WLCG Worldwide LHC Computing Grid

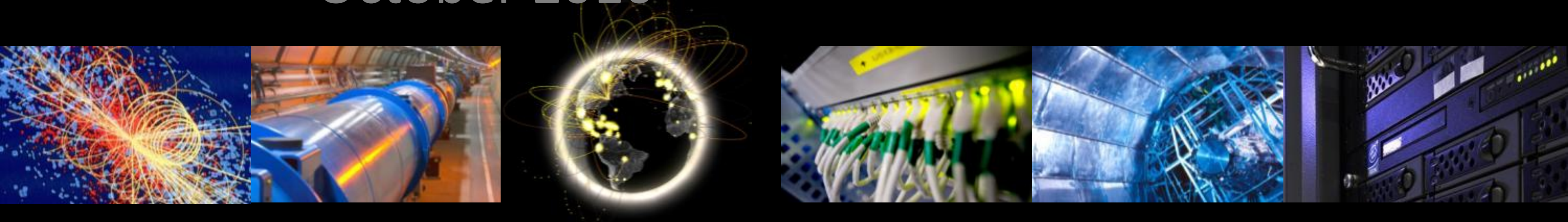
ROC LA Regional Operation Center Latin America

From Service Challenges to First Physics

Jamie Shiers

I ROC-LA Workshop

October 2010



Introduction

- What follows is a **personal** view of the challenges and successes in setting up today's worldwide distribution computing infrastructure for the world's largest scientific machine
- Although it is – historically – a story about grid computing the main messages are about how **collaboration** can lead to success
- It involves many people and sites around the world who have worked together to make something possible – something **great**

Elements of the Story

- Rather than follow the strict timeline implied by the title I propose to pick a few key messages to share with you
- The first is why this **style** – not especially grid – of distributed computing and collaboration makes sense (compared to other models) – including motivation for collaboration
- The second is about **service**: design, implementation, deployment, operation… and usage
- The final one concerns the future – our **future**
- But firstly, to set the scene, a few words about LHC computing…

The 3rd Decade of LHC Computing

- We are now entering the **3rd** decade of LHC Computing which is marked by the successful use of the Worldwide LHC Computing Grid for extended data taking since the restart of the LHC at the end of **March 2010**
- I will mention briefly the first decade – which prepared the ground for the deployment of the grid – focus mainly on the second and outline some challenges for the third

