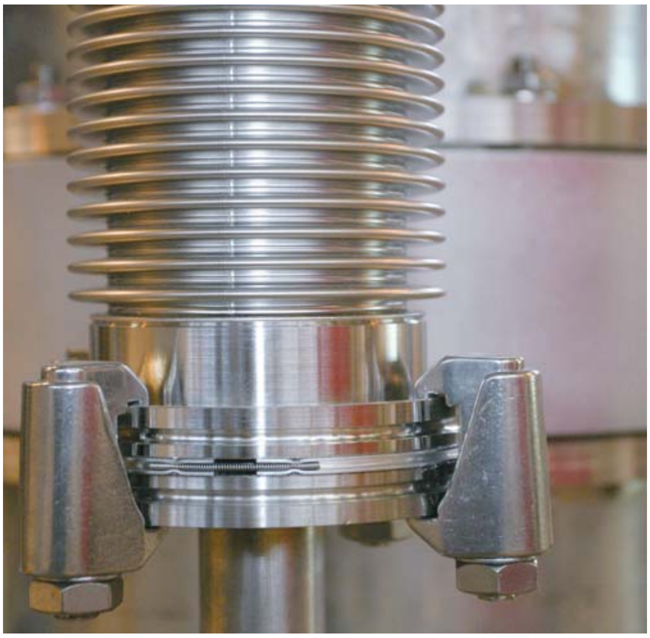

*Figure 2.1* *The challenge now facing CERN is as much about IT as it is about apparatus*

The mission visited CERN during 28-30 June 2004 with the scope including Grid applications and enterprise software. It provided three days of detailed interaction between the mission participants and CERN staff across three departments: Information Technology (IT), Technology Transfer (TT) and Experimental Physics in the ATLAS and ALICE experiments.

Qi3 is an organisation that specialises in the management of the intersection of markets and technologies, primarily in physical sciences, information and communication technologies (ICT) and healthcare. The company has built a reputation for effective 'technology translation' between public and private sectors and spin-out creation/licensing of a variety of technologies. Qi3 proposed this mission as it already has experience of CERN technologies in its role as UK Technology Transfer Coordinator for CERN. This gives the team strong existing links with the IT and TT departments at CERN and the UK IT and technology translation networks.

## 2.2 Report

The purpose of this report is to provide information gathered during the course of the mission. For further information and support please refer to the contact points listed in Appendix F.

This report contains information and opinions gained from the mission. All photographs are copyright Ted Ridgway Watt and all other images are copyright of CERN unless otherwise stated.

The report structure includes an overview and a section on the role of physics data. Readers interested in generic applications of the IT at CERN should proceed directly to chapter 5.

## 2.3 Mission aims

To bring CERN's distributed systems technologies to the attention of a sufficiently high level in industry in order to support the transfer of the technologies. This will provide acknowledged additionality to existing activities, by enhancing and extending the PPARC awareness and dissemination programme funded by the Office of Science and Technology (OST), for example by leveraging dissemination to a broader audience in the ICT sectors.

This will serve to further develop a TT environment where UK companies can benefit from a facility which is unique to Europe. As a major investor in CERN, the UK stands to benefit in the following sectors:

- Biomedical/biopharmaceutics: data logging, storage, mining and processing
- Telecoms: real time call record data storage and retrieval for security, mobile access to large-scale distributed computing facilities
- Finance: large scale distributed e-business systems
- Computational simulation: 'grand challenge problems' in structural analysis and electromagnetics

In complementing the OST-funded activities, this mission will augment the CERN 50th Anniversary programme, which has already gained high-level governmental support.

## 2.4 Key questions

In order to address the mission objectives, the following key questions were identified:

1 CERN in general
- Who owns the IP for CERN IT developments?
- What are the regulations and policies for IP at CERN?
- How is TT managed at CERN?
- How can UK companies access technology opportunities at CERN?

2 Software
- Is software developed for high energy physics useful in broader industrial applications?
- How is software maintenance managed?
- How is CERN software licensed?
- What are CERN's software innovations?

3 Standards
- How does CERN participate in and support IT standards?

4 Hardware
- How is specialised hardware procured?
- How is general purpose hardware procured?

5 Networking
- What are CERN's special networking requirements?
- What are CERN's network innovations?

6 Science
- How does CERN IT support high energy physics research?
- How can CERN's IT be applied to other scientific fields?

7 Administrative computing
- How does CERN IT support the CERN administration?

# 3 MISSION OVERVIEW

## 3.1 Participants

The mission consisted of a group of companies and an academic partner who have considerable background in Grids, distributed IT, and the ICT industry in general. The group was led by Nathan Hill of technology commercialisation specialists Qi3, together with Peter Rice of the European Bioinformatics Institute as mission leader and Ted Ridgway Watt as the DTI International Technology Promoter (ITP). Industrial participants were Derek Greer (VEGA Group), Andrew Hide (LogicaCMG), Iain le Duc (CODASciSys), Harris Makatsoris (Orion Logic), Maziar Nekovee (BT Exact), David Palmer (QinetiQ) and Paul Perkins (Atos Origin). This group gave the mission a broad spectrum of interests in the application of distributed computing in industry, bioinformatics, telecommunications, space, defence and healthcare.

For further details please see Appendix A.

*Figure 3.1 Mission participants with Roberto Amendolia, Maximilian Metzger and Jean-Marie Le Goff*

## 3.3 Overview of CERN and its IT department

Established 50 years ago, CERN is the world's largest particle physics laboratory. The mission of the organisation is to provide physicists with accelerators and detectors to study particles at high energies, and to coordinate particle physics in Europe.

## 3.2 Itinerary

| | |
|---|---|
| Monday 28 June | Introduction and tour of CERN, including the ATLAS experiment and the LHC accelerator ring<br>Technology transfer at CERN; policy and collaborative projects with industry; intellectual property management<br>Introduction to the IT department |
| Tuesday 29 June | LHC Computing Grid (LCG) and the EGEE Framework 6 project<br>The CERN openlab for DataGrid applications<br>Grid technologies: Grid middleware, Grid deployment, network and Grid security, Grid data storage and management<br>Innovative general and administrative computing, e-business and web services |
| Wednesday 30 June | Data management from physics experiments to the Grid, including hardware and software for data reduction<br>Challenges for experimental data acquisition, exemplified by the ALICE detector<br>High performance networking |

Please see Appendix C for the full programme and CERN speaker list.

CERN has 20 member states and a number of observer states and organisations. The member states provide the core funding according to national economic performance indicators, and their delegates form the CERN Council which is the highest authority and has the ultimate responsibility for all important decisions. It controls CERN's activities in scientific, technical and administrative matters, approving the programmes of activities, adopting the budgets and reviewing expenditure. The Council is helped by the Finance Committee, composed of representatives from each of the member states, and by the Scientific Policy Committee of scientific advisors.

The UK was a founding member in 1954 and had over 400 scientific users in 2002. The UK financial contribution to CERN is through PPARC.

The Director General (or CEO), appointed by the Council every 5 years, manages CERN and is authorised to act in its name. He is assisted by a Directorate composed of the Chief Financial Officer and the Chief Scientific Officer.

CERN has 2,400 core staff, plus 7,000 visitors – scientific users and staff working on the various scientific experiment projects, in addition to service contractors. Typically there are 7,000 people on site at any time.


*Figure 3.2* *François Fluckiger – one of the key coordinators of the mission team's visit*

The major part of the CERN core budget of about 1 billion Swiss francs (£450 million) is spent on new machines such as the current LHC project, a particle accelerator that will probe deeper into matter then ever before. Due to switch on in 2007, it will ultimately collide beams of protons at an energy of 14 TeV. Beams of lead nuclei will also be accelerated, smashing together with a collision energy of 1,150 TeV.


*Figure 3.3* *Iain le Duc takes his life in his hands*

After a beam of energetic particles is produced in the LHC accelerator, it can be smashed into a fixed target or made to collide with a beam coming from the opposite direction. In both cases, a large number of new particles will be produced. The physicists who use CERN build detectors to count, trace and characterise each of these particles and to study the many possible interaction processes.

These detectors are known at CERN as 'experiments'. Each experiment is a separate international scientific collaboration, with its own budget and funding. The four experiments currently being constructed with the LHC are huge projects, each with staff levels comparable to CERN itself.

The ATLAS experiment will study LHC proton-proton interactions. The project has 2,000 collaborators in 150 institutes around the world. The primary purpose of the ATLAS detector will be studies of the origin of mass at the electroweak scale; it has been designed for a large range of possible masses for the as yet undiscovered 'Higgs boson'. The detector will also be used for studies of top quark decays and supersymmetry searches.

The Compact Muon Solenoid (CMS) experiment has 2,300 staff from over 150 institutes in 36 countries. This detector is a solenoid magnet with a strength of about 100,000 times the earth's magnetic field, the largest magnet of its type ever constructed. CMS will also aim to detect the 'Higgs boson' over its full theoretical mass range. CMS can also study colliding beams of lead or calcium nuclei to study the state of matter less than 1 second after the 'Big Bang'.

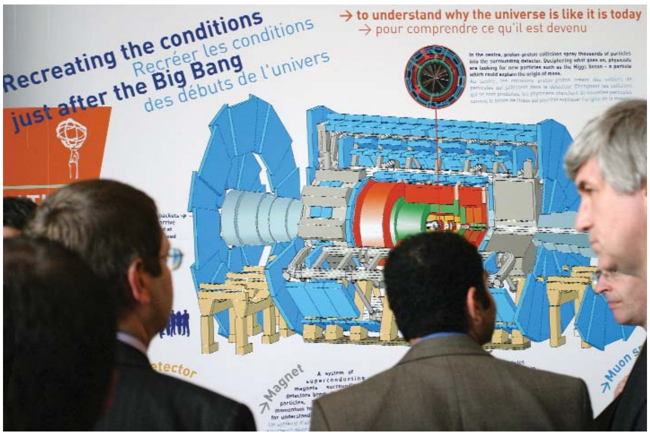


*Figure 3.4* *Mission participants learn about the LHC experiments*

The ALICE experiment has 1,200 collaborators from 85 institutes in 27 countries. ALICE will use beams of heavy nuclei to study matter in a state called a 'quark gluon plasma' (QGP) which is extremely dense and may exist in the hearts of neutron stars.

The LHCb experiment has 565 collaborators from 47 institutes in 15 countries. LHCb is less complex than ATLAS or CMS, and detects low-angle particles from proton-proton collisions.

In addition to the scientific computing needs, CERN's IT department is also responsible for the infrastructure and general services of CERN's administration. These responsibilities range from management of CERN's internal computing and networking to providing support for administrative computing for a major international organisation.

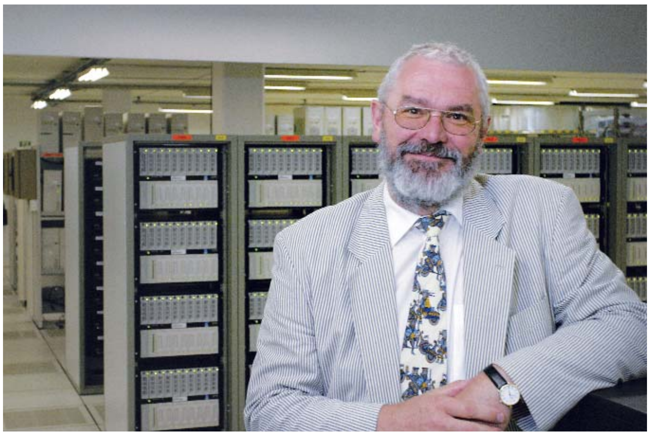

*Figure 3.5* *Wolfgang von Rüden heads up the IT department at CERN*

# 4 THE PLACE OF IT IN THE CHAIN OF PHYSICS DISCOVERY

## 4.1 Introduction

The production of data is essential to the effective conduct of experiments – indeed, IT systems for the generation, management and analysis of data are at the heart of physics. This makes CERN one of the world's most important IT development centres. This chapter describes where the data are generated and how these data are used.

The LHC has four experiments around its circumference (ATLAS, CMS, ALICE and LHCb). Whilst these experiments are different in terms of their objectives, there are significant similarities in the technical challenges, and the way in which these challenges have been addressed, so that just the ATLAS and ALICE experiments will be considered here.

The experiments are regions in which the two contra-rotating beams of the LHC are brought together in order to cause particle collisions. Each experiment has several detector systems which generally exist in coaxial layers and are relevant to specific particles or phenomena. An example is shown in Figure 4.1.

There are three significant challenges to the acquisition of experimental data:

1 Design of detectors and detector systems which must:

a Survive the harsh operating environment over the commissioning period of the LHC and the whole lifetime of the experiment

b Provide adequate accuracy for the experimental requirements

c Not impede the correct functioning of the other detector systems in each experiment

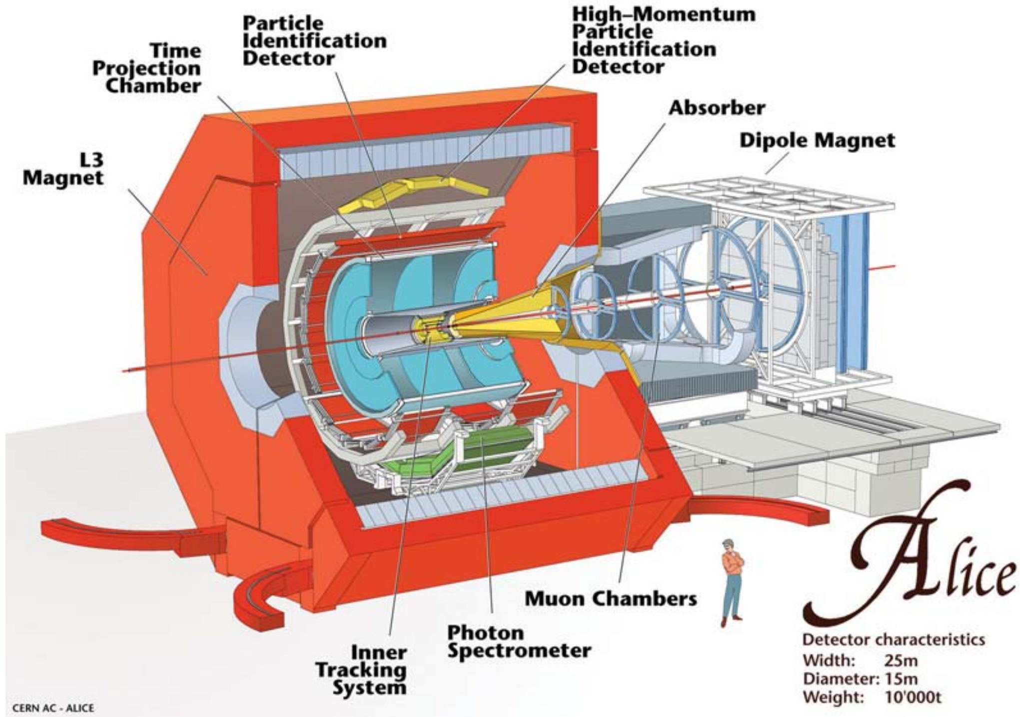


*Figure 4.1 The ALICE experiment schematic*

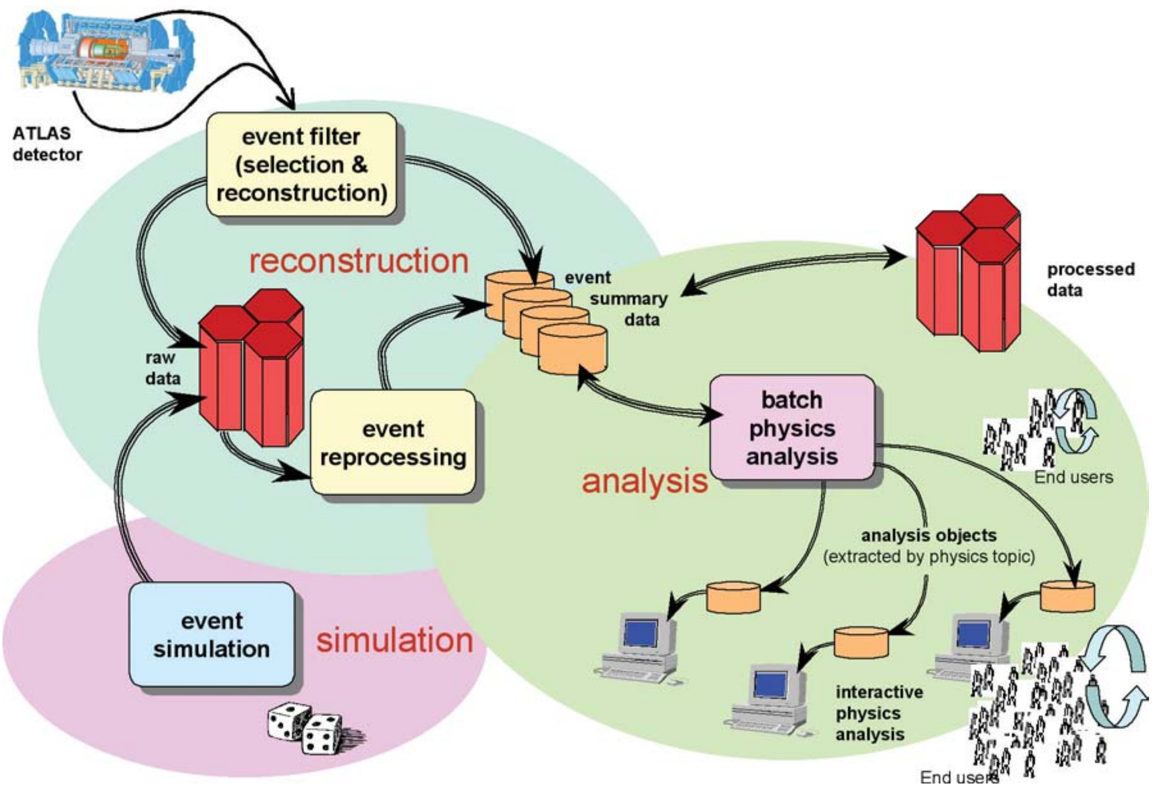


*Figure 4.2* *Overview of data flow from ATLAS LHC experiment*

2 **Filtering of event data to optimise the data flow from the sensors, noting:**
a That collisions occur at 40 MHz and result in a data readout rate of 100 GB/s
b Information whether a collision occurred only after 6 µs
c Information whether data are worth being recorded only after 100 µs
d Time available to readout data is 256 µs
e Final application environment can only be simulated (prior to commissioning of the LHC) besides the test beams activities

3 **Fusion and streaming of data to off-line processing facilities:**
a There are tens of millions of channels of information coming from the detectors in ALICE alone
b Trigger and event building hardware is located 100 m away from the detectors, leading to potential signal integrity issues on the optical media used

An overview of the global data flow from ATLAS is shown in Figure 4.2.

Essentially the data are collected from discrete sensors (front end electronics – FEE) and then a hierarchy of decisions (triggers) and data fusion produces a stream of data which is of interest. Subsequent processing further characterises the event and it is finally piped out from CERN as 'event summary data' (ESD) to a network of primary data centres ('tier-1 centres' – see Figure 5.5). The management of storage and data is covered in section 5.3.