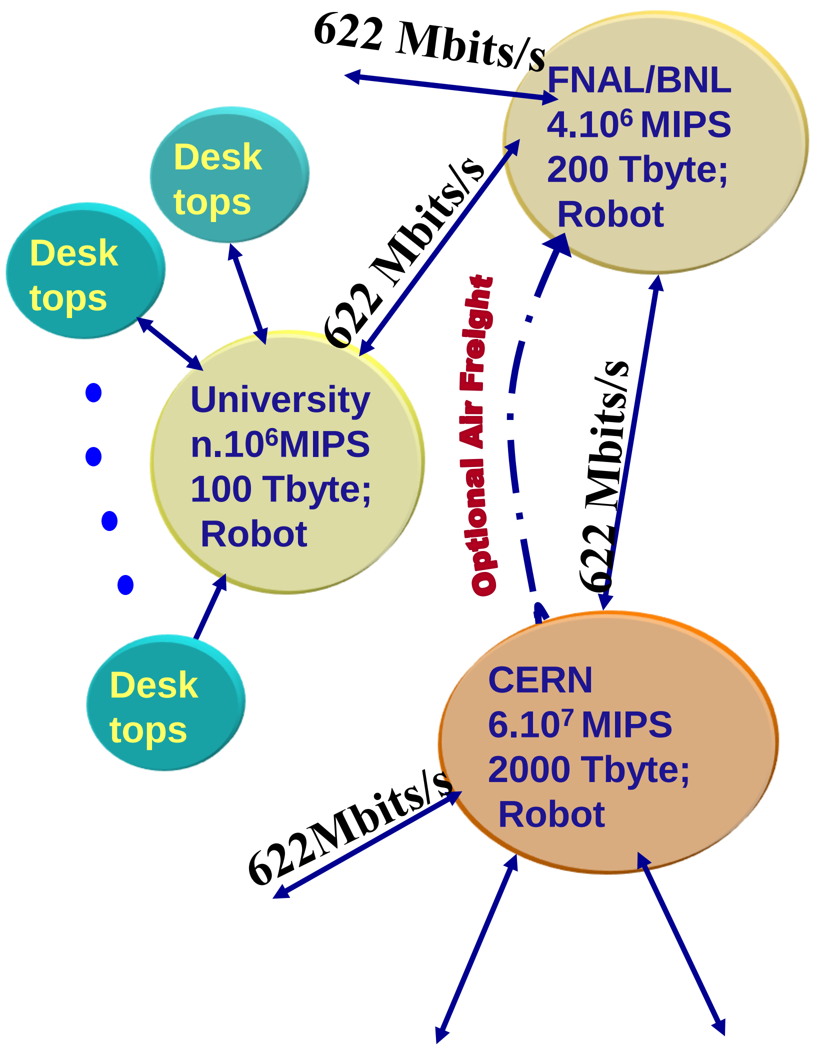
The First Decade

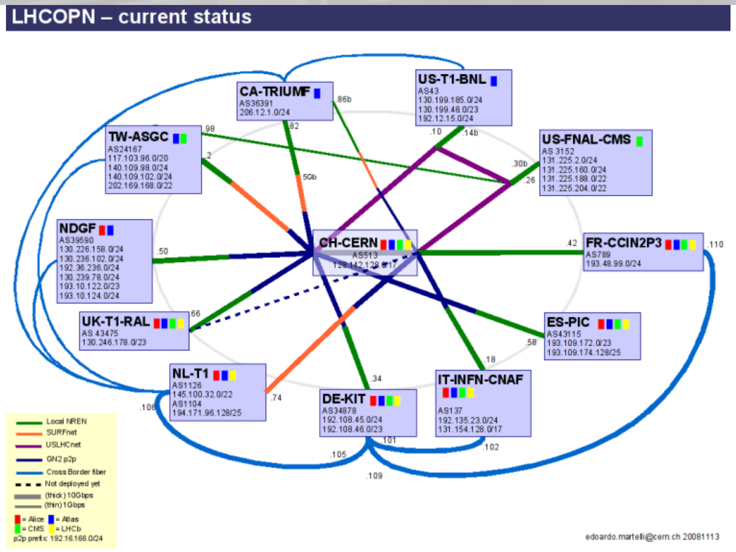
- Started at **CHEP 1992**, Annecy, France, where significant focus was on the challenges of the SSC and LHC – plus increasing focus on "industry standards" versus HEP-specific solutions
- Led to several years of **R&D** – object oriented analysis and design, object oriented languages and databases – and production use towards the end of the decade
- Co-existed with wide-scale LEP exploitation and a revolution in the IT world: Internet explosion, commodity PCs, the **Web**
- It ended with the elaboration of possible models for LHC Computing – the "MONARC proposal"

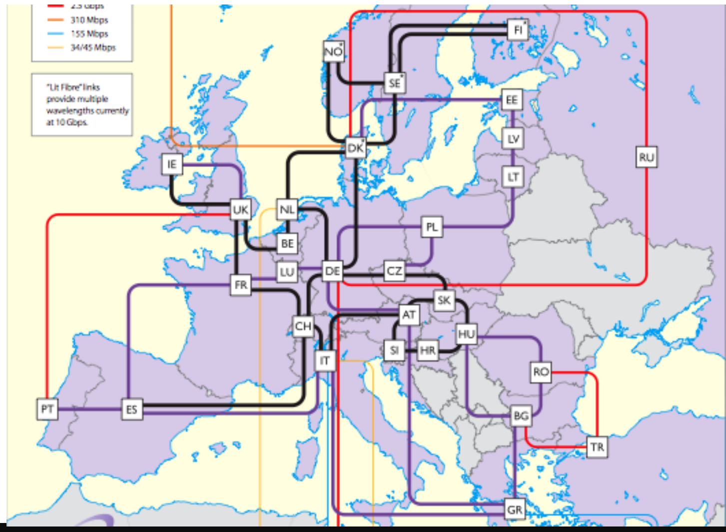
The MONARC model

- The MONARC project tried to define a set of viable models for LHC computing
- It proposed a hierarchical model with a small number of regional centres at the national level plus a larger number of local centres – Universities of Institutes
- This model – consisting of a Tier0, roughly 10 Tiers1 and some 100 Tier2s – is the basis of the today's production environment
- N.B. MONARC foresaw optional airfreight as an alternative to costly and low-bandwidth networking (622Mbps or less…)

The MONARC model vs today







Enter the Grid

- Around the turn of the millennium cracks were beginning to appear in the solutions proposed by the various R&D projects – and adopted at the 100TB-1PB scale by experiments from several labs across the world
  - Major data and software migrations necessary
- At the same time, Ian Foster *et al* were evangelizing a new model for distributed computing
- HEP took the bait: CERN was the lead partner in a series of EU funded projects and (W)LCG was born

The Second Decade

- Several generations of grid R&D and deployment projects: in Europe EDG followed by EGEE I, II and III (**EUR100M** of investment from EU) plus partner projects in other areas of the world
- 1st half of the decade included "**data challenges**" run by the experiments testing components of their computing models and specific services
- 2nd half: a series of "**service challenges**" that contributed to the ramp-up of the global service to be ready well prior to planned data taking
- **Moving targets**: computing models, experiment frameworks and underlying middleware and services all developed concurrently…

The Worldwide LHC Grid (WLCG)

- Simply put, this is the distributed processing and storage system deployed to handle the data from the world's largest scientific machine – the Large Hadron Collider (LHC)
- Based on grid technology – including the former **EGEE** infrastructure in Europe / Asia Pacific & OSG in the US
- WLCG is **more** than simply a customer of EGEE: it has been and continues to be a **driving** force not only in the grid domain but also others, such as storage and data management
- WLCG has always been about a **production** service – one that is needed 24 x 7 most days (362) per year
  - Much activity – particularly at Tier0 and Tier0-Tier1 transfers – takes place at nights and over weekends (accelerator cycle)

The WLCG Deployment Model

- WLCG is the convergence of **grid** technology with a specific deployment model, elaborated in the late 1990s in the "Modelling of Network & Regional Centres" (**MONARC**) project
- This defined the well-known hierarchy Tier0/Tier1/Tier2 that is now common to several disciplines and matches well to International Centre / National Centres / Local Institutes
- MONARC originally foresaw limited networking between Tier0/Tier1/Tier2s – with air freight as a possible backup to (best case) **622Mbps** links (cost!), as well as a smaller number of centres than we have today
  - We have redundant 10Gbps links: T0-T1 & also T1-T1, some of which are on occasion max-ed out!

- These base assumptions are currently being re-discussed

CCP 2008

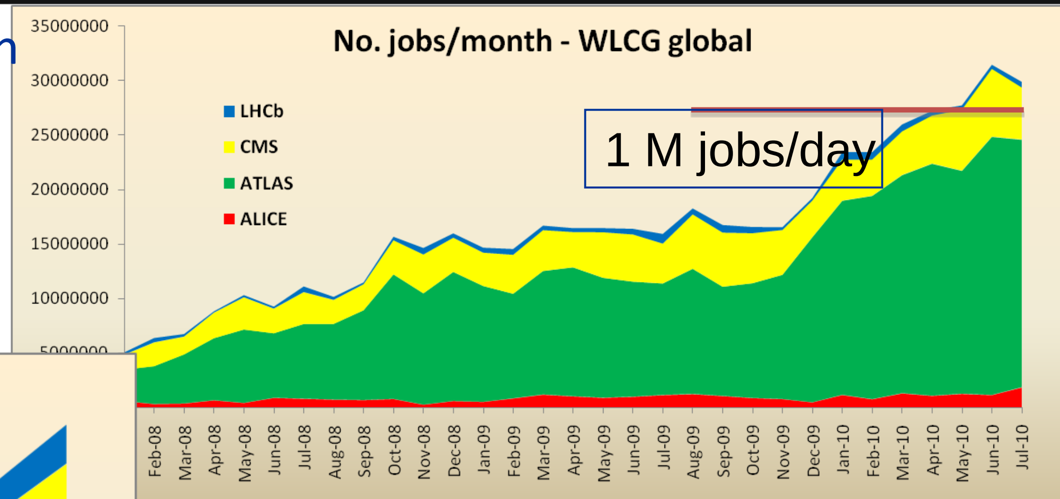
- In 2008 I attended a conference on Computational Physics in Ouro Preto
- I have a talk entitled “Grids Today, Clouds on the Horizon” (hence first picture)
- At that conference there were many other examples of “petascale” computing presented
  - And a definition of what petascale means…
- Together with discussions of the support models & costs