Figure 4.3 shows the data acquisition (DAQ) process of transmitting the sensor signals (output from the FEE) and assimilating the data via data concentrators. This process draws together data from different detectors and enables the software to build the event.

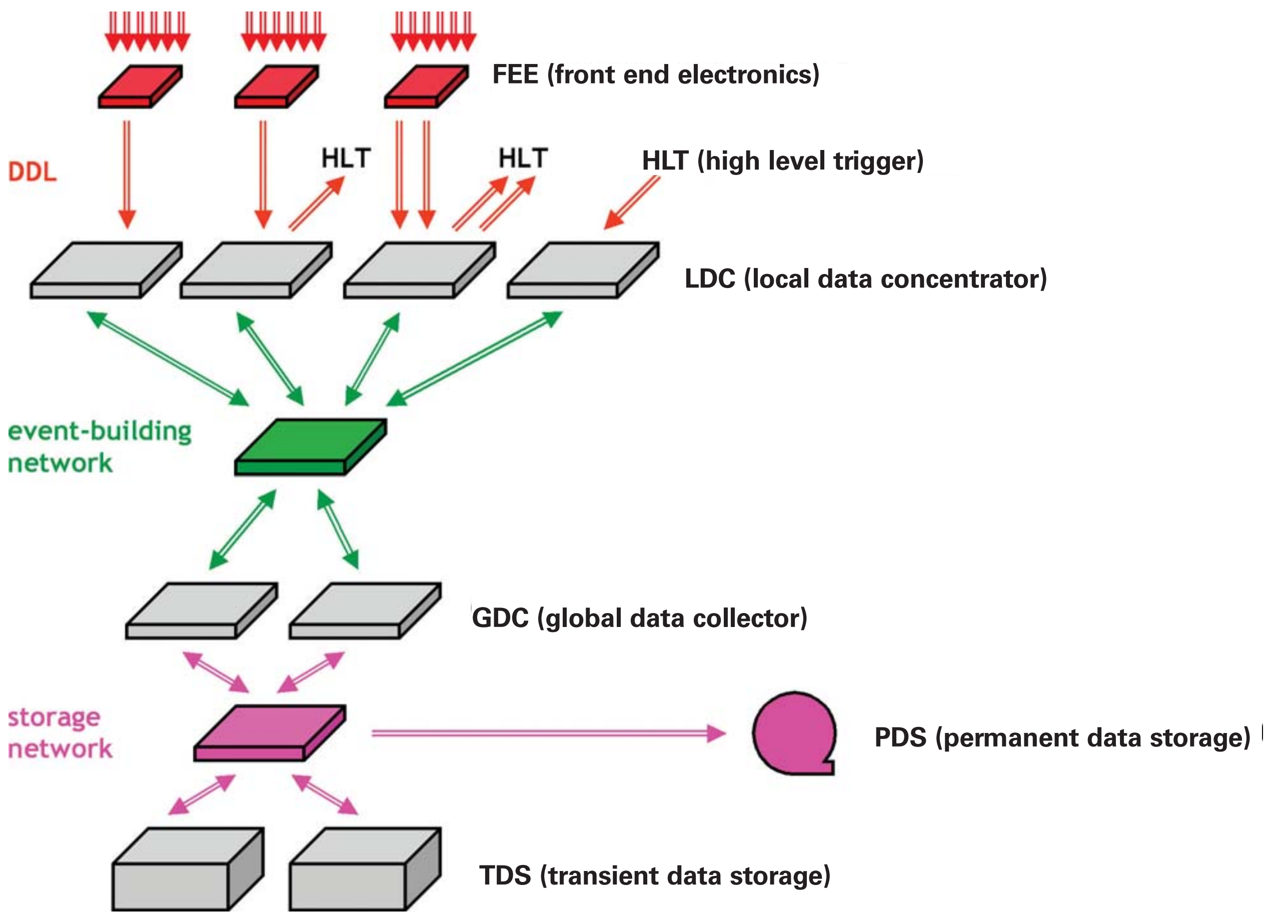


*Figure 4.3* *Data acquisition (DAQ) process flow*

## 4.2 Experimental data acquisition

### 4.2.1 Detectors

Detectors are subject to physical and performance constraints.

1 Physical constraints
a Volume; the electronics must fit within space which is determined by adjacent sensor systems.
b Transparency; the detection systems must not have a deleterious effect on subsequent sensors.
c Heat dissipation; the temperature within the experiment must be held within bounds in order that the sensors behave to specification and that the particles propagate as predicted. In the case of ALICE, the heat dissipation for the innermost detector (the pixel tracking detector) is 1 kW in a 5 $dm^3$ volume.

2 Electronic constraints
a Connections; although the experiments are physically large (ALICE dimensions shown in Figure 4.1), there are tens of millions of channels of data from the sensors. The silicon pixel tracking detector, for example, has 10 million channels of data within a cylindrical volume of 280 x 155 mm diameter, so CERN has had to explore effective ways of providing dense

*Figure 4.4* *Particle detectors for the LHC*

but robust connections which will withstand a hostile radiation environment of 250 krad. Techniques such as bump bonding have been used to achieve high connection densities. To address the radiation issue, the signal bus within the detector was made up of a five layer aluminium-copper (Al-Cu) structure, but the problem of creating vias (to connect different levels) led CERN to develop a technique where the layers were offset in a staircase and joined with wire bonds.

b Performance; the detectors are not accessible once the experiments are commissioned and must provide reliable service throughout the lifetime of the LHC. Some systems are able to accept a limited degree of failure through redundancy or interpolation.

c Calibration; the front end systems are generally capable of being recalibrated during the lifetime of the experiments. This enables the experiments to recover from limited failure or to alter (for example) triggering parameters.

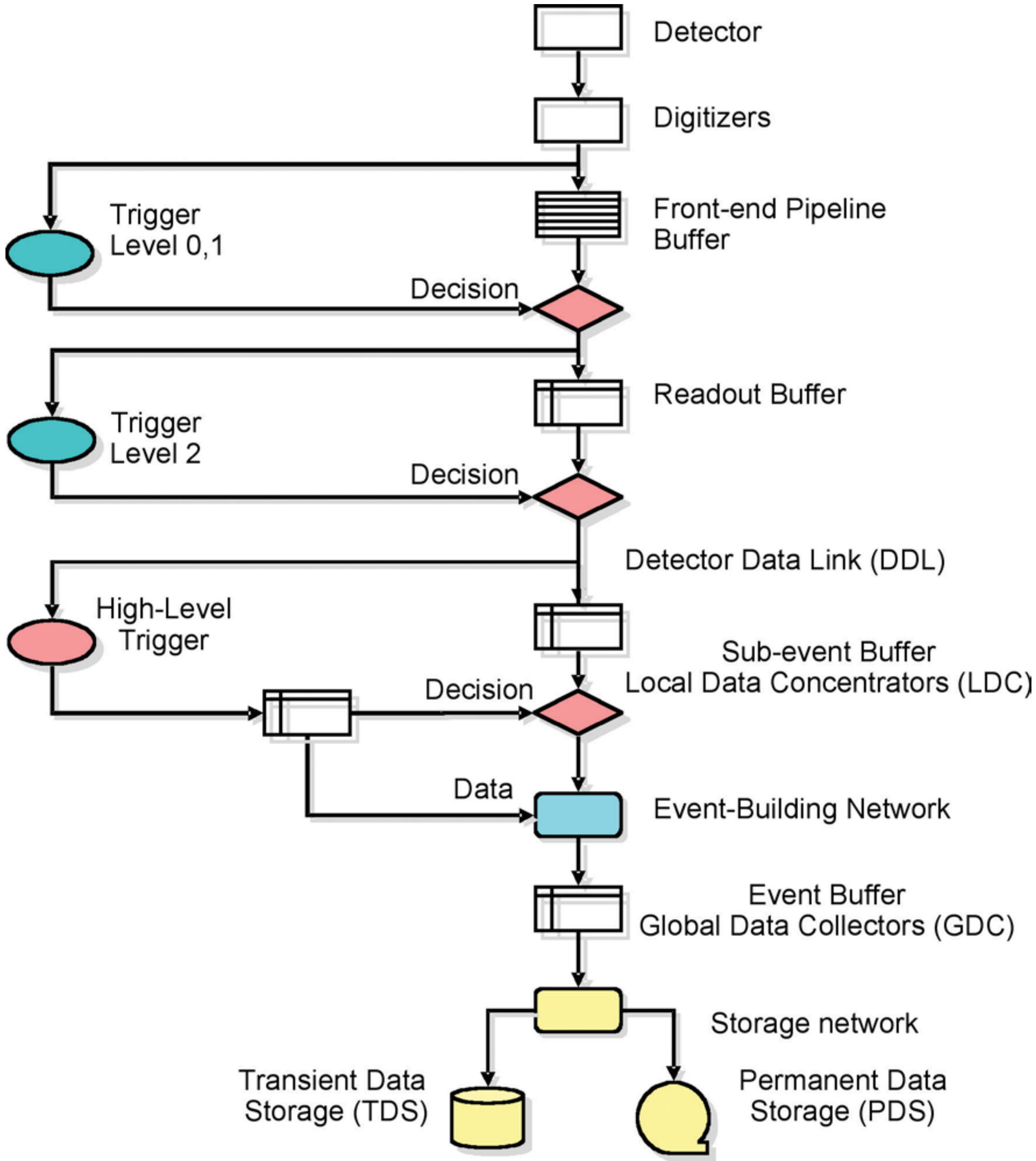


*Figure 4.5 Logical model of ALICE DAQ*

### 4.2.2 Triggering

Triggering is the process of identifying that events have been detected, ascertaining if the event was of interest and then deciding, based on other data, if the event should be piped out of the experiment.

1 Data acquisition (DAQ)
(logical model, as shown in Figure 4.5). The low-level triggers supply data to the high-level trigger which will decide whether an event is to be gated through to the data network or if it is not of interest. The purpose of the multi-layer triggering system is to reduce the datasets which propagate through to storage (see section 5.3). Some level of parameter recalibration will be possible after commissioning.

2 Hardware implementation
The hardware implementation of the triggering system in ALICE is described in section 5.1.

### 4.2.3 Event building (data fusion and streaming)

The ALICE DAQ software system comprises three elements, all developed in-house and subject to continued collaborative development:

| | |
|---|---|
| DATE | Data flow, detector readout and event building |
| AFFAIR | Performance monitoring; uses feedback from the other data management systems in order to track the performance of data streaming |
| MOOD | Data quality monitoring |

Once the data have been flagged as 'of interest', then they are output to ground level and then to the data management system – see section 5.3.

## 4.3 Conclusions

The acquisition and management of data from the LHC detectors is a significant challenge which has been met by CERN by using a combination of in-house hardware and software and external collaboration. Key points emerging from the mission are:

- Processing is performed by third party, readily available hardware provided by supply contracts to CERN from member states' industries and other large multinationals but the network architecture is defined by CERN.

- Detectors, which are largely complete for LHC experiments, have been designed by CERN in conjunction with collaborating laboratories and universities. However, there are future developments planned which may yet provide opportunities for suppliers and collaborators:
  - *DWDM at 1 TB/s*
  - *FPGA and optical components at 10 and 40 GB/s*
  - *Hardware improvements*
  - *Evaluation of DAQ performance at system and component levels*

- In general, CERN's supply contracts are for relatively small numbers of high-value items, which may favour small contractors. For these companies, the PPARC team supporting TT and supply contracts are a valuable source of information and contacts.

# 5 DISTRIBUTED COMPUTING

## 5.1 Data processing hardware

The challenge of the data processing hardware is to manage data which can stream at many gigabits per second and resolve these data into structured event summary data which are piped out for storage and analysis.

The entire data system cannot be tested in its final state until the LHC is commissioned and so a number of test beds have been developed in order to simulate experimental conditions and verify correct functioning of the hardware and software.

In general, the hardware is composed of commercially available equipment which has been networked as highly parallelised computer resources linked with gigabit channels. With reference to the overview in chapter 4, the hardware overview for ALICE is as follows:

Platforms

- Single/dual/quad CPU
- Rack-mount and VMEbus form-factor
- PCI or PCI-X slots for read-out-receive cards (RORCs) and fibre channel adaptors
- Linux operating system (Red Hat)
- DATE software including custom drivers

Networking

- Gigabit Ethernet adaptors/switches
- Fibre channel adaptors/switches

Data storage

- Disk array (RAID level 0 or 5)

Infrastructure

- KVM (keyboard, video, mouse) system
- PDUs (power distribution units)

*Figure 5.1* *A tiny part of CERN's computing resource – British companies have been successful in supplying some of these requirements*

### 5.1.1 Platforms

The computing resources are comprised of 32-bit and 64-bit processors, primarily dual-CPU.

The local data collectors (LDCs) and global data collectors (GDCs) (see Figure 4.3) are primarily dual Xeon processors, currently provided by Broadberry and Elonex, whilst the DATE DAQ management software may run on 64-bit dual Opteron machines (Newisys 2100).

*Figure 5.2* *Bank of Itanium processors*

In general, CERN has found that dual 32-bit processors currently achieve the optimum performance against cost.

32-bit processors run Linux Red Hat 7.3, whilst the 64-bit processors use UnitedLinux 1.0 or Scientific Linux CERN 3 (SLC3).

### 5.1.2 Network infrastructure

The event-building network uses Gigabit Ethernet, with an aggregate bandwidth of 2.5 GB/s. CERN has evaluated network performance by looking at transfer rate against TCP socket size, comparing alternative Ethernet controller systems (3C996 – BCM5700 and Pro/1000 MT server – 82545EM) and established peak performance at 128 kB socket size.

### 5.1.3 Data storage

In this context, data storage is from the DAQ system and not analysis and archival storage in tier-1 centres. A summary of the transient storage system follows:

Connection technologies
- DAS (direct attached storage)
  – has scaling problems
- NAS (network attached storage)
  – not conclusive
- SAN (storage attached network)
  – good results with FC (see below)

Disk arrays
- Product from DotHill
  – 12x 36.7 GB based on FC disks
  – averaged >140 MB/s write speed in benchmarking
- Product from Infortrend
  – 12x 300 GB based on IDE disks
  – averaged <120 MB/s write speed in benchmarking
- Going to test products based on SATA technology
- Dual 2 Gbps ports, rack-mount
- Configuration: RAID level, file system (ext2, xfs), etc

Fibre channel fabric
- Host bus adaptor (HBA): products from QLogic
- Switch: product from Brocade with 15 ports

### 5.1.4 Conclusions

ALICE DAQ is based on standard technology, as:

- Technology is driven by the market
- Performance increases by almost constant price
- CERN has chosen supply from different vendors, to assess performance
- CERN's experimentation and evaluation of very high-throughput data transmission protocols – beyond conventional TCP/IP – may prove valuable to UK industry facing challenges of ultra high rate data management.
- Fast evolution cycles mean that continual technology watch is essential

Reliable performance to specification is only achievable through evaluation of computing and networking elements:

- Doing functional tests and benchmarking
- Working closely with the manufacturer
- A structured, documented approach

### 5.1.5 Opportunities for UK industry

*Key potential for partnership and supply*
- Need to follow the CERN procurement regulations (see section 7.2)
- Installation is phased up to 2007 and new opportunities will arise
- Hardware integration, following the model successfully used up to the present to develop an innovative data management and dispersal system

## 5.2 Grid middleware

The major motivations for the development of Grid computing technology at CERN are the capturing, management, storage, retrieval and distribution of the large amounts of data generated by the LHC. This will generate data of the order of petabytes per year (1 petabyte (PB) = $1 \times 10^9$ megabytes).

This amount of data requires a large pool of computational resources that exist in a number of large data centres, internally at CERN and at partner universities and research institutes around the globe. Physicists will then be able to search and download data that is of interest to their research. To marshal such a large network, or Grid, of resources CERN is in the process of building distributed middleware, by combining existing, new and emerging technologies, which will provide the services to control those resources and manage data across the Grid.

### 5.2.1 LHC Computing Grid

Those challenges led to the LHC Computing Grid (LCG) project. The aim of the project is to prepare, deploy and operate the computing environment for the experiments to analyse the data from the LHC detectors. This involves building and operation of an holistic LHC computing service, which includes the development of Grid middleware and the development of common tools and frameworks for physics analysis and simulation. It also involves the collaboration of researchers, software engineers, national computing centres, and other research networks which will ultimately operate as a single virtual LHC computing centre.

CERN's involvement in the Grid collaboration is aimed at feeding in to LCG and establishing the performance required for LHC's operation.

### 5.2.2 EGEE – Enabling Grids for e-Science in Europe

Over the past ten years, there has been a number of distributed software technologies developed at CERN to handle, in an efficient and cost effective manner, the large amounts of data produced by high energy physics (HEP) experiments. CERN recently launched the project EGEE, Enabling Grids for e-Science in Europe. EGEE is a seamless Grid infrastructure for the support of scientific research. It aims at the integration of current national, regional and thematic Grid efforts and the provision to researchers in academia and industry of round-the-clock access to major computing resources, independent of geographic location. EGEE is a European-funded project, which involves 70 leading institutions federated in 28 national Grids. The EU part of the funding available to this project for the period 2004-2005 is about €32 million. The project aims at a combined capacity of 8,000 CPUs – the largest international Grid infrastructure ever assembled. Although the project is still at its initial phases since its beginning in April 2004, it ultimately aims to benefit a large number of domains in addition to HEP. Examples include healthcare, industrial applications, bioinformatics, chemistry, astronomy, nanotechnology, climate modelling and others. In the case of healthcare, already there is exchange of ideas between EGEE and the MammoGrid project, also funded by the EU.