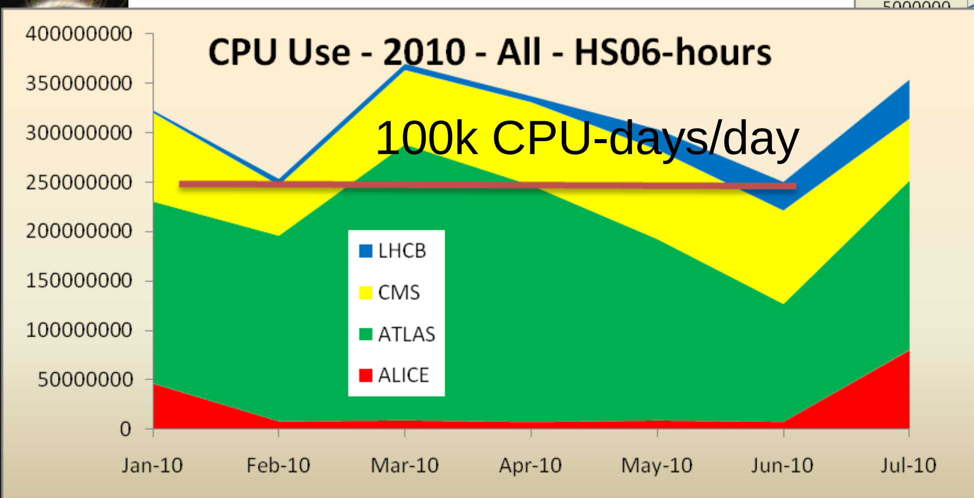
The WLCG Scale

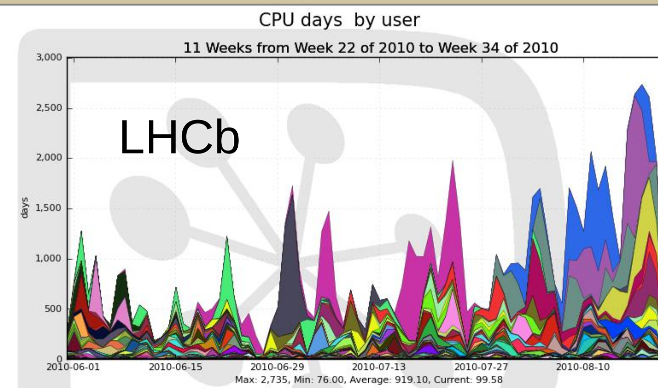
- Deployed worldwide: Americas, Europe & Asia-Pacific
- Computational requirements: **O(10^5)** cores
  - Currently close to **1M** analysis jobs / day & increasing
- Networking requirements: routinely move **1PB** of data per day between grid sites – significant intra-site requirements
  - Single VO transfers CERN-Tier1s >**4GB/s** over sustained periods (~days)
- Annual growth in stored data: **15PB**
  - Old calculation: # copies & location(s) of data may well be revised in coming months as well as trigger rates & event sizes
- Sum of resources at each tier approximately equal
  - 1 Tier0, ~10 Tier1s, ~100 Tier2s
- Sum of issues (tickets) at each tier (service metric) also ~equal!
  - A few: rarely as many as 5 per VO per day (OPS meeting)
- Significant responsibility devolved to the Tier1s & Tier2s

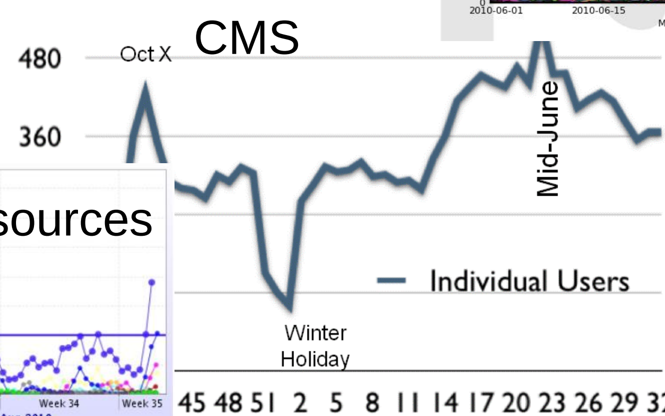
WLCG Usage

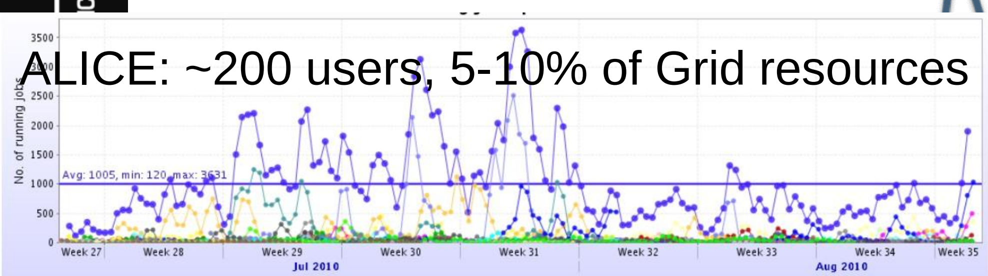
- Use remains consistently high
  - 1 M jobs/day; 100k CPU-days/day











Large numbers of analysis users

- CMS ~500,
- ATLAS ~1000,
- LHCb/ALICE ~200

Alternatives to Grids

- A classic alternative is based on Supercomputers
- Pros:
  - Very nice; very powerful; relatively easy to manage;
- Cons:
  - Application has to be extensively modified (read "re-written") for specific supercomputer;
  - Learning curve very steep;
  - SC lifetime short – cycle repeated often;
  - SCs are a scarce and expensive resource inaccessible to many disciplines / people;
  - None of the "pros" of grid computing apply…

Alternatives to SCs

- A possible alternative is based on grids of commodity clusters / hardware
- Pros:
  - Low(-er) cost of entry;
  - Hardware and support costs spent locally;
  - Scalable and extensible;
  - Very low "common denominator" – for us it is "Linux" – highly beneficial for very long-live projects such as the LHC
- Cons:
  - It ain't so easy to setup; it ain't so easy to use…
- But WLCG is proof of existence at petascale…

Why Spend Locally?

- The ability to spend locally (materials, personnel) is an enabler for (inter)national collaboration
- It avoids "brain-drain" (and money) and helps create / support locally thriving institutes / universities of excellence
- We all know why this is needed: if you don't invest in the future you don't have one…

Recap

- We have seen how distributed computing infrastructures can handle petascale problems at a fraction of the cost than alternatives
- And much more acceptable in terms of direct and indirect benefits…
- But what of "The 3rd law of grid computing" – namely a "non-trivial level of service"…