EGEE will examine, combine, re-engineer and package existing middleware technologies (eg Globus/OGSI, Virtual Data Toolkit, EDG etc), experience of partners involved in large-scale distributed software and technologies already developed and used in LCG to ultimately provide robust and supportable components and influence related emerging standards (WS-RF). In addition it is expected that the limited Grid operation experience available at present will be expanded. Since April 2004 the focus has been the implementation of a

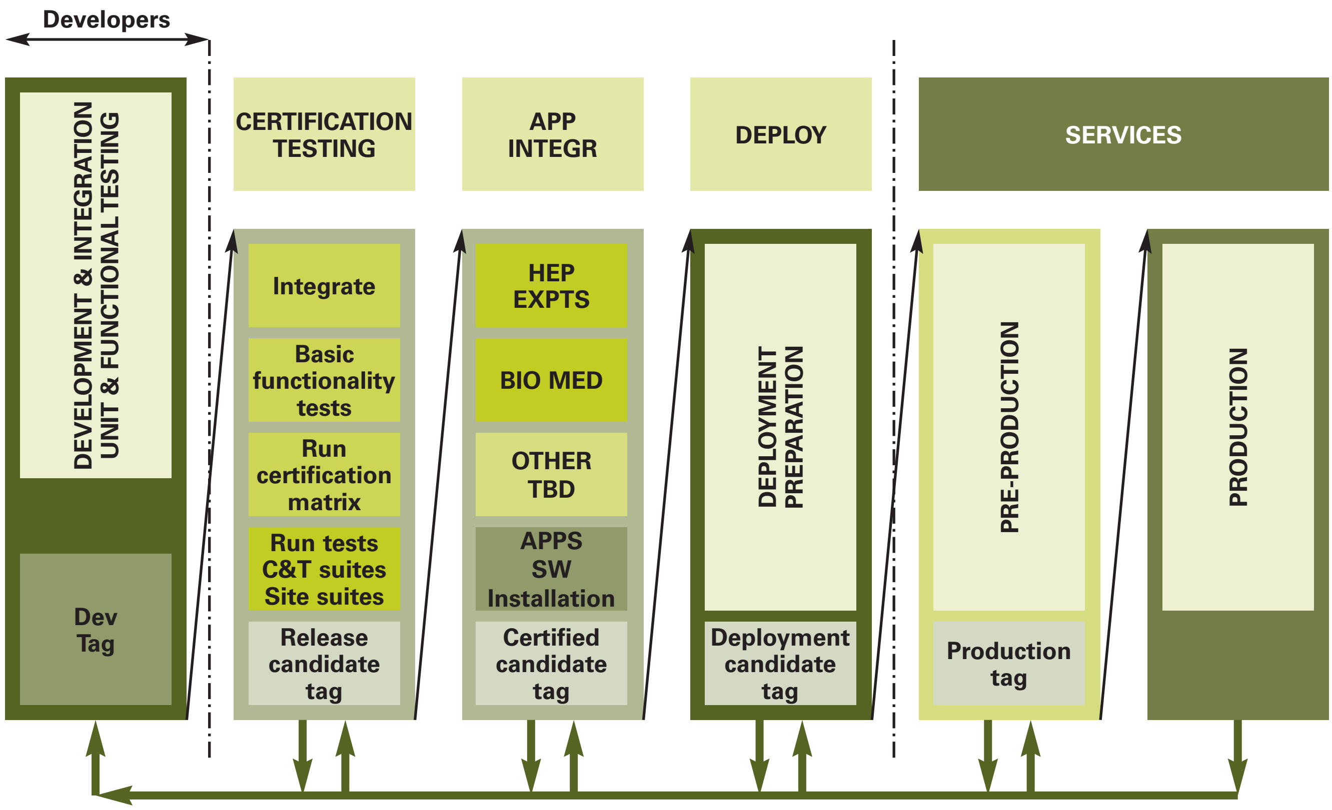


*Figure 5.3* *Grid certification, testing and release cycle*

production Grid service based on the LHC infrastructure running LCG-2 middleware. However, there is a parallel development (starting from components from existing Grid middleware such as AliEn, VDT, EDG etc) of a more generic next generation Grid facility based on new standards such as web services. This will result in the ultimate replacement of the LCG-2 on the production facility in spring 2005.

At present the existing EGEE middleware, which is based on LCG-2, is distributed under an OS licence developed by the EU DataGrid. Under this licence there are no restrictions on usage for commercial or scientific purposes other than an acknowledgment. Application software maintains its own licensing scheme. The same approach will be used for the new middleware. See section 5.5.2 for collaborative details relating to EGEE.

### 5.2.3 Deployment and security issues

The development and deployment of Grid infrastructure involves a large number of engineers and partner institutes. For this reason, CERN developed a certification, testing and release process, which is summarised in Figure 5.3.

The LCG Grid is a virtual organisation that comprises tiers and a large number of 'stakeholders'. Each tier includes a number of data centres according to their computational resources, management policies and computer platforms used. In the UK, RAL is a tier-1 data centre. Imperial College, London is a tier-2 data centre. The LCG Grid deployment board has been set up to address policy issues requiring agreement and negotiation between resource centres. Working groups that comprise technical experts who focus on specific issues are set up either for a short term or are ongoing. Such groups are responsible for security issues, operations, resources, support etc. The Grid deployment board approves recommendations of such groups and ensures implementation of approved recommendations that will lead to interoperability of the participating data centres' infrastructures, and improvements in operation and security of the Grid.

The LCG Grid is a highly connected and novel infrastructure that comprises a large number of distributed 'farms' of computational resources, massive distributed storage capacity, complex and heterogeneous and dynamic environment. The LCG Grid is thus subject to security risks. Such risks include:

- Launch of attacks on other sites
- Illegal or inappropriate distribution or sharing of data
- Disruption by exploitation of security holes
- Damage caused by viruses, worms etc

Security is thus a major concern, and it has to be addressed both by the Grid middleware and also by the application of non-conventional policies. The LCG security group is responsible for Grid security, and although some procedures for registration, authentication, account management and auditing are in place, they are mostly manual and not fully addressed by existing Grid management software. However, some issues will be addressed during the EGEE project.

### 5.2.4 Opportunities for UK industry

*Key potential for technology transfer*
EGEE offers an open industry forum, which primarily focuses on knowledge dissemination in relation to Grid technology. However, companies can take advantage of the OS licensing of existing and emerging Grid middleware technologies from CERN. The type of OS licences used by CERN does not restrict use of such technologies in a commercial product. In addition, CERN can offer consulting services that relate to the IP and know-how developed at CERN, to transfer such technologies to industry. Examples include EGEE, ROOT and Castor.

*Key potential for collaborative partnerships*
There are several channels for collaboration between CERN and industry that lead to knowledge dissemination, TT and procurement opportunities. The CERN openlab is a framework for collaboration with industry, and its objective is the evaluation, integration and validation of cutting edge technologies. CERN's policy is to procure and integrate standard commercial components where possible, and develop new technology when necessary. The CERN openlab is a test-bed of commercially available components. Industrial members can then have access to a state of the art, large scale distributed environment where they can test and benchmark their latest products. Currently there are two membership tiers, partner and contributor. There are five partners and one contributor at present, all primarily supplying hardware with the exception of Oracle who supply its relational database management system (RDBMS) and IBM supplying a compound file and storage system.

*Opportunities for UK supply*
Finally, procurement opportunities will arise in the latter half of 2005 and the beginning of 2006, when CERN will be expanding their data centre resources to ramp up to LCG's operational target. Such opportunities will primarily relate to hardware infrastructure.

## 5.3 Storage and data management

### 5.3.1 Introduction

The data storage needs at CERN will grow hugely when the LHC comes on line. CERN predicts a need to be able to process and store 12 – 14 petabytes per annum from 2007.

The data are stored in the form of one raw data file per collision, with further files of processed data for 'interesting' data – the more interesting, the more processing is needed to extract the interesting information and to detect misleading data. Each raw data file will hold about 2 GB of data. CERN currently uses Oracle to implement the


*Figure 5.4* *CERN will need to manage the storage of 12 petabytes per year*

catalogues of all the raw and processed data files. The total number of files is expected to be in the region of 40 million for the LHC.

Another aspect of CERN's needs is the requirement to not only receive and store these large volumes of data, but also to disseminate the data systematically to several external organisations, with each receiving possibly only a subset. This philosophy allows the data captured by CERN to be backed up at sites external to itself (provided that all data go to at least one external site), whilst giving the external organisations local access to data that they are interested in.

In this hierarchy (Figure 5.5), CERN is termed the tier-0, with the external organisations as the tier-1. Beyond tier-1 is tier-2, which are centres for collaborative analysis and further data dissemination.

In terms of the volume of data, CERN expects that the space needed throughout the three tiers shown in the diagram will be about 80 PB (ie a five-fold redundancy overall).

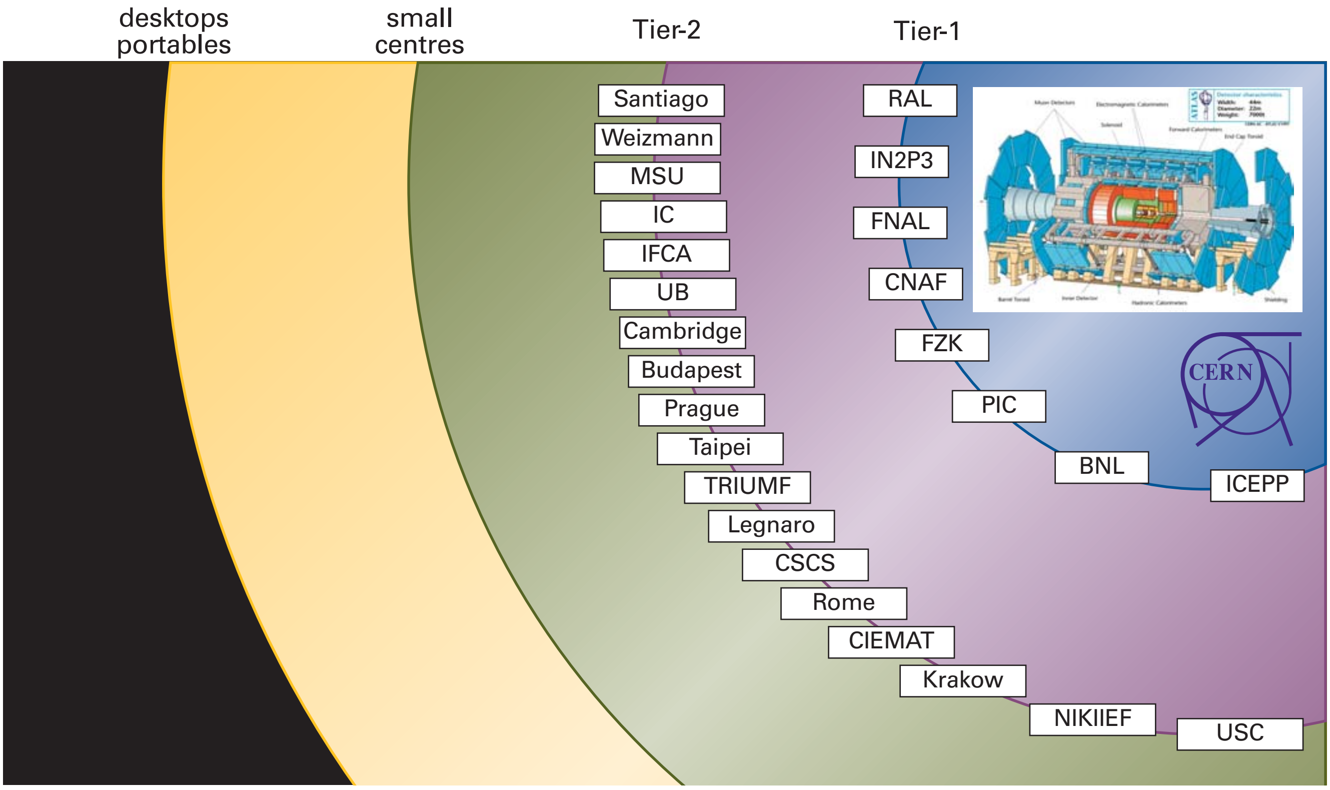


*Figure 5.5* *The tier hierarchy of data transfer*

As a percentage of the storage needs, the raw data is expected to consume about 90%, the processed data about 9% and the metadata about 1%.

CERN intends to make use of the Grid, in particular the EGEE, to both distribute data and to process them.

### 5.3.2 CERN's technology requirements

CERN use a mixture of commercial-off-the-shelf (COTS) software and bespoke software developed in-house. The main COTS product is Oracle, which is used to hold the metadata for the catalogues of all the data files held throughout the tiers 0 and 1, as well as some of 2. The catalogues are themselves hierarchical in nature, in that a file that is to be found in a tier-1 or tier-2 store is identified in the CERN catalogue only down to the site; the Grid catalogue then provides the further resolution of the exact location. All this mapping is handled transparently to the science applications (in part in the CERN software, in part in the EGEE Grid middleware). It is also used for many other applications related to detector construction, monitoring, calibration, file transfer, job production bookkeeping and so forth.

Interoperability is also a fundamental need. In many cases, it is far more efficient to transport an application to the data, than the other way around. However, that means that applications must be 'certified' for one or more platforms. Then the Grid middleware can choose the most efficient means of getting the data and application together. However, how this interoperability/certification can be achieved is unclear at present.

The tier-0 and tier-1 sites employ high availability hardware, software and procedures. However, the tier-2 sites are smaller and cannot offer the same levels of service as the tier-1. Providing some means of data transmission guarantees and processing

*Figure 5.6* *Mission leader Peter Rice listens as Wolfgang von Rüden explains CERN's storage requirements*

job re-tries is needed. CERN is looking at this in the context of EGEE, but have not yet solved the problem. In addition, tier-2 sites will need to have light-weight maintenance solutions applied to them (given their 'small' service status).

With the number of data movements and an average of a five-fold redundancy of data throughout tiers 0 to 2, optimisation of data movements across the Grid becomes essential – to minimise network loads. Again, this is a factor needing to be addressed in the EGEE middleware.

The physical stores at CERN premises are currently large arrays of PC servers, but might be based on high-speed storage access networks (SANs) in the future.

### 5.3.3 Opportunities for UK industry

*Key potential for technology transfer*

Three applications developed by CERN were identified for TT within the data storage dissemination area:

- POOL = Pool Of persistent Objects for the LHC
- CASTOR = CERN Advanced STORage management system
- ROOT = Data querying tool

POOL and CASTOR work together to satisfy the storage and cataloguing needs of CERN for the raw and processed data, POOL primarily for the online (disk) storage and CASTOR for the offline (tape and disk cache) storage.

ROOT is further downstream in the data cycle, satisfying their needs for data querying for data analysis.

## POOL

POOL is a project building upon existing object streaming and database technology at CERN, focusing on C++ object persistence for the LHC experiments. It provides the storage and retrieval of the LHC experimental data and associated metadata (catalogues) across a distributed environment within the context of the Grid, coping with databases of 100 PB or more. The data include the large raw data, with relatively small metadata, through the event summary data to the highly processed end data, with relatively large metadata. This requires a hybrid approach, with C++ object streaming technologies for the bulk raw data and transactionally-safe services for the metadata and their retrievals access.

CERN believes with good reason that a Grid-enabled hybrid such as POOL will succeed where the object databases of the 1990s failed and that technologies based on POOL will have very much wider application than that of the LHC. They are endeavouring to implement POOL in a generic and open way, using off-the-shelf solutions where possible and with a well-defined infrastructure.

A relational backend to POOL has recently been developed. It is implemented on SQLITE, MySQL and Oracle. It is expected that online detector conditions data will use this backend, as well as offline conditions data, future implementations of the file and metadata catalogues and possibly event level metadata. The database services for the initial production phase of the LHC are expected to be based on Oracle 10g Real Application Clusters. Oracle 10g includes a number of enhancements that are important to CERN and for handling scientific data in general, including support for ultra-large databases, together with automatic storage management, as well as efficient handling of IEEE floats and doubles.

## CASTOR

CASTOR is a hierarchical storage management system written by CERN that manages the physics data produced by CERN. This includes the data held by the tier-1 sites. CASTOR is already in use by some tier-1 sites, and others are evaluating it for their own needs as well as CERN's. Files may be stored on one or more tier-1 sites, as well as the tier-0 site, and may be in online disk storage or offline tape storage.

Applications access files under logical names through a standard file system programming interface, and CASTOR re-maps these to real physical locations in the disk cache. The actual physical stores (whether disk or tape based) also employ rapid access disk caches. Where a file is available in more than one store, CASTOR endeavours to optimise the data access between the application and the data that it needs (much like server and CPU cluster software optimises processing resources). Note that CASTOR is not a Grid data management system per se. It is a mass storage system local to a site (CERN in this case) and it has no relation to Grid replica management systems.

CASTOR also manages the consistency of all copies of all files held within the managed disk caches on a site, including locking the initial writing of a file. The CASTOR project is collaborating with several organisations in the development of the software, including: the Maui project, Platform Computing Inc, and Grid Storage Management GGF working group.

Since all experimental data applications use CASTOR's programming interface to access files, it is hoped that changes to CASTOR for use in other business applications may be kept largely to an equivalent layer of software. CASTOR is licensed as OS under the General Public Licence (GPL).

ROOT

ROOT is an efficient data storage and access system supporting structured datasets in very large distributed databases, with a sophisticated query system and scientific 2D and 3D visualisation tools. The graphical user interface (GUI) of ROOT is implemented on Unix and Windows.

One of the features that makes ROOT unique is that the database schemas are self-describing. As the application evolves, so do the schema employed. The data and the schema are related in the database, such that any future application accessing a range of old data will receive the data already translated (if needed) from the older schema to the one that it is using. ROOT is an OS project.

*Opportunities for collaborative partnerships*

CERN took a leading role in the European Data Grid (EDG), which was the precursor of the EGEE project. In particular, the Grid middleware used and developed by EDG had been used as the starting point of the middleware for the EGEE. The EU funded both of these projects, with the consequence that the EGEE project is open to European organisations. CERN is clearly open to cooperation with industry and academia in this area.

CERN's interest and leading role in EGEE points towards the possibility of forming partnerships with industry and academia for various aspects of the middleware interoperability - in particular, security of the data on a widespread (possibly public) Grid.

In this context, security means ensuring that the information coming back from the Grid can be authenticated as valid. The raw data going into the Grid have little intrinsic value, and if they should be snooped in passing, there is little harm; whereas the processed data coming back that provides information about specific collisions and the particles produced has to be guaranteed to be authentic. This authentication of returned information is common to many business and academic sectors; whereas the lack of interest in the privacy of the incoming data is generally foreign to most business sectors. Privacy can be guaranteed in different ways, including standard encryption techniques, and can be entirely independent of the results authentication.

Clearly, another type of partnership exists for industry to use CERN as a proving ground for data capture, storage and dissemination technologies (both hardware and software). This type of partnership is already happening in the openlab, to identify how they will solve their various data, communications and processing problems foreseen in the LHC project; data storage and dissemination being such challenges.

CERN has recognised that it has not yet fully solved the problem of storing and accessing the 140 TB of metadata foreseen in the data catalogues. Likewise, it has not solved the problem of moving 40 Gbps of data from CERN (tier-0) to the tier-1 sites situated around the world on a 24/7 basis.

### 5.4 High-performance networks

The data generated by CERN's HEP experiments is expected to increase from terabytes per year in 2004 to tens of petabytes per year by 2007, when the LHC becomes functional. At that stage, CERN needs very high aggregate requirements for computational power (roughly 100,000 of today's fastest PCs) and data storage (12–14 PB a year) but can only provide a fraction of necessary resources locally. To meet this challenge, CERN is using Grid technology to distribute the required computational power and data storage among collaborating institutions. This solution, however, requires using very high-speed network infrastructures worldwide between LHC Grid centres. Connectivity requirements for LHC are composed of a number of modes:

1. ‘Buffered real-time’ for the CERN to tier-1 raw data transfer (~10 Gbps)
2. ‘Peer services’ for tier-1 to tier-1 and tier-1 to tier-2 for the background distribution of data products (2.5–10 Gbps)
3. ‘Chaotic’ for the submission of analysis jobs to tier-1 and tier-2 centres that may imply some ‘on-demand’ data transfer (~0.1–2.5 Gbps)

CERN is thus a lead user of state-of-the-art networking technology and network infrastructures. It is a principal driver and co-developer of a next generation global optical network, and at the forefront of exploring high-throughput transport protocols for multi-Gbps data transfer rates. CERN’s activities in the area of high-performance networking were partly carried out within DataTAG and its sister project DataGrid (see section 5.5.2). CERN was the prime contractor and manager of the DataTAG project. The DataGrid project came to an end in March 2004 but work is being carried out within its successor, the EU-funded EGEE project.

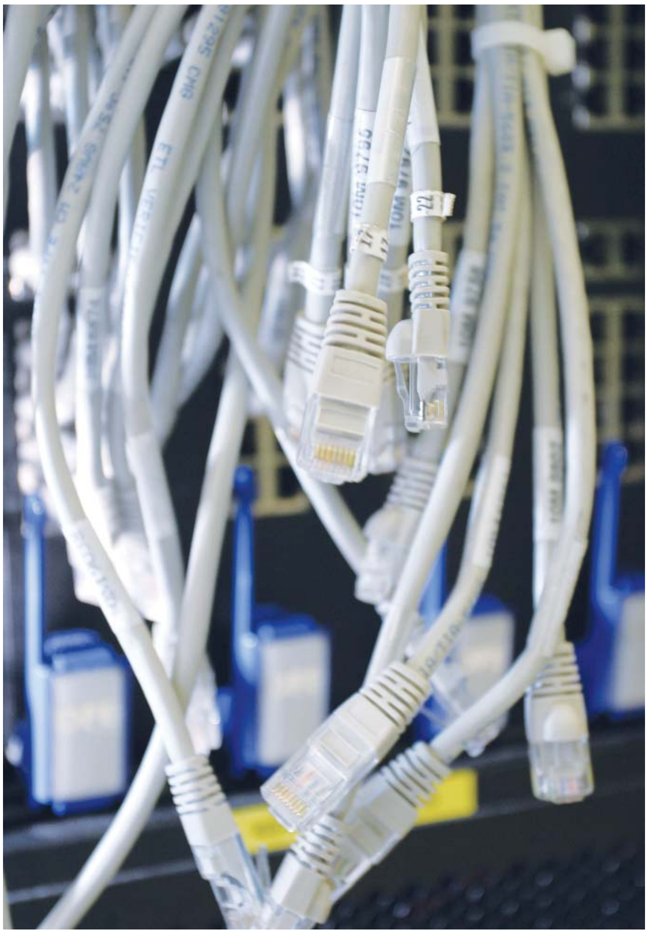

*Figure 5.7 CERN’s IT requirements are never ending*

In this section we briefly describe the next generation high-speed networking infrastructures that are currently being deployed and co-developed by CERN, and some of the achievements in deployment of high-throughput transport protocols. We then proceed to highlight potential areas of technology transfer to and collaboration with UK companies in this area, as well as potential supply opportunities.

### 5.4.1 GÉANT

CERN is a lead user of GÉANT, a multi-gigabit pan-European data communications network, reserved specifically for research and education use. The project also covers a number of other activities relating to research networking. These include network testing, development of new technologies, and support for some research projects with specific networking requirements.

The deployment of the new IPv6 internet protocols and related services on GÉANT was achieved in April 2003. In addition, a number of other service developments have been initiated, to provide IP quality of service (QoS), IP multicast, and virtual private networking (VPN). Network performance measurement tools and network security are the focus of significant development work.

### 5.4.2 TransLight and lambda Grids

CERN is actively involved in an international collaboration called TransLight which aims to provide next generation high-speed optical networking between Europe and the USA. The aim is to develop a system of optical networks for high-end e-science applications needing transport capacities of multiple gigabits per second for many hours. TransLight will make use of a fully dedicated light path, technically known as a lambda in networking terms, capable of greater than 10 Gbps bandwidth. TransLight is being designed to have a powerful simplified core, with highly distributed peer-to-peer management systems. This is in contrast to traditional networks, which are complex at core and governed by centralised management systems. TransLight's switched optical networks promise vastly increased transport capacity with predictable latency, determined largely by the speed of light, and development of new methods of bandwidth provisioning that offer control of light paths, their characteristics, and traffic behaviour to the application level.

### 5.4.3 Managed Bandwidth – Next Generation (MB-NG)

This is a collaborative project with the UK core e-Science programme, CISCO and UKERNA to look at multi protocol label switching (MPLS) for the development of managed bandwidth and QoS in high-speed networking.

### 5.4.4 Transport protocols for high performance

Today's internet relies on the transmission control protocol (TCP) for its working. TCP allows thousands of users to share bandwidth in a stable manner by deploying a congestion control mechanism. TCP's congestion control mechanism, however, results in poor utilisation of networks with large bandwidth-delay product (BDP), hence limiting throughput (note that a single TCP flow can reach very high throughputs, however sustaining such throughputs with standard TCP requires unrealistically low packet losses).

CERN network researchers are currently experimenting with several new transport protocols and techniques for tuning TCP and user datagram protocol (UDP) to efficiently use optical-fibre networks. These include Fast TCP and GridFTP, which uses multiple TCP flows. Recent advances include:

- In February 2002, an international team of physicists and computer scientists transferred 1 TB of data across 10,037 km in less than an hour from SLAC in Sunnyvale, CA, to CERN, sustaining a TCP single-stream rate of 2.38 Gbps. This throughput is equivalent to transferring a DVD movie in less than 18 seconds.
- In October 2003, during the ITU Telecom World Event in Geneva, a CERN/Caltech team transferred 1.1 TB of data across 7,000 km in less than 30 minutes, an average rate of 5.44 Gbps. This is equivalent to transferring a DVD movie in approximately 7 seconds. The record was set through the efforts of the DataTAG and FAST projects.
- In May 2004, the DataTAG project set its latest land-speed record, transferring data at nearly 1 Gbps between CERN and Chicago using the new IPv6 internet protocol.

### 5.4.5 Opportunities for UK industry

*Key potential for technology transfer*
The technologies and techniques that are currently being co-developed by CERN for next generation optical networks could have an important potential for transfer to UK providers of networks and telecommunications operators. The most immediate potential in this area is in knowledge transfer, which could be achieved through collaborative partnerships.

### *Key potential for collaborative partnerships*

Existing collaborative partnerships such as openlab have the potential to be very useful to product-oriented companies such as the major vendors of data storage devices, database management products, networking switches and routers. In the area of high-speed networking, a similar form of collaboration could be beneficial for UK network providers and operators who are looking into future opportunities in the space of network services for Grid-type applications. Possible forms of collaborations include testing of equipment in the extreme transport scenarios of LHC, knowledge transfer from CERN in the area of high-speed optical networks, and research collaboration in the area of high-throughput protocols. Finally, with current development of the Grid, network security (including user authentication and identification procedures) is becoming a pressing issue. This is a typical area, which is very well advanced in the private sector, but much less advanced in the academic and publicly funded research community, including CERN. Network security is another potential area for collaboration between CERN and UK companies.

### *Key potential for UK supply*

CERN is highly innovative in acquiring and integrating the most advanced transmission systems and routing equipment available to create a network that remains at the forefront of research networking developments. It is thus believed that, as a lead user, CERN has considerable potential for supply of high-performance networking equipment, such as switches and routers from UK based companies. Another key potential for UK suppliers, in particular in the telecommunication sector, could be provision of network monitoring and management tools, including control planes and signalling, as well as billing models and procedures.

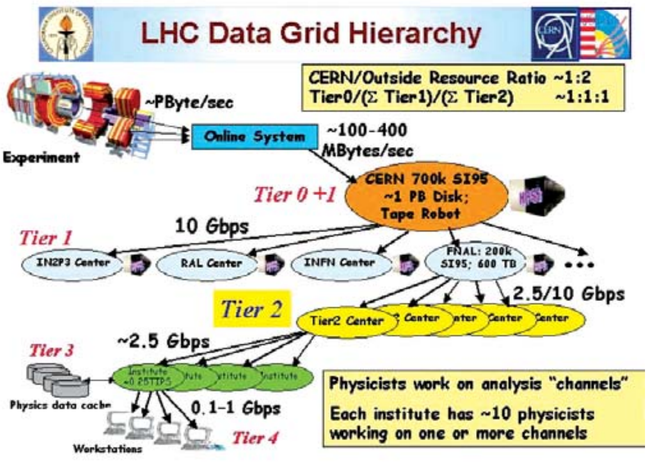


*Figure 5.8* *The LHC data grid hierarchy and its data transfer rates*

## 5.5 Major IT collaborative projects for physics data

In this section, we describe the various major IT collaborative projects that exist for physics data and focus upon the application of distributed computing technology. It is noted that the remit and structure of these various projects is well documented on the web (and, indeed, acknowledged that these resources have been used as the principal source of information in compiling this section). URLs are, therefore, included where appropriate.

Furthermore, in focusing upon (and differentiating between) PPARC and CERN supported projects, the mission team acknowledges that many projects are funded from a variety of sources and have collaborative links that extend beyond either of these two organisations. Again, further details of project partners' involvement can be found in the URLs below.

### 5.5.1 PPARC-supported e-Science projects for physics data

The PPARC e-Science programme[1] forms part of a much wider UK programme. The full UK programme includes e-Science developments in almost all of the Research Councils

1 *http://www.pparc.ac.uk/rs/fs/es/lettr/programme_components.asp*

together with a Core Programme focused on the development of robust Grid middleware, building a UK e-Science Grid, and increasing the involvement of UK industry in Grid technologies.

The major components of the PPARC e-Science programme are described below. However, it is noted that, together with the UK core e-Science programme, PPARC is funding the UK Grid Support Centre and the Grid Starter pack. In addition, PPARC is pursuing collaborative projects with EPSRC, MRC and BBSRC and allocates 10 e-Science Studentships each year to work on a variety of innovative IT developments relevant to the PPARC e-science programme. There is also co-funding provided to support UK computer scientists working in the EDG project. The National e-Science Centre (NeSC) is based in Edinburgh and there are regional e-Science centres in Belfast, Cambridge, Cardiff, London, North-East, North-West, Oxford and Southampton.

#### GridPP
Building the computing system for the LHC. Specifically, this programme provides the UK contribution for the EDG project and implements the UK components of the EDG test bed. The test bed developments are aimed at providing a production quality platform for the use of both the LHC experiments and many other HEP experiments (BaBar, CDF, D0, Minos, UK-QCD). Approximately 30% of GridPP resources are invested directly at CERN as the UK contribution to the LHC Computing Project.

#### AstroGrid
The development and application of Grid technologies for the development of 'virtual observatories'. The main aims of the programme are to develop a working data Grid in the UK, integrating key UK data sets, and to prepare the way for the extremely large data sets anticipated from future facilities. The VISTA telescope has been chosen as presenting particular challenges in the coming decade. This programme brings together a wide range of astronomical and solar physics communities, and the AstroGrid consortium is a collaborator in two EU funded Grid projects, AVO and EGSO. AstroGrid is also pursuing collaboration with the US NVO project.

#### GridoneD Java on the Grid
The development of Grid-enabled problem solving environments based on the existing Cactus and Triana systems. This work forms part of the EU GridLab project. The PPARC-funded effort focuses on the application of these technologies for the gravitational wave community.

### 5.5.2 CERN-supported projects for physics data

#### DataGrid
DataGrid[2] was a project funded by the EU. The objective was to build the next generation computing infrastructure providing intensive computation and analysis of shared large-scale databases, from hundreds of terabytes to petabytes, across widely distributed scientific communities. After a very successful final review by the EC, the DataGrid project came to completion at the end of March 2004 (as a part of the recently completed Framework 5 programme). Many of the products of the DataGrid project (technologies, infrastructure etc) have fed into its successor EGEE (see following section). Its history and achievements are described below.

The European DataGrid had the aim of setting up a computational and data-intensive Grid of resources for the analysis of data coming from scientific exploration. Next-generation science will require coordinated resource sharing, collaborative processing

*2 http://eu-dataGrid.web.cern.ch/eu-dataGrid/*

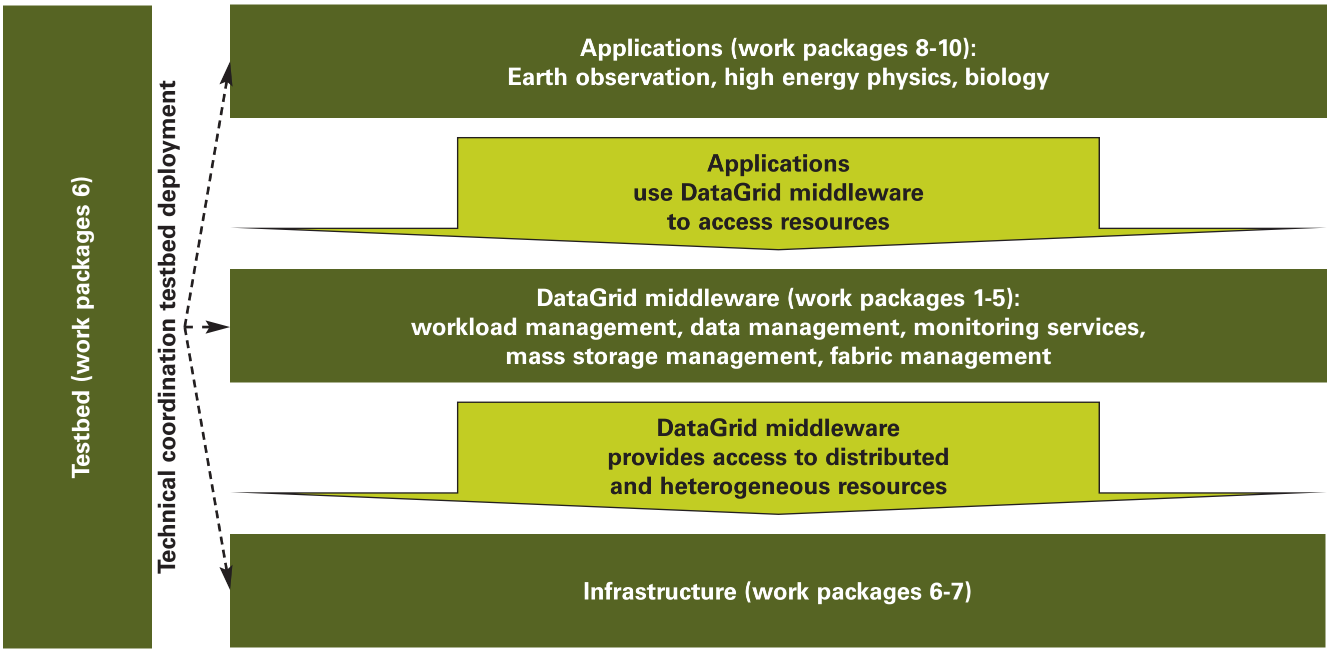


***Figure 5.9*** *The structure of DataGrid*

and analysis of huge amounts of data produced and stored by many scientific laboratories belonging to several institutions. The main goal of the DataGrid initiative was to develop and test the technological infrastructure that will enable the implementation of scientific 'collaboratories' where researchers and scientists will perform their activities regardless of geographical location. It has also allowed interaction with colleagues from sites all over the world as well as the sharing of data and instruments on a scale not previously attempted. The project has devised and developed scalable software solutions and testbeds in order to handle many petabytes of distributed data, tens of thousands of computing resources (processors, disks, etc), and thousands of simultaneous users from multiple research institutions.

The DataGrid initiative was led by CERN and was an ambitious project. Its development benefited from many different kinds of technology and expertise. The project spanned three years, from 2001 to 2003, with over 200 scientists and researchers involved, and EC funding of €9.8 million.

Figure 5.9 illustrates the structure and scope of the project and the interactions between the work packages which are detailed in the website mentioned in footnote 2.

## EEGE (Enabling Grids for e-Science in Europe)

Following on from DataGrid (as part of the EC 6th Framework Programme), the EGEE project[3] brings together experts from 70 organisations and 28 countries with the common aim of building on recent advances in Grid technology and developing a service Grid infrastructure in Europe which is available to scientists 24 hours a day. See section 5.2.2 for a description of the technical aspects of EGEE.

The project aims to provide researchers in academia and industry with access to major computing resources, independent of their geographic location. EGEE focuses on attracting a wide range of new users to the Grid. The primary objective for CERN is that EGEE will develop the underpinning technologies to enable deployment of the LCG (see section 5.2.1 for an overview of LCG).

**3** *http://egee-intranet.web.cern.ch/egee-intranet/gateway.html*

The project concentrates on three core areas:

- To build a consistent, robust and secure Grid network that will attract additional computing resources
- To continuously improve and maintain the middleware in order to deliver a reliable service to users
- To attract new users from industry as well as science and ensure they receive the high standard of training and support they need

The EGEE Grid will be built on the EU Research Network GÉANT and exploit Grid expertise generated by many EU, national and international Grid projects to date.

Funded by the EC, the EGEE project community has been divided into 12 partner federations, consisting of 70 partner institutions and covering a wide range of both scientific and industrial applications. Two pilot application domains have been selected to guide the implementation and certify the performance and functionality of the evolving infrastructure. One is the LCG supporting physics experiments and the other is Biomedical Grids, where several communities are facing equally daunting challenges to cope with the flood of bioinformatics and healthcare data. With funding of over €30 million from the EC, the project is one of the largest of its kind. EGEE is a two-year project conceived as part of a four-year programme, where the results of the first two years will provide the basis for assessing subsequent objectives and funding needs. Project Director, Fabrizio Gagliardi, said: *'EGEE will make Grid technology available on a regular and reliable basis to all European science, as well as R&D. Like the World Wide Web, which was initially developed for specialised scientific purposes, the impact of the emerging Grid technology on European society is difficult to predict at this stage but is likely to be huge.'*

## LHC Computing Grid (LCG) project

The goal of the LCG project[4] is to meet the unprecedented computing needs of LHC by deploying a worldwide computational Grid service, integrating the capacity of scientific computing centres spread across Europe, America and Asia into a virtual computing organisation.

The job of the LCG project is to prepare the computing infrastructure for the simulation, processing and analysis of LHC data for all four of the LHC collaborations. This includes both the common infrastructure of libraries, tools and frameworks required to support the physics application software, and the development and deployment of the computing services needed to store and process the data, providing batch and interactive facilities for the worldwide community of physicists involved in LHC.

The requirements for LHC data handling are very large, in terms of computational power, data storage capacity, data access performance and the associated human resources for operation and support. It is not considered feasible to fund all of the resources at one site, and so it has been agreed that the LCG computing service will be implemented as a geographically distributed Computational Data Grid. This means that the service will use computing resources, both computational and storage, installed at a large number of Regional Computing Centres in many different countries, interconnected by fast networks. Special software, referred to generically as Grid middleware, will hide much of the complexity of this environment from the user, giving the illusion that all of these resources are available in a coherent virtual computer centre.
This is an emerging technology that is at present receiving substantial R&D support from agencies that fund computing developments, and is exciting considerable interest from industry.

4 *http://lcg.web.cern.ch/LCG/*

The first phase of the project, from 2002 through 2005, is concerned with the development of the application support environment and of common application elements, the development and prototyping of the computing services and the operation of a series of computing data challenges of increasing size and complexity to demonstrate the effectiveness of the software and computing models selected by the experiments. This first phase will conclude with the production of a Computing System Technical Design Report, providing a blueprint for the computing services that will be required when the LHC accelerator begins production. This will include capacity and performance requirements, technical guidelines, costing models, and a construction schedule.

A second phase of the project is envisaged, from 2006 through 2008, to oversee the construction, commissioning and first years of operation of the initial LHC computing system.