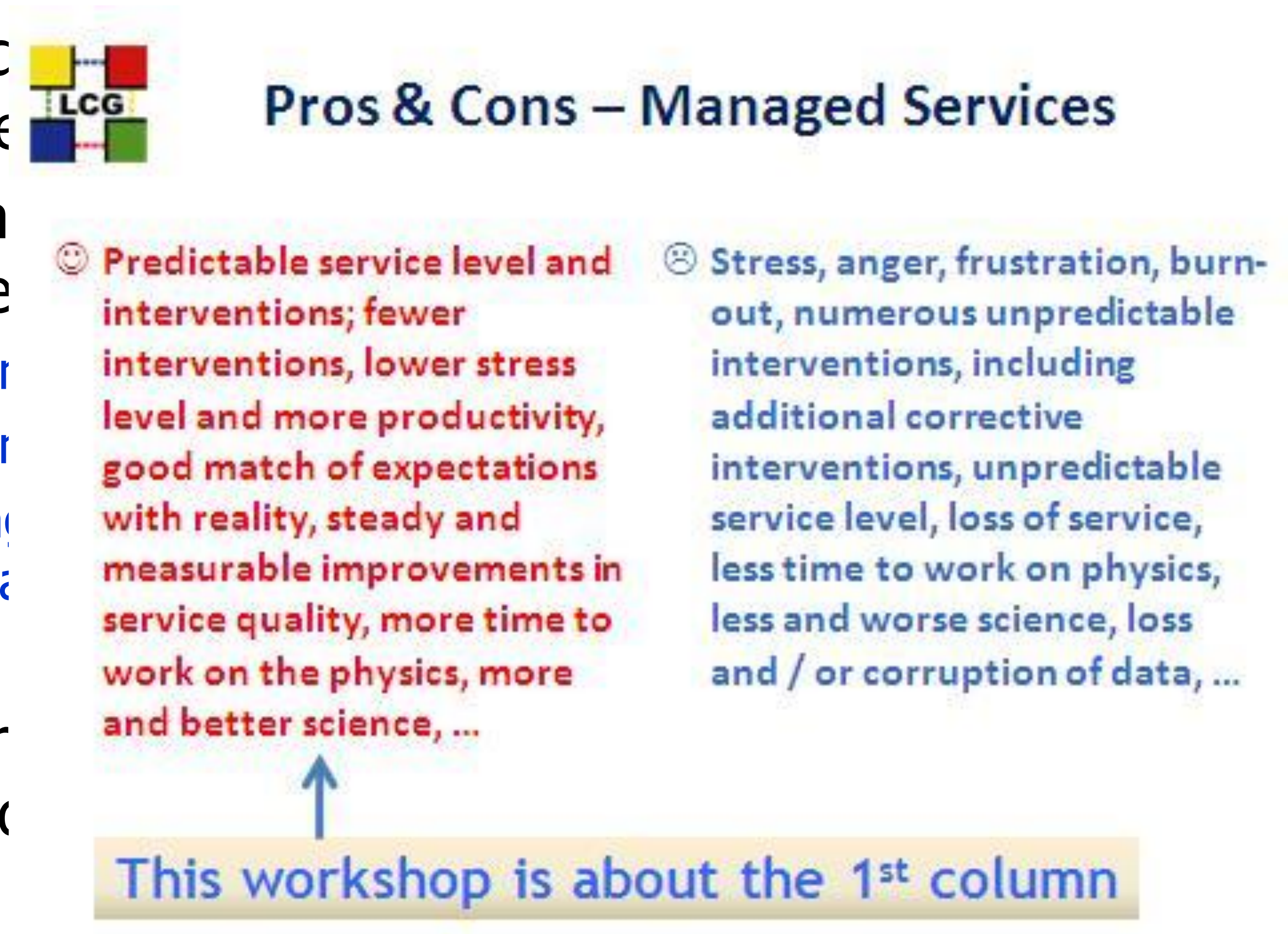
The Service is the Challenge

(not the Service Challenges)

What Does Service Deployment Mean?

Production Services

- In November 2007 we had a workshop focussing on service reliability
- This followec deploy the se
- Much of wha "Common Se
  - Not deployi
  - Not deployi
  - Not countin production a
  - Etc.
- Can we affor providers anc




Fragile!
Please do not touch
Ne touchez pas s.v.p.