The human and material resources required for Phase I of the project come from a variety of sources, including: resources at CERN funded both by the CERN base budget and by special voluntary contributions from countries participating in the LHC programme; industrial contributions, including resources provided by members of the CERN openlab for DataGrid applications; resources managed by the LHC experiments, at CERN and elsewhere – this is particularly important in the area of applications software development; resources provided by national funding agencies at LHC Regional Computing Centres; technology R&D projects funded by the EC and other national and regional funding agencies.

With so many elements to be managed and coordinated, the execution of the project has been organised in four different areas:

- Applications
- Computing fabrics
- Grid technology
- Grid deployment

### DataTAG

The DataTAG project (completed: March 2004) was another project working towards the common goal of providing transparent access to the massively distributed computing infrastructure that is needed to meet the challenges of modern data-intensive applications.

The fundamental objective of the DataTAG project was to create a large-scale intercontinental Grid testbed involving the European DataGrid project, several national projects in Europe, and related Grid projects in the USA. The project addressed some forefront research topics such as the design and implementation of advanced network services for guaranteed traffic delivery, transport protocol optimisation, efficiency and reliability of network resource utilisation, user-perceived application performance, middleware interoperability in multi-domain scenarios, etc.

Two types of intercontinental network connections were used: a new link for high-performance network service and data transfer application development, and the other a set of existing production links for interoperability test activities which do not require separation from production traffic. The new intercontinental connection proposed in this project was a European counterpart of GriPhyN, PPDG and iVDGL projects. This infrastructure will be made available to other EU Grid projects with similar requirements.

The major operational goals were:

1 To establish a global high performance intercontinental Grid testbed based on a high speed transatlantic link connecting

existing high-speed national and international Grid testbeds. This infrastructure will be made available to other EU Grid projects with similar requirements through the GÉANT backbone and National Research Network infrastructure.

2 To demonstrate advanced end-to-end network services across multiple domains, including:
- Demonstration and optimisation of reliable high-performance data transfer over long distance high bandwidth-delay network connections, aiming for a minimum of 100 Mbps end-to-end throughput in the early stages and Nx100 Mbps in later stages (where 5 < N < 10 at least)
- Provisioning and demonstration in a multi-domain network scenario of effective and consistent end-to-end network 'value added' services including QoS, managed bandwidth and advance network reservation
- Estimation of the real usefulness of methodologies for traffic differentiation through tools for user-perceived performance measurement

3 To develop bulk data transfer validation and applications performance monitoring tools including:
- Long-term infrastructure monitoring
- Monitoring of network performance delivered to the user (throughput and delay)
- Application-level performance monitoring

4 To insure interoperability mechanisms between US and European Grid domains, and study the scalability issues which arise, including:
- Interoperation of Grid security mechanisms
- Interoperation of information services to permit inter-Grid resource discovery
- To provide the European research community with an intercontinental test infrastructure for advanced network and application test

## 5.6 IT collaborative projects with industry

### 5.6.1 Introduction

CERN has a track record of working with partners to both develop new techniques and technologies to address their needs and to exploit these in the commercial arena. CERN's primary need is for industrial partners to work with them to meet the needs of the physics community, and one current focus of this is on the use of the Grid through the openlab for research; however, it is also working with industrial partners in major IT organisations including Microsoft, Sun and Oracle on applications, services and databases.

### 5.6.2 MammoGrid

The MammoGrid project focused on building a demonstration level pan-European database of mammography using Grid technologies. The project was not directed towards development of Grid technologies, rather towards the application of them in the domain of medical analysis of mammograms. One of the key drivers for this project was the need for high bandwidth communication to handle the very large images in situations of critical clinical needs.

The project developed a number of services, layered to provide medical services for analysis of mammograms to the client, backed by a layer of Grid aware services interfacing to the Grid middleware service layer. In this respect the technology is portable at the lower two levels for applications that require high-volume data storage and retrieval, distributed queries and handlers for results and remote activities. Specific application layers will have to be written for handling specific data models or providing specific services through the Grid. MammoGrid had a strong UK participation, with involvement from Oxford and Cambridge Universities and two UK companies.

### 5.6.3 CRISTAL

CRISTAL (Cooperating Repositories and Information Systems for Tracking Assembly Lifecycles) is a distributed workflow management tool developed by CERN with collaboration from a number of universities, including the University of West England at Bristol, to support the production and assembly phases of the CMS ECAL detector for the HEP experiment.

CRISTAL is based around an object oriented database management system, originally with a CORBA based distribution functionality although this is latterly being replaced in the next generation with Grid-based services. Meta modelling is used for description of components, tasks, collections and compositions stored in the database. This approach lends itself to development of the system to support any workflow system, providing suitable analysis is done to establish the core classes.

Currently the technology has been spun out through the French company Agilium who have developed specific extensions to the core product, particularly in the interface area. CERN continues to develop the kernel of the product to meet their own requirements and to integrate it further with their Grid infrastructures. The Agilium experience showed that there is a need to transfer the knowledge base completely and to further develop the technology based on specific business processes.

### 5.6.4 openlab

The CERN 'openlab for DataGrid applications' is a collaboration between CERN and a number of industrial partners. One of its core activities is evaluation of advanced IT, and it provides data-intensive environments using distributed architectures for integration and experimentation. The primary objective of these activities is to provide potential computing and data storage solutions to the scientific community who will be working at and with CERN, specifically with respect to LCG. However, there are opportunities for industry to develop solutions within the openlab environment that may have broader application in diverse commercial areas with similar intensive data storage/availability and processing requirements.

The principal mechanisms for involvement with the openlab are:

a As Partners, where a commitment of three years funded by a contribution of CHF 2.5 million over the period is required (£1 = CHF 2.26 at the time of publication). This contribution can be in cash or kind


*Figure 5.10* *openlab is a collaboration between CERN and industrial partners*

*Figure 5.11* *Mission participants and members of CERN's IT and TT departments in the openlab area*

b As Contributors, where the commitment is for only one year funded by a contribution of CHF 250,000 over the period. Again, this can be in cash or kind

CERN negotiates these contributions on a case-by-case basis. Typically companies provide staff seconded to CERN over the period, or supply equipment for use in the openlab, or a combination of both.

The current Partners in openlab are:

- IBM
- Hewlett-Packard
- Oracle
- Intel
- Enterasys Networks

Within openlab, thematic projects are developed between partners. The current project, opencluster, brings together the Partners, along with a Contributor (Voltaire), in order to:

- Build an ultra-high performance computer cluster
- Link it to the DataGrid and test its performance
- Evaluate potential of future commodity technology for LCG

### 5.6.5 Opportunities for UK industry

*Key potential for technology transfer*
There are opportunities for the use of CRISTAL in many activities where there is the use of detailed workflow and a need for control of manufacturing information.

*Key potential for collaborative partnerships*
The MammoGrid technology is potentially exploitable where the provision of high volumes of data for critical activities is required. Development will be required for application-level interfaces with the Grid services interface layer, and potential exploiters should be aware that the performance benefits are to a degree dependent on the performance of the underlying communications network.

In seeking to join the openlab, companies should consider how their contribution may be of value to CERN and how they will realise return on their investment. Application-specific technologies may benefit from involvement with the data and processing scale of the openlab environment; however, unless the application is of use to the CERN scientific community it may not be considered suitable for the collaboration. More generic high performance computing, data storage and networking technology is likely to be highly suitable, and UK companies developing and manufacturing such should seriously consider involvement. The contribution criteria for involvement may put off smaller companies or departments; however, CERN may be open to discussion on different models of involvement, providing appropriate levels of benefit and commitment can be identified.

## 5.7 Physics software systems

We present in this section a review of the software systems adopted by CERN in response to its undeniably challenging requirements. These include (but are not restricted to) alignment with the timescales and restraints of a major infrastructure project of unmatched scientific significance. The quantity and rate of production of data requires novel thinking and application throughout the project life cycle and across all disciplines. However, it may be useful to reflect firstly upon the similarities (both environmental and application-specific) shared by CERN and other experimental physics groups.

### 5.7.1 Similarities to other large-scale experimental physics challenges

#### Application similarities

Application similarities include the following:

- The need to capture, reconstruct, simulate, model and disseminate data for value-adding analysis
- The multi-layer approach (centralised pipeline to remote established research institutes before transfer of reduced data to a user's desktop)
- The need to carry out trade-off studies between data reduction as near to the source as possible (so reducing file transfer and database requirements downstream) against the risk of 'hard-wiring' inflexibility which inhibits future upgrade and repair

#### Environmental similarities

Environmental similarities include the following:

- The need to collaborate (funding, design, build, analysis) multinationally (requires collaborators to have experience of complex third-party relationships and working within such a complex environment)
- Any involvement in the project automatically transfers kudos, prestige and exposure (ie it is an ideal 'showcase' environment in which to trial/demonstrate capability)
- All involvement is high profile (bringing both risks and benefits)

### 5.7.2 The solution adopted by CERN

The mission team gladly acknowledges the presentation given by CERN's Pere Mato Vila as the source material for this section.

#### Physics signatures and rates

The key is that interesting physics events (or rather, their signatures) are buried in a huge background of uninteresting events: ~1 in $10^5$ to $10^9$ of all recorded events.

Note also that a different processing channel is required depending upon the type of physics (and therefore the type of signature) one is searching for. As to the quantity of data, the figures for LHC are quite astounding: bunch collisions occur at the rate of 40 MHz, resulting (as a function – inter alia – of collision cross sections) in some $10^9$ collisions per second.

#### Data processing and data sets

HEP main data are organised in events (particle collisions) and simulation, reconstruction and analysis programs process 'one event at a time'.

- Events are fairly independent of each other
- Trivial requirements for parallel processing

Event processing programs are composed of a number of algorithms selecting and transforming 'raw' event data into new 'processed' event data and statistics.

- Algorithms are mainly developed by physicists
- Algorithms may require additional 'detector conditions' data (eg calibrations, geometry, environmental parameters, etc)
- Statistical data (histograms, distributions, etc) are typically the final data processing results

### 5.7.3 Software structure and frameworks

#### Structure

The foundation is a library (widely used tools such as STL, CLHEP, GSL). Built on top of this library are framework toolkits (these include, for example, simulation and visualisation tools). Above all of this are applications which implement the required algorithms (such as high level triggers, reconstruction, simulation and analysis).

Frameworks
Frameworks are developed by experiments:

- General architecture of the event processing applications
- To achieve coherency and to facilitate software re-use
- Hide technical details to the end-user physicists (providers of the algorithms)
- Applications are developed by customising the framework
- By the 'composition' of elemental algorithms to form complete applications
- Using third-party components wherever possible and configuring them

Software components and domains

- Foundation libraries (basic types, utility libraries, system isolation libraries)
- Mathematical libraries (special functions, minimisation, random numbers)
- Data organisation (event data, event metadata (event collections), detector conditions data)
- Data management tools (object persistency, data distribution and replication)
- Simulation toolkits (event generators, detector simulation)
- Statistical analysis tools (histograms, n-tuples, fitting)
- Interactivity and user interfaces (GUI, scripting, interactive analysis)
- Data visualisation and graphics (event and geometry displays)
- Distributed applications (parallel processing, Grid computing)

Core libraries and services

The goal is to provide the software infrastructure, basic frameworks, libraries and tools that are common among the LHC experiments. This involves a requirement to select, integrate, develop and support foundation and utility class libraries. It is also necessary to develop a coherent set of basic framework services to facilitate the integration of LCG and non-LCG software. This results in:

- Foundation Class Libraries (basic types are STL, Boost, CLHEP, utility libraries, system isolation libraries, domain specific foundation libraries)
- Basic Framework Services (component model, reflection, plug-in management, incident (event) management, distributed computing, Grid, scripting)

### 5.7.4 LHC data management requirements

There is an increasing focus on maintainability and change management for core software due to long LHC lifetime. In other words it is necessary to anticipate (during the architecture design phase) the scope of changes in technology to ensure that software is able to adapt quickly to changes in environment and physics focus. Common solutions will simplify considerably the deployment and operation of data management in centres distributed worldwide – this has resulted in the establishment of POOL – the common persistency framework. Its development has required the involvement of the LHC experiments from the beginning, in order to assist its specification:

- Some experiment teams participate directly in POOL
- Some teams work with third party software providers to integrate POOL into experiment frameworks

## POOL

POOL provides persistency for C++ transient objects. It supports transparent navigation between objects across file and technology boundaries without requiring the user to explicitly open files or database connections. It follows a technology-neutral approach:

- Abstract component C++ interfaces
- Insulates experiment software from concrete implementations and technologies
- It uses a hybrid technology approach combining
  - *Streaming technology for complex C++ objects (event data)*
  - *Event data – typically write once, read many (concurrent access simple)*
  - *Transaction-safe relational database (RDBMS) services, for catalogues, collections and other metadata*
- It allows data to be stored in a distributed and Grid-enabled fashion and be integrated with an external File Catalogue to keep track of the file physical location, allowing files to be moved or replicated

## Distributed analysis

Analysis will be performed with a mix of 'official' experiment software and private user code.

It is a key prerequisite of any solution that user code can execute and provide a correct result wherever it 'lands'. This problem is made more complicated by the fact that:

- Input datasets not necessarily known a-priori
- Possibly very sparse data access pattern when only a very few events match the query
- Large number of people submitting jobs concurrently and in an uncoordinated fashion resulting in a chaotic workload
- Wide range of user expertise
- Need for interactivity – requirements on system response time rather than throughput
- Ability to 'suspend' an interactive session and resume it later, in a different location

## COTS

Throughout this entire programme, extensive use has been made of external COTS products:

- Foundation libraries (STL, Boost, CLHEP, Zlib,…)
- Mathematical libraries (NagC, GSL, CLHEP,…)
- Data organisation (Oracle, MySQL, XercesC,…)
- Data management tools (ROOT, Oracle, MySQL, EDG,…)
- Simulation toolkits (Pythia, Herwig, Geant4, Fluka,…)
- Statistical analysis tools (ROOT, GSL,…)
- Interactivity and user interfaces (Qt, Python, ROOT,…)
- Data visualisation and graphics (Coin, OpenGL,…)
- Distributed applications (PROOF, Globus, EDG,…)

# 6 GENERAL AND ADMINISTRATIVE COMPUTING

## 6.1 Internet services

### 6.1.1 Introduction

CERN operates across a geographically dispersed community on a high visibility important mission for which high quality information services are required. The provision of these services must be consistent and reliable. These services have been developed in-house using standard commercially available systems with bespoke service level applications with secure access where required. In the case of the online web based services the overall provision has achieved high standards and the look and feel is right up there with the best. What comes across strongly is the breadth of services available online to the CERN community. It is an excellent example of implementation on a large scale across a geographically distributed network.

### 6.1.2 Collaborative services

CERN provides a range of services across its entire network that includes messaging, videoconferencing and calendar services. Messaging services are via email which includes a flexible spam filter service tailored to the individual. The interface to the filter controls is via a web service on the desktop and allows for tailored filter levels including dump, quarantine for review and pass. The level of traffic experienced by the system is ~1 million emails per day, some 80% of which is spam. This is handled by 20 servers servicing 14,000 accounts.

Sharing of information is via shared directories, although CERN is looking for a more dynamic and flexible means of sharing and finding information. This would need to be an integrated facility with facilities for working collaboratively on the information across a geographically distributed domain. CERN also operates videoconferencing facilities with ~1,500 conferences per year.

### 6.1.3 Windows desktop services

The provision of a consistent desktop environment to all its users is a key driver for the IT department, which it achieves by using core sets of application software and a series of online services that provide a range of commonly needed activities. These include setting up resources on a local machine, accessing data storage facilities, installing software packages, computing systems documentation and online training for software and services.




*Figure 6.1 CERN Windows Services*

On the server side the Windows environment is serviced by 130 servers which provide 16 TB of disk space together with core facilities for providing web and mail services. The client side is extensive with 5,000 managed desktops running either Windows 2000 or Windows XP. The flexible nature of

CERN's computing environment supports 10,000 user accounts on these platforms. The support to these services is operated on a 24/7 basis by a small engineering team.

### 6.1.4 Web services

The Internet Services team provide standardised facilities for authoring and hosting of web pages to the CERN community. This is achieved via ~30 dedicated servers which handle some 1.2 million transactions per day. The support to these services is again operated on a 24/7 basis. Users are provided with online facilities which can be accessed via the internet for creating entire websites, with detailed guidance available for most activities. Users can also manage their sites with online monitoring of the status of the web servers, access details of websites and manage accessibility all using secure connections supported by HTTPS.

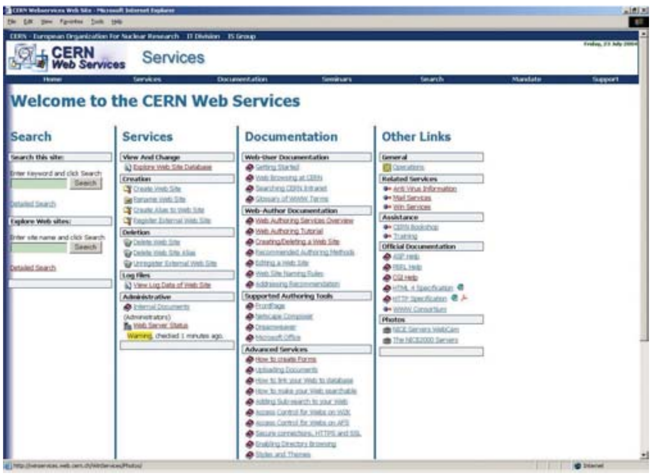




*Figure 6.2 CERN Web Services*

### 6.1.5 Opportunities for UK industry

*Key potential for technology transfer*
A number of CERN's Internet services could be adopted at little cost by UK companies.

*Key potential for collaborative partnerships*
Collaborative opportunities in developing internet services are likely to be limited, however even low level activities with the CERN team would provide good opportunities for education in large-scale implementation across distributed domains.

*Key potential for UK supply*
The CERN team have developed their internet services to a very high standard and have a high level of expertise in house, therefore it is unlikely there will be significant opportunities for supply into this area by UK companies.

There are opportunities for supply of advanced collaborative tools to CERN, the requirements for which are at this stage likely to be more aspirational than firm commitments.

In the area of videoconferencing, opportunities will arise to supply facilities to CERN or other organisations associated with physics research. However, the levels of increase in usage are unclear, with figures given indicating around 15% likely increase in use this year.

## 6.2 Document handling

### 6.2.1 Electronic document handling

Like any large organisation, CERN has a large number of official procedures – currently there are around 100: everything from buying material, requesting training, to going on holiday. These procedures are used by everyone who works at CERN, amounting to around a quarter of a million documents each year. These documents each require completion, authorisation and processing – often by several departments. The complexity is further increased by CERN's status and its international collaborations.

Over the past 10 years, CERN's IT department has replaced paper-based administrative procedures with its own customised Electronic Document Handling (EDH) system. Despite its name, EDH is an e-business tool, rather than a document management tool. EDH is implemented as a web application, handling approximately 2,000 users per day on low-cost Sun servers, with users using their desktop PCs or workstations as client systems.

The core of this system is a set of web forms built from Java components, a workflow engine using an Oracle database server, and a back-end permissions system. Although the EDH system currently uses Oracle's 'Oracle workflow' product, this was a recent replacement for an in-house solution and could be substituted with limited effort, estimated at about six months.

Documents are authorised electronically by notifying the appropriate staff and waiting for confirmation, or by prior authorisations. Each business process at CERN is modelled by a workflow, ensuring that the correct people are informed and can approve each stage electronically. The EDH automated workflow system allows the majority of documents to be processed in a single day, with over 70% of purchase requests processed within 30 minutes. Because EDH includes a permissions sub-system, even the granting of authorisation can be included within a workflow.

The EDH system IP is owned by CERN. There was a TT agreement (an exclusive right to sell, with special approval from CERN's Finance Committee) with a company but the parent company had problems and the agreement was cancelled.

The mission delegates agreed that EDH had far more to offer than had been expected. In particular it was acknowledged that the system had benefited from being driven by the users' needs, with Oracle as the enabling component. The system delivers real benefit to CERN's project management; given that this is on an extraordinary scale, the potential for UK industry is significant and this was reflected in the interest of several participants.

### 6.2.2 CERN Document Server Software (CDS)

According to CERN Convention, ARTICLE II, paragraph 1: 'The results of CERN experimental and theoretical work shall be published or otherwise made generally available' (1 July 1953).

The dissemination of material from CERN started out with standard library facilities and led, in the early 1990s, to the invention of the World Wide Web.

Since then, CERN has gone on to develop CDS[5], which offers:

- **Digital Library Application[5]**
  - *Enables an organisation to run an electronic preprint server, online library catalogue or document system on the web*
  - *Uses a commercial software package (Aleph500) for most internal library-related activities*
  - *Use of CDSware for all user-related activities:*
    - a Document Submission (400 collections)
    - b Document Harvesting (from more than 80 different sources)
    - c Document Search, Baskets, Alerts, etc (with a near-instantaneous response time for up to 1 million records)
  - *CERN specifics:*
    - a Open Access: it complies with the Open Archives Initiative metadata harvesting protocol (OAI-PMH)
    - b Non-traditional bibliographic records: eg documents with thousands of authors from large collaborations
    - c Complex Search combination required: eg combined metadata/full text/reference search

*5 http://cdsware.cern.ch (including a demonstration of Digital Library Application)* d Electronic Long Term Archive requirements (being a nuclear site, CERN has extreme archiving requirements)

– *The CDSware is free software, licensed under GNU General Public Licence (GPL)*

- CDS Agenda[6]
  – *Keeps track of most meetings at CERN*
  – *Agenda with sessions, talks and all associated material*
  – *Launched in 1999 by LHC experiments*
  – *More than 10,000 agendas and 70,000 documents at CERN*
  – *Distributed as GNU GPL*

- Integrated Digital Conferencing[7]
  – *INDICO financed by EU: 'European Solution for managing and saving Conference Content in long term'*
  – *Main Innovations: Multiple Conferences management; Conference Archiving solution; Multimedia preservation*
  – *Only in HEP and HEP-related topics: about 100 conferences per month*
  – *Considerable cost savings expected*

### 6.2.3 Opportunities for UK industry

*Key potential for technology transfer*

The opportunities for TT are primarily in out-licensing of the systems.

EDH was the unexpected star of the mission, with delegates surprised at its scope and real-world application. Departing from the current dependence on Oracle WorkFlow would require significant effort, in the order of six months, but it was seen as having strong potential and deserving of further assessment and consideration.

CDSware is currently being used or being considered by:

- San Diego Supercomputer Center, USA
- University of Missouri-Columbia, USA
- Fundao Osqaldo Cruz (Ministry of Health) Rio de Janeiro, Brasilia
- ISDN-ENSSIB, France
- Montreal International, Canada
- Bologna University, Italy
- ETH Zurich, Switzerland
- EPFL Lausanne, Switzerland
- UN Population Fund, New York, USA
- Instituto de Insvestigacions Electrica, Mexico
- Casalini Libri, Italy
- HBZ-NRW, Germany
- Aristotie University of Thessaloniki, Greece
- RERO: Consortium of all public libraries of Suisse Romande (100), Switzerland
- University of Geneva, Switzerland
- IN2P3, France
- NiKHEF, Holland
- Fermilab, USA
- ICTP, Italy

*6 Demo at http://agenda.cern.ch*

*7 Demo at http://cern.ch/indico*

# 7 CERN TECHNOLOGY TRANSFER (TT) AND IT SERVICES

## 7.1 TT at CERN

In March 1999, following a request by the CERN Member States, the Finance Committee, Scientific Policy Committee and Council asked CERN to introduce an active technology transfer (TT) policy with the aim to establish CERN's technological competence in European industrial and scientific environments. In addition, they asked CERN to demonstrate clear returns to industry on the results obtained from the considerable resources made available to particle physics research. TT can thus be seen as forming an integral part of CERN's principal mission of fundamental research.

The TT Service, which is part of the Education and Technology Transfer (ETT) unit at CERN, was set up to identify, promote, protect and transfer technologies developed at CERN. Its mission is to proactively and cost effectively transfer technology and knowledge to industry with first option to national industry of member states. The TT group's objective is to generate some revenue for CERN with minimal, if any, use of the budget for HEP research. The TT Service is headed by Jean-Marie Le Goff.

At present, the TT Service has listed about 160 technologies which are maintained in CERN's TT database. Some of them are ready for transfer now. Typically, however, since most of those technologies relate to HEP, they may require proof of concept and demonstrators to extend these applications beyond HEP.

CERN's basic policy for TT involves technology licensing, as opposed to full IP transfer. That is primarily because most technologies relate to HEP and are often reused and extended for CERN's own needs beyond a project's life cycle. TT can also involve consulting services from CERN relating to CERN's IP and know-how as required, and only when there is no direct competition with industry. Licensing of technology is usually non-exclusive. However, when required, the TT Service will offer IP protection mechanisms which depend on the type of technology and how this technology will be further used in CERN's primary activities, ie HEP. Those technologies that do not directly relate to HEP can be fully transferred to industry. Typical licensing terms include royalty fees payable to CERN, while CERN retains a perpetual licence to use and further develop transferred technology for HEP for CERN's internal use. CERN will not hold any equity in a spin-off company that exploits technology transferred to that company.

CERN has always had the intent to build and maintain strong and good relationships with industry, primarily that of member states. That means the TT Service will put in place the right mechanisms to protect any technology that is transferred to industry. For example, in a collaborative project with industry, it can be agreed at the beginning of the project what can be published or what is patentable material.

*Figure 7.1* *Nathan Hill discusses TT at CERN with the Secretary General, Maximilian Metzger*

An important restriction that constrains TT is that no technology that originates from CERN will be used directly in military applications. Another restriction is CERN staff joining startups, which is more relaxed for renewable staff such as postdoctoral researchers or fellows. However, it is recognised that TT cannot be effectively carried out if people are not involved. For this reason, CERN will ensure that part of staff time may be allocated in a TT project until the process is complete, especially for staff that were involved in the original development of the technology. Arrangements can be made for staff from industry to work together with CERN's staff to facilitate TT.

Other TT agreements include partnering with industry and performing R&D based on CERN's IP. Projects of this type are funded primarily by industry and are targeted to partner interests.

Another type of TT activity includes pre-competitive R&D projects, usually publicly funded, typically by the EU or national funding agencies. The results of such projects are shared and jointly exploited.

The income generated by all TT activities is then reinvested in TT activities or used to fund other activities and developments internal to CERN.

There are three main phases involved in the generation of new IP and know-how and its commercialisation (Figure 7.2)

Phase 1 involves R&D in the area of HEP which then leads to HEP applications. Such projects are typically financed by funds allocated to HEP research. Phase 2 involves the application of technologies developed or under development in Phase 1, to non-HEP domains and the development of a prototype or demonstrator which is of interest to an industrial partner. Phase 2 activities are funded by the TT Service and industry. The final stage is the development of a commercial product and the creation of a spin-out company if this is applicable. This stage is a fully commercial activity and is thus funded by industry. In this phase, CERN staff are usually involved on a consulting basis to ensure TT to industry. Activities in Phase 3 can also be funded by national sources to enhance dissemination of CERN's technologies and facilitate

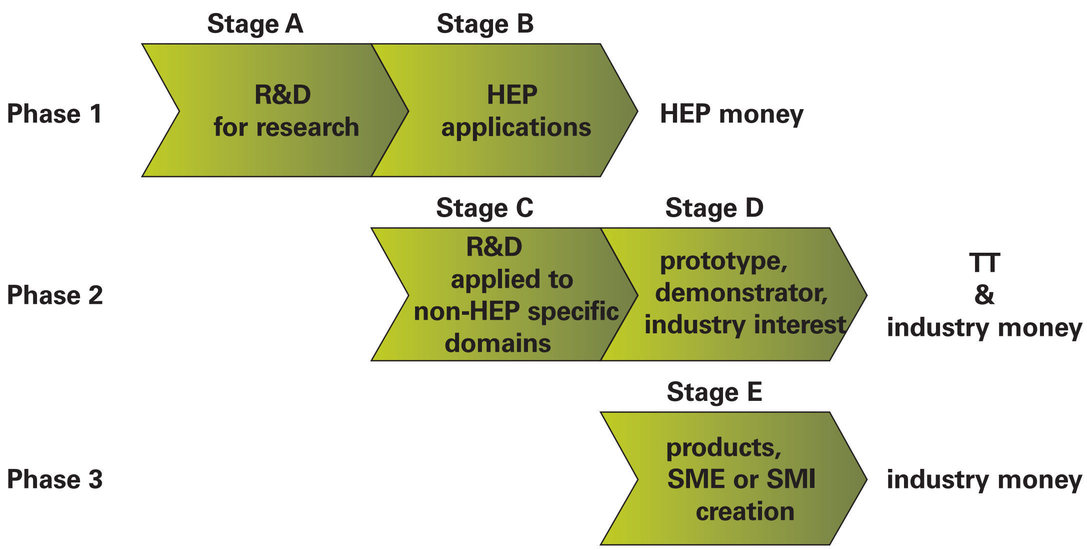


*Figure 7.2 Technology development and exploitation phases*

transfer of such technologies to industry. An example is the PIPSS programme offered by PPARC in the UK.

### 7.1.1 TT and IT

CERN's IT-specific skills include:

- High-performance computing involving round the clock operation and large-scale integration of computer resources
- Administrative computing solutions to support the large number of permanent staff, visitors and participating universities and institutes in Member States
- General purpose computing solutions such as email, desktop services, web solutions
- Large-scale distributed software engineering and Grid technologies

Those activities involve fast-changing technologies, and development on those technologies requires collaboration with external partners. This results in the need for collaboration and partnering with industry, and such collaborations are the main channel for TT in IT at CERN.

There are four main channels for such TT:

Software licensing
- Via collaboration agreements
- Via OS licensing such as GPL, LGPL, EDG, Apache

Collaboration with industry
- openlab framework (see section 5.6.4)
- Industry-led projects

Collaboration with other institutions
- EU-funded collaborative projects
- Other publicly funded projects
- Bilateral collaboration with industry

Outreach activities
- Schools and seminars on ICT
- Conferences and publications
- Other dissemination activities

For licensing software, the preferred route is to 'open-source' the core (kernel) of a software application and allow others to use the code as a library in their own applications. However, although a given SW module may have an OS licence, it can also be licensed bilaterally to a partner, allowing the distribution of modified versions under proprietary licences.

### 7.1.2 Support for contracts, TT and partnerships for UK companies

The size and complexity of CERN has made it essential for there to be a defined route for organisations seeking to partner with CERN.

Companies seeking advice and information on procurement opportunities are able to contact the UK Industrial Liaison Officer (ILO) (see Appendix F for details). Each member state has appointed such an ILO, whose role is to provide contacts and advise of opportunities at all levels of business.

The establishment of an active TT unit at CERN has led to the formation of an external TT Network of national TT Coordinators. This has received support and funding from the UK OST through PPARC. This post is held by Qi3, and details of the UK TT Coordinator (UKTTC) for CERN are also to be found in Appendix F.

Outreach to companies for TT purposes is initiated through a series of subject-specific technology and market surveys, followed by brokering meetings and workshops on themes selected as of particular interest for UK companies. Individual organisations are welcome to call on the services of the UKTTC, which are free of charge.

The UKTTC and ILO work closely together, and both encourage direct technical liaison with the relevant department in CERN and official contact through the purchasing department (in the case of procurement) and the TT unit (in the case of collaborative partnerships and IPR).

### 7.1.3 Case studies (see also section 5.6)

The CRISTAL and MammoGrid projects have been successfully transferred to industry and commercialised. Other successful examples include the small animal PET scanner project, which is currently in production, and the 3D axial brain PET scanner with long tube project.

## 7.2 Procurement of IT goods and services

### 7.2.1 Introduction

CERN has been established for 50 years and has an established methodology to determine when it will procure goods/services as compared to internal provision, and a commercially rigorous process for the subsequent procurement of the goods/services.

Overall in 2003, CERN procured CHF 18 million of computer systems, and issued CHF 6 million in computer system rental and maintenance service contracts. A further proportion of the CHF 537 million goods procurement and CHF 151 million service contracts will also have contained a proportion of IT-related activity. IT expenditure will increase significantly over the next couple of years.

This section is intended to outline the type of IT activities which CERN is undertaking in the next few years and to describe the procurement process in which UK companies will have to engage should they wish to provide goods or services.

### 7.2.2 CERN IT goods and services

The majority of IT procurement is coordinated through the IT department in support of the various services and projects required at CERN. The IT department has over 400 staff, with a further 50 fixed-term staff mainly in support of large project developments. This complement is supplemented by a large number of visiting staff seconded to CERN for learning or research activities by a wide variety of physics institutes and research councils (including PPARC).

CERN operates a desktop office Windows environment, but with in-house application developments being primarily in Java/C++ on Unix/Linux platforms.

The IT activities undertaken in CERN have a functional breakdown into:

- **Infrastructure and general services**
  This is the provision of all IT desktop, network, internet, database and back-end office applications to support the general activities of all staff based at CERN and their corporate information systems
- **Physics services**
  These activities can be generally categorised as providing the mathematical and scientific support software and IT infrastructure which will be utilised by the major projects. The primary focus is the development of Grid infrastructure
- **Support to major IT projects**
  There are currently two major projects – the LCG and the EGEE

The philosophy for IT procurement would appear to be that CERN will procure the optimum available hardware and productionised software upon which the infrastructure and physics applications are developed. Where CERN is driving the technology (eg high-speed networks, databases), it is helpful for potential suppliers to enable CERN to trial the equipment. This enables a relationship to be built such that suppliers can influence the technical specifications of subsequent procurements and CERN can ensure a pool of competitive suppliers. This can have an effect beyond the supply of IT to CERN itself. As CERN is at the centre of particle physics research in Europe, other collaborating institutes are likely to procure similar products, and hence a successful relationship with CERN may enable wider and easier market access. This is likely to be particularly applicable for the provision of computing, storage and network facilities in the LHC tier-1 Grid.

In contrast, the development of customised or bespoke applications is undertaken in-house or in collaboration with a number of international institutes. This approach is driven by the availability of seconded manpower, and the challenges of working with complex specifications in a customer-supplier relationship.

CERN has a small number of IT-related service contracts. Most notable is the multi-year provision of desktop support services to the CERN community, which has recently been won by a UK-based company.

### 7.2.3 CERN IT procurement process

The CERN procurement principle is that contracts are awarded to the lowest priced bidder which satisfactorily complies with the issued technical, commercial and programmatic requirements. There is no concept of best-value-for-money evaluation.

This principle is only amended when the industrial return to a CERN member country is 'poorly balanced'. CERN measures the return coefficient of a Member State which is defined as the ratio between that Member State's percentage share of the value of all supply contracts and that Member State's percentage contribution to the CERN budget over the same period. If the coefficient is less than 0.92 for supply contracts, or less than 0.40 for industrial services contracts (these are 2004 figures), then the member state is described as 'poorly balanced'.

In such cases, the lowest bidder from a poorly balanced country will be invited to match the price offered by the winning bidder and thereby be awarded the contract. It is important to note that in the current financial year, the UK is a poorly balanced country for supply contracts in 2004. Hence UK companies will obtain a small advantage on tender evaluation in the short term.

The tender release procedure varies dependent upon the size of the contract:

- In the case of small tenders (< CHF 200,000) CERN will not necessarily publish the tender, but will approach a restricted list of 3 to 5 companies that have been internally identified as being able to provide the goods/services
- For larger tenders (> CHF 200,000) CERN publish on their website an advance announcement of the forthcoming tender. Interested companies are invited to submit a pre-qualification 'market' survey. On the basis of the received answers, CERN will select up to 10 companies to receive the full tender (up to 15 companies if the value is > CHF 750,000)

The process above applies both to IT-related procurements and all other procurements. UK firms wishing to reply to a CERN market survey are advised to discuss their intentions with the relevant PPARC contact point. The URL for CERN purchasing is: http://spl-purchasing.web.cern.ch/spl-purchasing

### 7.2.4 Opportunities for UK industry

*Key potential for technology transfer*

It is clear that CERN has a technology lead in several areas including Grid software, data storage and high-speed networking. There are also several in-house administrative computing packages which might be used elsewhere in similar academic/semi-academic heterogeneous environments. The specific TT opportunities have been detailed earlier in this report.

### *Key potential for collaborative partnerships*

Existing collaborative partnerships such as openlab have the potential to be very useful to product-oriented companies such as the major vendors of data storage devices, database management products, networking switches and routers. CERN benefits from trialling the available market products and the companies obtain a cost-effective intensive test environment.

The potential benefits for software engineering companies to participate in the currently proposed collaboration packages are not so high. A different paradigm will need to be found. An option may be to choose a particular service which is of interest to the commercial markets and which is also needed as part of the LHC Grid. Alternatively, participation with CERN on European-wide EU programmes in which Grid technology is prototyped in non-particle physics environments would have some advantage. The key question is how far Grid technology is from being applied to the commercial markets.

Worryingly, there are no accepted common Grid software standards as yet – though work is ongoing in this area. Major investment by individual software engineering companies is therefore at risk of being superseded by the future direction of the LHC particle-physics-led Grid activities.

A potential collaboration opportunity with CERN is in the area of advanced IT security services and authentication systems applied to the Grid which may provide an important area of procurement in many IT markets in the next 3-5 years.

### *Key potential for UK supply*

It is not believed that CERN holds any similar large potential for the supply of IT services or bespoke software procurements.

The following areas are believed to show considerable potential for supply from UK based companies:

- Dual processor PCs (CERN requires several thousand PCs linked together in processor farms)
- Storage devices (disk cache and tape-based systems capable of storing up to 10 PB per annum)
- Database management systems (primarily Oracle)
- Routers, network devices and communications (capable of 1-10 GB/s)

These procurements will likely be made in the period 2005-2007. Suppliers should also target the LHC tier-1 centres.

# 8 ANSWERS TO KEY QUESTIONS

## 8.1 CERN in general

*Who owns the IP for CERN IT developments?*

Many software developments are in-house, with the IP owned by CERN. There are numerous international collaborations, each of which has its own IP agreement.