WLCG Operational Targets

Time Interval	Issue (Tier0 Services)	Target
30’	Operator response to alarm / call to x5011	99%
1 hour	Operator response to alarm / call to x5011	100%
4 hours	Expert intervention in response to above	95%
8 hours	Problem resolved	90%
24 hours	Problem resolved	99%

Above targets approved by WLCG Overview Board

Time Interval	Tier1 Services	Target
1 working day	All services	95%

Time Interval	Tier2 Services	Target
1 working day	All services	90%

Targets discussed at WLCG Grid Deployment Board

Collaboration – motivation

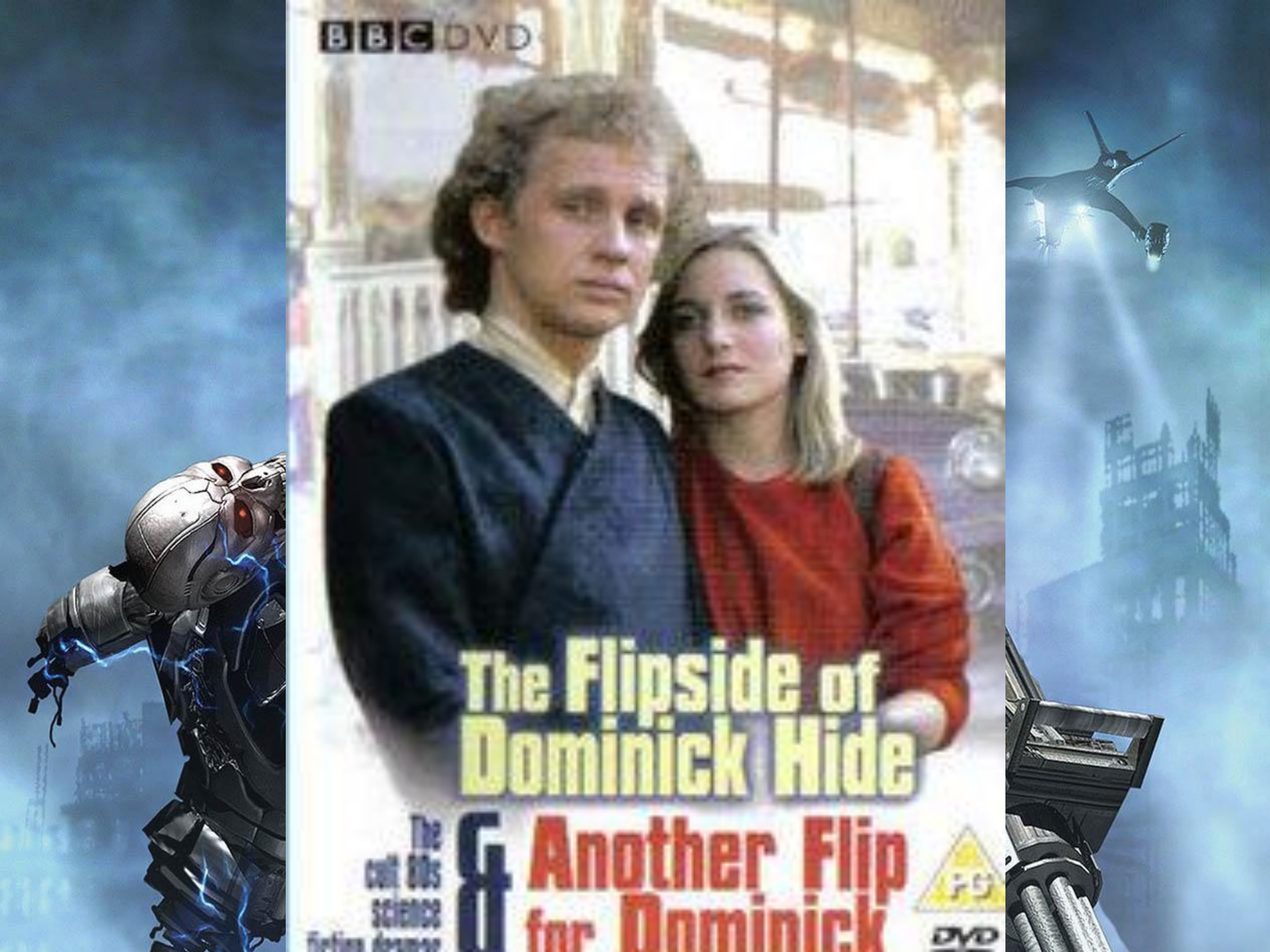
Collaboration Example: CERNLIB

- The CERN Program Library is a large collection of general purpose programs maintained and offered in both source and object code form on the CERN central computers. Most of these programs were developed at CERN and are therefore oriented towards the needs of a physics research laboratory. Nearly all, however, are of a general mathematical or data-handling nature, applicable to a wide range of problems.
- The library contains several thousand subroutines and complete programs which are grouped together by logical affiliation into several hundred program packages. 80% of the programs are written in FORTRAN and the remainder in assembly code, or C usually with a FORTRAN version also available. The language supported is currently Fortran.
- Wikipedia entry

The World Wide Web

- often, abbreviated as **WWW** and commonly known as **The Web**, is a system of interlinked hypertext documents contained on the Internet. With a web browser, one can view web pages that may contain text, images, videos, and other multimedia and navigate between them by using hyperlinks.
- Using concepts from earlier hypertext systems, British engineer and computer scientist Sir Tim Berners Lee, now the Director of the World Wide Web Consortium, wrote a proposal in March 1989 for what would eventually become the World Wide Web.[1] He was later joined by Belgian computer scientist Robert Cailliau while both were working at CERN in Geneva, Switzerland. In 1990, they proposed using "HyperText [...] to link and access information of various kinds as a web of nodes in which the user can browse at will",[2] and released that web in December.[3]
- "The World-Wide Web (W3) was developed to be a pool of human knowledge, which would allow collaborators in remote sites to share their ideas and all aspects of a common project." [4] If two projects are independently created, rather than have a central figure make the changes, the two bodies of information could form into one cohesive piece of work.

Investment in the future?


BBC DVD
The Flipside of Dominick Hide
The
cult 80s
science
&
Another Flip
for Dominick
PG
DVD

- If you cut investment in the future – e.g. in Research and Development – it is **exactly the same** as a Terminator coming back.

- **Exactly.**

➢ **Funding that encourages collaboration and collaborations that has/have a long term benefit to science and society typically reaps benefits orders of magnitude larger than the initial investment(s).**

Grid computing – benefits

- Grid computing has been shown to efficiently support worldwide virtual organisations
- It allows local investment and the return on that investment is also seen locally – a big plus for international science
- The costs – including support – are very modest compared with alternative “peta-scale” solutions
- Also matches well the timescale of such projects – typically >> lifetime of SuperComputer generations
- LHC measured in decades… (or careers…)

LHCC Referees…

- *"First experience of the World-wide LHC Computing Grid (WLCG) with the LHC has been positive. This is very much due to the substantial effort invested over several years during the intensive testing phase and all Tier centres must take credit for this success. The LHCC congratulates the WLCG on the achievements."*

May 2010

The Economist, July 29th 2010

- *"More striking still is the speed with which the raw data are being processed. The freshest batch emerged from the LHC on July 18th and were moulded into meaningful results by July 21st, in time for the Paris conference. Not long ago this process would have taken weeks, says Fabiola Gianotti, the spokeswoman for ATLAS, one of the four main LHC experiments.*
- *One reason is the development of the Grid, a computing network CERN hopes will prove a worthy successor to its previous invention, the World Wide Web. The Grid lets centres around the world crunch the numbers as soon as they come out of the machine."*

Summary

- WLCG delivers a **production** service at the Terra & peta scale, it permits local investment and exploitation, solving the "brain-drain" and related problems of previous solutions
- It has taken **longer** & been a lot **harder** than foreseen
- Many of the basic ingredients work well and are applicable beyond grids; **complexity** has not always been justified
- Expect some **changes**, e.g. in Data Management + a more network-centric deployment model, in coming years: these changes need to be adiabatic