*What are the regulations and policies for IP at CERN?*

As an international scientific research organisation, CERN's mission is to make everything freely available to the scientific community. Much software is available under OS arrangements.

An important restriction is that, as a nuclear physics research organisation, CERN cannot license technology for military applications.

In general, British companies should approach the UK TT Coordinator in the first instance for information.

*How is technology transfer managed at CERN?*

CERN has an Education and Techology Transfer (ETT) Unit, reporting to the Director General. Its mission is to identify, protect, promote and transfer technologies developed at CERN, with preference given to the European member states including the UK.

*How can UK companies access technology opportunities at CERN?*

The UK has an industrial Liaison Officer (ILO) to support contract opportunities and a Technology Transfer Coordinator (UKTTC) to support partnerships and TT.

## 8.2 Software

*Is software developed for high energy physics useful in broader industrial applications?*

Software is developed for CERN and its scientific collaborators. The nature of CERN's industrial partnerships facilitates broader application of software developed. The most significant past example of general use for CERN software is the development of the World Wide Web.

*How is software maintenance managed?*

Software is maintained by the relevant CERN department (in general, the IT or the PH department).

*How is CERN software licensed?*

CERN software is generally made available to the scientific community under OS licensing.

*What are CERN's software innovations?*

CERN is currently an early adopter and key player in Grid middleware. CERN leads the LCG project to distribute and analyse data from the LHC. CERN also collaborates on the adoption of Grid technology in other areas through various projects, including the GridPP project with UK particle physicists.

CERN's administrative computing includes an interesting workflow system which has previously been licensed for commercial development.

## 8.3 Standards

*How does CERN participate in and support IT standards?*
CERN is a leader in the development of Grid middleware and in high speed networking. No standards are yet widely accepted in Grid computing.

## 8.4 Hardware

*How is specialised hardware procured?*
CERN collaborates actively with manufacturers to drive the technical specifications of high-technology equipment, according to the needs of high-energy physics research, and in return has a pool of competitive suppliers and control over the quality of hardware.

*How is general purpose hardware procured?*
General purpose hardware, available from a broad range of suppliers, has a tender procedure with the contract going to the lowest priced bidder. For CERN Member States which are providing less than their expected share, the lowest national bidder is offered the further opportunity to match the lowest bid. This offers an advantage to UK bidders in the current financial year.

## 8.5 Networking

*What are CERN's special networking requirements?*
When CERN's LHC experiments begin to generate data in 2007, data will be generated at extremely high rates. The data is filtered in real time, passed to the central Grid at CERN to be stored on disk and tape, and copied to national Grid centres.

*What are CERN's network innovations?*
CERN is working on high-bandwidth networking for data collection from the LHC experiments.

## 8.6 Science

*How does CERN IT support high energy physics research?*
HEP experiments are supported both by CERN and by the large experiment project teams.

*How can CERN's IT be applied to other scientific fields?*
CERN technology is generally applicable to biomedical research, telecommunications, and large-scale distributed computing, including bioformatics and e-business.

## 8.7 Administrative computing

*How does CERN IT support the CERN administration?*
The CERN IT department provides the computing infrastructure needed for CERN staff and visitors. It also provides customised software for CERN administration, and for CERN's international collaborations.

# 9 CONCLUSIONS

The mission was given open access to CERN's staff and know-how, in a programme which covered many state-of-the-art technology advances, and was able to identify not only those technologies which have potential for valuable transfer to UK industry but also the instruments for doing so. This is in no small part thanks to the efforts of the CERN IT and TT teams and their colleagues. This welcoming and open attitude is a strong advertisement for UK industry to look closely at opportunities for collaborating with CERN, not only for TT but also for the development and implementation of data management, Grid technologies and e-business applications.

The mission team drew the following conclusions as a result of their visit:

### General

- IT developed in pursuit of CERN's HEP goals is readily transferable to generic IT applications. Companies should endeavour to look beyond the complexities of HEP to the applications pertinent to their domains
- CERN is a developer and user of IT to suit its focused applications and not as an academic pursuit in its own right. The emphasis of the IT department is on practical implementation
- CERN has a skills base in Grid deployment and a team of experts that is clearly at the leading edge in this field
- CERN has developed means to protect IP, to license software and to work in successful international collaborative partnerships
- The UK has developed support for TT and contract opportunities. In this respect, it is more proactive than many other Member States and this represents an opportunity for UK companies
- The IT department is well organised and able to work with industry

### Distributed computing

- CERN is deploying Grids now – not just talking about it
- The UK is well regarded as a significant player in the LCG and openlab
- Middleware and high-speed networking are well advanced at CERN
- The openlab and EGEE provide means for industry to engage with these advances
- Network and Grid security were highlighted as a specific opportunity for collaboration; a workshop will be held in October 2004 to explore this further

### Distributed applications

- The e-business system EDH was the surprise star of the mission and could be adopted by UK companies through licensing
- CRISTAL also has significant potential for industry
- A number of CERN's internet services are well-developed and may be used by companies through OS licences

### Supply

- The UK is performing well in the supply of hardware and desktop services to CERN
- The procurement regulations at CERN are currently weighing in favour of UK companies because of our overall poor return
- There are plenty more opportunities in this area for supply to LHC and its four experiments
- There are also opportunities in supply of high-speed data acquisition electronics to the experiments

*Figure 9.1* *Mission participants standing on the beam line of the LHC*

- Opportunities in software were harder to define, with CERN and its collaborating institutes supplying many of its needs
- Companies should also consider sales to the tier-1, 2 and 3 centres
- openlab is the place to benchmark products

Partnership
- The UK is taking an active approach towards TT
- PPARC is also supporting TT through its PIPSS scheme
- UK companies are not taking sufficient advantage of the opportunities presented through the support offered – this represents a risk to UK leadership in Grid deployment

# Appendix A

## MISSION PARTICIPANTS

Nathan Hill
Qi3 – Industry Coordinator for PPARC & UK Technology Transfer Coordinator for CERN
*nathan.hill@qi3.co.uk*
*www.qi3.co.uk*

In 1999, Nathan started his own business, Qi3, which is a specialist technology commercialisation business, acting to support the sales and marketing activities of technology companies and startups. As part of this activity, Nathan works for PPARC as Industry Coordinator and UK Technology Transfer Coordinator for CERN, building partnerships between the PPARC-funded research groups and UK industry.

Nathan studied Physics at Oxford University (specialising in Condensed Matter) and then moved to Germany for three years, selling cryogenic and superconducting magnet systems for Oxford Instruments. He returned to Oxford Instruments in 1992, and as Marketing & Strategy Director headed up the marketing activity for research and novel industrial applications of cryogenics and superconductivity. He moved to Cambridge in 1996 as Managing Director of two companies, one specialising in superconducting thin-film detectors and SQUIDs, the other a manufacturer of scanning tunnelling microscopes. During this period he evaluated numerous companies for acquisition or disposal. Much of Nathan's work is in the development of 'physics to healthcare' opportunities and strategic alliances. Recently he was invited to speak at the International Patent Licensing Seminar in Tokyo. Nathan now supports a number of companies in startup or rapid growth modes.

Qi3 was the coordinating body for this mission. Qi3 is particularly suited to lead this mission as it already has a contract supported by OST through PPARC as UK Technology Transfer Coordinator for CERN. This gives the team strong existing links with the IT and TT departments at CERN and the UK IT and technology translation networks.

Ted Ridgway Watt
International Technology Promoter
DTI Global Watch Service
Pera Innovation Ltd
*ted.ridgway-watt@pera.com*
*www.globalwatchonline.com/itp*

Ted joined the International Technology Promoters (ITP) element of DTI's Global Watch Service in 2001, to address the Information Technology, Electronics and Communications (ITEC) sector, with a focus on Europe. His experience of TT in the European region complements the network's coverage of the Asia-Pacific and North America regions.

Europe has established a globally competitive performance in the ITEC sector, and Ted's role is to help innovative UK companies benefit from overseas technology developments by identifying partnership opportunities, facilitating access to overseas technology providers, and assisting in TT projects.

The UK already has extensive capability and a wide competence base in electronics and communications but the increasing performance demands of future generations of communications systems underline not only the need for timely acquisition of technology but also effective exploitation.

Ted believes that the ITP approach of forging technology partnerships and encouraging inward TT projects is a valuable way to establish overseas links that will help the UK and its partners.

Ted came to the Global Watch Service from the Defence Diversification Agency, where he was the Technology Diversification Manager for the South East of England, responsible for TT projects between public research institutes and SMEs. Prior to that he was based at the DERA office in Brussels, responsible for developing collaborative research and TT projects with European partners.

The DTI Global Watch Service is managed by Pera Innovation Ltd.

Peter Rice
European Bioinformatics Institute (EBI) – Mission Leader
*pmr@ebi.ac.uk*
*www.ebi.ac.uk*

EBI is a non-profit academic organisation that forms part of the European Molecular Biology Laboratory (EMBL). EBI is a centre for research and services in bioinformatics, and manages databases of biological data including nucleic acid, protein sequences and macromolecular structures.

EBI's mission is to ensure that the growing body of information from molecular biology and genome research is placed in the public domain and is accessible freely to all facets of the scientific community in ways that promote scientific progress.

Peter is an academic group leader at EBI, developing Grid and web services to integrate bioinformatics applications, databases and workflows, and to make them available as services to the research community. Peter is also the originator of the EMBOSS open source (OS) bioinformatics project.

Platform technologies include Grid technology, web services, workflows, application integration, biological data integration, OS software, and bioinformatics.

Derek Greer
VEGA Group Plc
*derek.greer@vega.co.uk*
*www.vega.co.uk*

VEGA is an established consulting and technology company that implements strategy for international businesses and government organisations. It works across the public and private sectors, with clients in government, defence, space and commercial markets and delivers services worldwide with offices in the UK, Germany and France.

Derek is the Head of Engineering Services with VEGA. He is responsible for management of VEGA's engineering and software support activities with clients in the aerospace and related high-technology markets. Derek has a PhD in Satellite Engineering and Space Science, and has 20 years' technical and project management experience, primarily in the Space sector encompassing both in-orbit and ground systems. Derek has worked in the UK, the Netherlands and Spain.

A key facet of Derek's activities is to understand and promote the application of knowledge and technology to deliver benefits to VEGA's clients. He has, therefore, an interest in following the emergence of next generation internet/Grid technologies, in particular their applications outside academic environments.

Andrew Hide
LogicaCMG
*andrew.hide@logicacmg.com*
*www.logica.co.uk*

LogicaCMG is a major international force in IT services and wireless telecoms. It provides management and IT consultancy, systems integration and outsourcing services to clients across diverse markets including telecoms, financial services, energy and utilities, industry, distribution and transport and the public sector. Formed in December 2002, through the merger of Logica and CMG, the company employs around 20,000 staff in offices across 34 countries, and has nearly 40 years' of experience in IT services.

Andrew is an ex-RAF Officer with a Doctorate in Condensed Matter Physics. He has worked in a variety of commercial, academic and military environments (most recently the Defence Communications Services Agency in the UK and the Isaac Newton Group of Telescopes in the Canary Islands) in which he has specialised in Spacecraft, Ground Segment and Instrumentation technologies. He is currently a senior consultant within LogicaCMG's Space and Defence division engaged as a bid manager.

His interest in Grid technology is principally in investigating the potential for the development of generic applications for complex, secure and high-throughput data processing and dissemination in the fields of defence, navigation and earth observation.

Iain le Duc
CODASciSys
*Iain.le.Duc@scisys.co.uk*
*www.scisys.co.uk*

CODASciSys (formerly Science Systems) is a publicly listed company formed in 1980 and employing over 850 people in the UK and throughout the world. SciSys is the software and consultancy division of the Group that builds technology solutions. Its Space & Defence (S&D) division focuses on both the space and ground segments of, largely European, satellite programmes and Battlefield Information Systems Applications and Synthetic Environments of defence.

SciSys S&D pride themselves on keeping up-to-date with all technology areas that may have relevance to their business sectors. Iain is the Principal Consultant leading this activity within S&D, having a special interest in emerging technologies and development methodologies. In particular, Grid technologies have much to offer SciSys business sectors. SciSys has been involved in the UK Government's e-Science Grid programme since 2000 exploring the potential for Grid-based software technologies for databases and implementing Grid parallel processing prototypes. SciSys has established links with the e-Science and CEOS-Grid programmes underway in the UK. SciSys has been a major partner in the recently completed European Space Agency (ESA) SpaceGrid project and was responsible for developing software implementations of two separate aspects: spacecraft-plasma interactions and solar system research.

Iain is also involved with studies into how European organisations can work better in 'collaborations' or 'vertical organisations' for ESA, with a major aim being to improve the use of Earth Observation data access and usage. This is a multi-dimensional problem covering sectors as diverse as forestry, agriculture and Global Monitoring for Environment Security; and employing technologies as diverse as peer-to-peer, Grid (as used in DataGrid), distributed document control, sensor web, semantic web, videoconferencing, web services and workflow management.

Dr Harris Makatsoris
Orion Logic Ltd
*h.makatsoris@orionlogic.com*
*www.orionlogic.com*

Harris is a founder and managing director of Orion Logic Ltd, a UK-based software and services vendor of adaptive supply chain optimisation and execution software.

Orion Logic's offerings comprise software technology and technical consulting services that enable operational control across extended manufacturing networks, enabling event capturing, event resolution management, cycle time compression and waste elimination. The company's target market includes semiconductor/high-tech in addition to other target industries such as metalworking and textiles.

Harris holds a degree in Mechanical Engineering and a doctorate degree in Computer Aided Systems Engineering from Imperial College, London, with a specialisation in Production & Supply Chain Optimisation. He has published a number of papers in the area of production, supply chain and virtual organisation management and optimisation including the latest book entitled 'Evolution of supply chain management, Symbiosis of Adaptive value Networks and ICT', Kluwer Academic, 2004.

Grid computing and other technologies developed at CERN are of significance to Orion Logic because they are aligned with Orion's technological development plans. For those reasons, Orion regards a relationship with CERN of importance with particular focus on the following:

- Collaborative development of a commercial Grid middleware for large scale optimisation and simulation primarily targeting supply chains and other optimisation applications
- Commercialisation of the e-business platform developed at CERN
- Use of CERN's Grid infrastructure as a testbed for Orion Logic's technologies

Dr Maziar Nekovee
BT Exact
*maziar.nekovee@bt.com*
*www.btexact.com*

BT Group is one of Europe's leading providers of telecommunications services. Its principal activities include local, national and international telecommunications services, higher-value broadband and internet products and services, and IT solutions. In the UK, BT serves over 20 million business and residential customers with more than 29 million exchange lines, as well as providing network services to other licensed operators.

Maziar gained his first degree in electrical engineering from Delft University of Technology in the Netherlands and his PhD in theoretical and computational physics from the University of Nijmegen, also in the Netherlands. Prior to joining BT in 2001, he was a research fellow at the Centre for Computational Science, Queen Mary College, London, where he worked on high-performance computing and visualisation of classical and quantum many-particle systems.

Maziar is a senior research scientist at BT Group's Research and Venturing division where he leads a collaborative e-Science project on Grid computing. Areas of research include high-performance multicast and unicast data transport for the Grid, mobility and peer-to-peer networks. Maziar Nekovee is the author of over 30 scientific papers in international peer-reviewed journals, a number of patents, and several publications on broader issues in science and technology.

David Palmer
QinetiQ
*wdpalmer@qinetiq.com*
*www.qinetiq.com*

David is a Business Group Manager in the Command & Intelligence Systems centre at QinetiQ Limited with an extensive background in software and systems engineering across a range of industries.

QinetiQ is a science and technology powerhouse with 8,000 scientists and engineers at 62 working locations in the UK, the largest security and defence research and technology business in Europe. It is a company with expertise that encompasses many fields, including energy and propulsion, sensors and electronics, chemicals and materials, software and systems, telecommunications and data, security, geosciences and space, human factors and medical science. While maintaining its strong link with the UK MoD forged as the Defence Evaluation and Research Agency, almost 20% of QinetiQ's business is non-defence.

David's group conducts R&D into new ways to integrate people and IT for knowledge management and collaborative working. This work includes semantic web, software agents, machine learning, natural language processing, peer-to-peer and Grid computing.

Paul Perkins
Atos Origin
*Paul.Perkins@AtosKPMGConsulting.co.uk*
*www.atosorigin.com*

Atos Origin is a leading international IT services company and a global provider of business consulting and technology integration services. Atos is the exclusive IT provider for the Olympic Games.

Common to all Atos Origin solutions is a unique full service, full life cycle, 'design, build and operate' approach. Atos craft front-to-back office, end-to-end integrated services for clients across all functional, technical and geographic parameters. These can be tailored to meet all local and regional requirements and are delivered worldwide.

Paul has supported Oracle implementations at BT, ntl and Vodafone. He is currently at Canon Amsterdam, consulting on a pan-European implementation.

Paul's interest in Grid technology is principally in investigating the potential for transferring and popularising this technology for commercial use. Utility Computing is one of the products being offered by Atos, but at present it has many limitations and the market is very small. He wants to develop, expand and demonstrate that this brand can be a viable solution and can be exploited in the many fields Atos are supplying in both public and private sector industries.

# Appendix B

## CERN PARTICIPANTS

Director General's Office

| Name | Function | Group | e-mail |
|---|---|---|---|
| Maximilian Metzger | Secretary General | SG-DI | maximilian.metzger@cern.ch |

SG/ETT Unit Participants

| | | | |
|---|---|---|---|
| Jean-Marie Le Goff | Technology Transfer Service Group Leader | ETT-TT | jean-marie.le.goff@cern.ch |
| Marilena Streit Bianchi | TT Operations, Section Leader | ETT-TT | marilena.streit-bianchi@cern.ch |
| Roberto Amendolia | TT Promotions, Section Leader | ETT-TT | salvator.amendolia@cern.ch |
| Nelly Castilla | TT Promotions, group member | ETT-TT | nelly.castilla@cern.ch |

IT Department Participants

| | | | |
|---|---|---|---|
| Wolfgang von Rüden | Department Head | IT-DI | wolfgang.von.rueden@cern.ch |
| John Ferguson | Deputy Department Head | IT-DI | john.ferguson@cern.ch |
| Les Robertson | LCG Project Leader | IT-LCG | les.robertson@cern.ch |
| François Fluckiger | Departmental Technology Transfer Officer/ Manager of the openlab/Director of CERN School of Computing | IT-DI | francois.fluckiger@cern.ch |
| Reinoud Martens | Administrative Information Services (AIS) Group Leader | IT-AIS | reinoud.martens@cern.ch |
| Jamie Shiers | Databases (DB) Group Leader | IT-DB | jamie.shiers@cern.ch |
| Tony Cass | Fabric Infrastructure and Operations (FIO) Group Leader | IT-FIO | tony.cass@cern.ch |
| Ian Bird | Grid Deployment (GD) Group Leader/ LCG Deployment Area Manager & EGEE Operations Manager | IT-GD | ian.bird@cern.ch |
| Alberto Pace | Internet Services (IS) Group Leader | IT-IS | alberto.pace@cern.ch |
| Mick Draper | User and Document Services (UDS) Group Leader | IT-UDS | mick.draper@cern.ch |
| François Grey | Strategy and Communications Officer | IT-DI | francois.grey@cern.ch |
| Robert Jones | Technical Director, EGEE project | IT-EGE | robert.jones@cern.ch |
| Olof Bärring | Architecture and Data Challenges (ADC) Group Member, CA section | IT-ADC | olof.barring@cern.ch |
| Derek Mathieson | Administrative Information Services (AIS) Group, EB Section Leader | IT-AIS | derek.mathieson@cern.ch |
| James Purvis | Administrative Information Services (AIS) Group, Deputy Group Leader and FLP Section Leader | IT-AIS | james.purvis@cern.ch |
| Jurgen de Jonghe | Administrative Information Services (AIS), PM Section Leader | IT-AIS | jurgen.de.jonghe@cern.ch |
| Jean-Yves Le Meur | User and Document Server (UDS) Group, CDS Section Leader | IT-UDS | jean-yves.le.meur@cern.ch |
| Jean-Michel Jouanigot | Communication Systems Group, Campus Network Leader | IT-CS | jean-michel.jouanigot@cern.ch |

| Name | Function | Group | e-mail |
|---|---|---|---|
| Olivier Martin | Communication Systems Group,<br>EU Data Tag Project Leader,<br>External Networks Leader | IT-CS | olivier.martin@cern.ch |
| James Casey | GRID Deployment Group member,<br>GIS section | IT-GD | james.casey@cern.ch |
| Ian Neilson | GRID Deployment Group member,<br>GIS section, LCG Security Officer | IT-GD | ian.neilson@cern.ch |
| **FI Department Participants** | | | |
| Anders Unnervik | Purchasing & Industrial Services Group<br>Deputy Group Leader | FI-PI | anders.unnervik@cern.ch |
| **PH Department Participants** | | | |
| John Harvey | Software Development for Experiments<br>(SFT) Group Leader | PH-SFT | john.harvey@cern.ch |
| Pere Mato Vila | Software Development for Experiments<br>Experiments (SFT) Group Member | PH-SFT | pere.mato@cern.ch |
| John Apostolakis | Software Development for Experiments<br>Experiments (SFT) Group Member | PH-SFT | john.apostolakis@cern.ch |
| René Brun | Software Development for Experiments<br>Experiments (SFT) Group Member | PH-SFT | rene.brun@cern.ch |
| Alberto Aimar | Software Development for Experiments<br>Experiments (SFT) Group Member | PH-SFT | alberto.aimar@cern.ch |
| Wisla Carena | ALICE Trigger/DAQ (AID) Group,<br>Group Leader | PH-AID | wisla.carena@cern.ch |
| Pierre Vande Vyvre | ALICE Trigger/DAQ (AID) Group Member | PH-AID | pierre.vande.vyvre@cern.ch |
| Klaus Schossmaier | ALICE Trigger/DAQ (AID) Group Member | PH-AID | klauss.schossmaier@cern.ch |
| Fabio Formenti | Electronic Design (ED) Group Leader | PH-ED | fabio.formenti@cern.ch |
| Alex Kluge | Electronic Design (ED) Group Member | PH-ED | alex.kluge@cern.ch |

# Appendix C

## MISSION PROGRAMME

Programme Coordinators
Mission: Nathan Hill, Qi3; CERN: François Fluckiger

Monday 28 June 2004

| Time | Organiser | Theme | Topic | CERN speakers/participants | Min | Place |
|---|---|---|---|---|---|---|
| 08:30 -09:00 | N Hill | Welcome Presentation of Mission by Mission Leader | Welcome Presentation of Mission by Mission Leader | | 30 | Bd 61, Room B |
| 09:00 -10:00 | ETT | Presentation of CERN | Introduction to CERN and TT | Maximilian Metzger Jean-Marie le Goff | 60 | |
| 10:00 -12:00 | ETT | Visit of Atlas | Organised by ETT Unit | NA | 120 | Atlas site |
| 12:00 -13:15 | ETT | Lunch | Organised by ETT Unit CERN Participants tbc by ETT | Roberto Amendolia François Fluckiger Jean-Marie le Goff Maximilian Metzger Marilena Streit Bianchi Wolfgang von Rüden | 75 | Main Restaurant Glass Box |
| 13:15 | IT | Transport to Building 513 | Organised by IT | NA | 15 | |
| 13:30 -15:45 | ETT | ETT related activities and projects | IP issues in TT at CERN | Robert Abbott | 20 | openspace Room 513-R-070 |
| | | | The Mammogrid project | Roberto Amendolia | 30 | |
| | | | CRISTAL: A process and data management systems. A spinout from the CMS experiment | P Bayle CEO of AGILIUM | 30 | |
| | | | TT through procurement activities study | Marilena Streit-Bianchi | 20 | |
| | | | Discussion | Those above | 30 | |
| 15:45 -16:00 | IT | Coffee Break | NA | NA | 15 | openspace |
| 16:00 -17:30 | IT | IT Department (General) Procurement | IT Department: mission, goals, skills | Wolfgang von Rüden | 15 | openspace Room 513-R-070 |
| | | | General services and Administrative Computing | John Ferguson | 15 | |
| | | | Technology Transfer in IT | François Fluckiger | 10 | |
| | | | CERN procurement principles | Anders Unnervik | 20 | |
| | | | Discussion | Those above and François Grey | 30 | |
| 18:30 | N Hill | Reception at British Consulate | Transport organised by N Hill | NA | | British Consulate |

Tuesday 29 June 2004

| Time | Organiser | Theme | Topic | CERN speakers/participants | Min | Place |
|---|---|---|---|---|---|---|
| 08:30 -10:30 | IT | The LHC Computing and the CERN openlab | Introduction to Physics Computing | Les Robertson | 10 | openspace |
| | | | The LCG (incl timeline for procurement) | Les Robertson | 20 | |
| | | | openlab (overview) | François Fluckiger | 15 | |
| | | | openlab (technical) | Sverre Jarp | 15 | |
| | | | Discussion | Those above and John Ferguson Wolfgang von Rüden | 60 | |
| 10:30 -10:45 | IT | Coffee Break | NA | NA | 15 | openspace |
| 10:45 -12:45 | IT | Grid technologies and EGEE | EGEE (overview) | Robert Jones | 15 | openspace |
| | | | Grid Middleware | Robert Jones | 15 | |
| | | | Grid Deployment | Ian Bird | 15 | |
| | | | Grid Security | Ian Neilson | 15 | |
| | | | Discussion | Those above and John Ferguson Wolfgang von Rüden | 60 | |
| 12:45 -13:30 | IT | Lunch | (sandwiches and refreshment) | NA | 45 | openspace |
| 13:30 -14:30 | N Hill | Visitors closed meeting | NA | NA | 60 | openspace |
| 14:30 -15:30 | IT | Storage and Data Management Technologies | Grid Data Management | James Casey | 10 | openspace |
| | | | Databases and Pool | Jamie Shiers | 10 | |
| | | | CASTOR | Olof Bärring | 10 | |
| | | | Discussion | Those above | 30 | |
| 15:30 -15:45 | IT | Coffee Break | NA | NA | 15 | openspace |
| 15:45 -17:30 | IT | Innovative General and Administrative Computing | Innovative Desktop services (including NICE, web, print, anti-spam services) | Alberto Pace | 15 | openspace |
| | | | Workflow solutions, EDH and E-commerce | Derek Mathieson or James Purvis | 10 | |
| | | | Collaborative Project Management Tools (EVS) | Jurgen de Jonghe | 10 | |
| | | | Summarised overview of AIS activities | R Martens | 10 | |
| | | | Conference management (CVS, Indico) | Jean-Yves le Meur/ Mick Draper | 10 | |
| | | | Discussion | Those above and John Ferguson | 50 | |
| 17:45 -18:45 | IT | Visit of IT Computer Centre/Social Time | NA | NA | 60 | CC, Bd 513 |
| 18:45 | N Hill | Transport to Restaurant | NA | NA | 15 | |
| 19:00 | IT | Dinner at restaurant Au Renfort (in Sézegnin) | | CERN John Ferguson François Fluckiger John Harvey Jean-Marie le Goff Les Robertson Marilena Streit-Bianchi Pierre Vande Vyvre Wolfgang von Rüden | | Au Renfort Sézegnin |

Wednesday 30 June 2004

| Time | Organiser | Theme | Topic | CERN speakers/participants | Min | Place |
|---|---|---|---|---|---|---|
| 08:30 -10:30 | PH/SW | Physics Software Session coordinated by John Harvey | Introduction to Physics Software (requirements, basic applications, spectrum of components) of components) | Pere Mato Vila | 15 | openspace |
| | | | Geant4 | John Apostolakis | 15 | |
| | | | Root- Proof | René Brun | 15 | |
| | | | Software Engineering (code and release management, external libraries, testing processes and tools) | Alberto Aimar | 15 | |
| | | | Discussion | Those above | 60 | |
| 10:30 -10:45 | IT | Coffee Break | NA | NA | 15 | openspace |
| 10:45 -12:45 | PH/Alice | Physics on-line technologies: the ALICE experiment Session coordinated by Wisla Carena | Challenges of Data Acquisition at the LHC | Pierre Vande Vyvre | 30 | openspace |
| | | | Evaluation of computing technologies for the ALICE data acquisition | Klaus Schossmaier | 20 | |
| | | | Electronics integration of the ALICE Pixels tracking detector | Alex Kluge | 20 | |
| | | | Discussion | Those above and Wisla Carena Fabio Formenti | 50 | |
| 12:45 -13:30 | IT | Lunch | (sandwiches and refreshment) | NA | 45 | openspace |
| 13:30 -14:30 | N Hill | Visitors Closed meeting | NA | NA | 60 | openspace |
| 14:30 -15:30 | IT | High Performance Networking | High Performance LAN Networking | Jean-Michel Jouanigot | 15 | openspace |
| | | | High Performance WAN Networking | Olivier Martin | 15 | |
| | | | Discussion | Those above | 30 | |
| 15:30 -15:45 | IT | Coffee Break | | | 15 | openspace |
| 15:45 -16:30 | N Hill | Concluding Session | | From IT Dpt John Ferguson François Fluckiger François Grey Wolfgang von Rüden Other Dpt tbc | 45 | openspace |
| 16:30 | N Hill | Departure to airport | | | | |

# Appendix D

## LIST OF ILLUSTRATIONS

# Appendix E

## GLOSSARY OF TERMS

| | |
|---|---|
| µs | microsecond |
| 24/7 | 24 hours a day, 7 days a week |
| 2D | two dimensional |
| 3D | three dimensional |
| ALICE | A Large Ion Collision Experiment |
| AliEn | ALICE Environment (Grid framework) |
| ATLAS | A Toroidal LHC ApparatuS |
| BDP | bandwidth delay product |
| CASTOR | CERN Advanced STORage management system |
| CDS | CERN Document Server Software |
| CERN | The European particle physics laboratory (Switzerland) |
| CHF | Swiss franc (£1 = CHF 2.26 at time of publication) |
| CMS | Compact Muon Solenoid |
| COTS | commercial-off-the-shelf |
| CPU | central processing unit |
| CRISTAL | Cooperating Repositories and Information Systems for Tracking Assembly Lifecycles |
| DAQ | data acquisition |
| DAS | direct attached storage |
| DTI | Department of Trade and Industry (UK) |
| DWDM | dense wave division multiplex |
| EC | European Commission |
| EDG | European Data Grid |
| EDH | Electronic Document Handling (system at CERN) |
| EGEE | Enabling Grids for E-science in Europe |
| ESA | European Space Agency |
| ESD | event summary data |
| ETT | Education and Technology Transfer (department at CERN) |
| EU | European Union |
| FEE | front-end electronics |
| FPGA | field programmable gate array |
| GB | gigabyte (1 x $10^3$ MB) |
| Gbps | gigabit per second |
| GDC | global data collector |
| GÉANT | CERN's multi-gigabit pan-European data communications network |
| GPL | General Public Licence (www.gnu.org/copyleft/gpl.html) – a standard OS software licence |
| GUI | graphical user interface |
| HBA | host bus adaptor |
| HEP | high-energy physics |
| HPCN | high-performance computer networking |
| IP | (1) intellectual property; (2) internet protocol |
| IPR | intellectual property rights |
| IT | information technology |
| ITP | International Technology Promoter (DTI) |
| kB | kilobyte |
| krad | kilorad |
| KVM | keyboard, video, mouse |
| LCG | LHC Computing Grid |

| | |
|---|---|
| LDC | local data collector |
| LHC | Large Hadron Collider |
| LHCb | LHC beauty experiment |
| MB | megabyte |
| MB-NG | Managed Bandwidth – Next Generation (collaborative project) |
| MHz | megahertz |
| MPLS | multi-protocol label switching |
| NAS | network attached storage |
| NeSC | National e-Science Centre (UK) |
| OGSI | Open Grid Service Infrastructure |
| OS | open source |
| OST | Office of Science and Technology (UK) |
| PB | petabyte (1 x $10^9$ MB) |
| PC | personal computer |
| PDS | permanent data storage (on magnetic tape) |
| PDU | power distribution unit |
| PET | positron emission tomography |
| PIPSS | PPARC Industrial Programme Support Scheme |
| POOL | Pool Of persistent Objects for the LHC |
| PPARC | Particle Physics and Astronomy Research Council (UK) |
| QGP | quark gluon plasma |
| QoS | quality of service |
| RAL | Rutherford Appleton Laboratory (UK) |
| RDBMS | relational database management system |
| ROOT | Object Oriented Data Analysis Framework |
| RORC | read-out-receive card |
| s | second |
| SAN | storage attached network |
| SLC3 | Scientific Linux CERN 3 |
| SME | small or medium enterprise |
| SQUID | superconducting quantum interference device |
| TB | terabyte (1 x $10^6$ MB) |
| TCP | transmission control protocol |
| TDS | transient data storage (on disk) |
| TeV | tera-electronvolt |
| TT | technology transfer |
| UK | United Kingdom |
| UN | United Nations |
| US(A) | United States (of America) |
| VDT | Virtual Data Toolkit |
| VPN | virtual private network(ing) |
| WS-RF | Web Services Resource Framework |

# Appendix F

## KEY CONTACT POINTS

**Qi3 - Industry Coordinator for PPARC and UK Technology Transfer Coordinator for CERN**
St Johns Innovation Centre
Cowley Road
Cambridge CB4 0WS
United Kingdom
Contact: Nathan Hill
T +44 (0)1223 422 405
nathan.hill@qi3.co.uk

**UK/CERN Industrial Liaison Office**
Contact: Ian Stagg
UK/CERN Industry Coordinator
T +44 (0)1208 851 581
ian.stag@pparc.ac.uk

**Particle Physics & Astronomy Research Council (PPARC)**
Polaris House
North Star Avenue
Swindon
Wiltshire SN2 1SZ
United Kingdom
Contact: Susan Lansdowne
Business & Partnership Schemes Manager
T +44 (0)1793 442 056
F +44 (0)1793 442 036
susan.lansdowne@pparc.ac.uk

**CERN Technology Transfer Unit**
T +41 22 767 8544
industry.liaison@cern.ch

**BEAMA**
Westminster House
3 Albert Embankment
London SE1 7SL
United Kingdom
Contact: Jan Fillingham, Head of Exhibitions
T +44 (0)20 7793 3024/3025
F +44 (0)20 7793 3054

**British Consulate-General**
Trade Development Section
37 – 39 Rue de Vermont
1211 Geneva 20
Switzerland
Contact: Eleanor Baha
T +41 22 918 2421
F +41 22 918 2322
eleanor.baha@fco.gov.uk

# Appendix G

## MAJOR RELEVANT CERN WEBLINKS

| | | |
|---|---|---|
| About this visit | Event description page | http://cern.ch/it-dep-tt/Events/events-list.asp |
| | Agenda and Presentations | http://agenda.cern.ch/fullAgenda.php?ida=a042909 |
| IT Department Information | IT Home page | http://cern.ch/it-dep |
| | Technology Transfer in IT site | http://cern.ch/it-tt |
| | LCG Project Site | http://cern.ch/LCG/ |
| | EGEE Project Site | http://cern.ch/egee-intranet/gateway.html |
| | Openlab Project Site (Management) | http://cern.ch/openlab-mu-internal/ |
| PH Department Information | PH Home page | http://cern.ch/ph-dep/ |
| | Software Group Site | http://cern.ch/ph-sft/ |
| | Geant4 Site | http://cern.ch/wwwasd/geant4/geant4.html |
| | Root | http://root.cern.ch/ |
| | Alice Experiment | http://cern.ch/Alice/AliceNew/ |
| Secretary General/ ETT Information | SG Home page | http://cern.ch/dsu/ |
| | CERN Technology Transfer Home Page | http://cern.ch/ett-div/TT/zTT.html |

# Appendix H

## ACKNOWLEDGMENTS

The mission team would like to thank:

- DTI Global Watch Service
- British Consulate-General Geneva
- All representatives of CERN as listed in Appendix B, in particular:
  – François Fluckiger
  – Jean-Marie Le Goff
  – Wolfgang von Rüden

The DTI's Global Watch Service provides support dedicated to helping UK businesses improve their competitiveness by identifying and accessing innovative technologies and practices from overseas.

**Global Watch Information**
*Global Watch Online* – a unique internet-enabled service delivering immediate and innovative support to UK companies in the form of fast-breaking worldwide business and technology information. The website provides unique coverage of DTI, European and international research plus business initiatives, collaborative programmes and funding sources.
*Visit: www.globalwatchonline.com*

*Global Watch magazine* – the website's sister publication, featuring innovation in action. Distributed free to over 16,000 UK organisations, this monthly magazine features the latest technology developments and practices gleaned from Global Watch Service activities around the world now being put into practice for profit by British businesses.
*Contact: subscriptions@globalwatchonline.com*

*UKWatch magazine* – a quarterly magazine, published jointly by science and technology groups of the UK Government. Highlighting UK innovation and promoting inward investment opportunities into the UK, the publication is available free of charge to UK and overseas subscribers.
*Contact: subscriptions@ukwatchonline.com*

**Global Watch Missions** – enabling teams of UK experts to investigate innovation and its implementation at first hand. The technology focused missions allow UK sectors and individual organisations to gain international insights to guide their own strategies for success.
*Contact: missions@globalwatchonline.com*

**Global Watch Secondments** – helping small and medium sized companies to send employees abroad or receive key people from another country. Secondments are an effective way of acquiring the knowledge, skills, technology and connections essential to developing a business strategically.
*Contact: secondments@globalwatchonline.com*

**Global Watch Technology Partnering** – providing free, flexible and direct assistance from international technology specialists to raise awareness of, and provide access to, technology and collaborative opportunities overseas. Delivered to UK companies by a network of 16 International Technology Promoters, with some 6,000 current contacts, providing support ranging from information and referrals to more in-depth assistance with licensing arrangements and technology transfer.
*Contact: itp@globalwatchonline.com*

For further information on the Global Watch Service please visit *www.globalwatchonline.com*

Printed in the UK on recycled paper with 75% de-inked post-consumer waste content

First published in September 2004 by Pera Innovation Limited on behalf of the Department of Trade and Industry



URN 04/1422