

Synchrotron Grid Workshop

On behalf of the ESRF :

benvingut

bienvenue

bem-vindo benvenuto

velkommen

welcome καλώς ὄρισεσ

välkommen tervetuloa

willkommen

Why the GRID ?

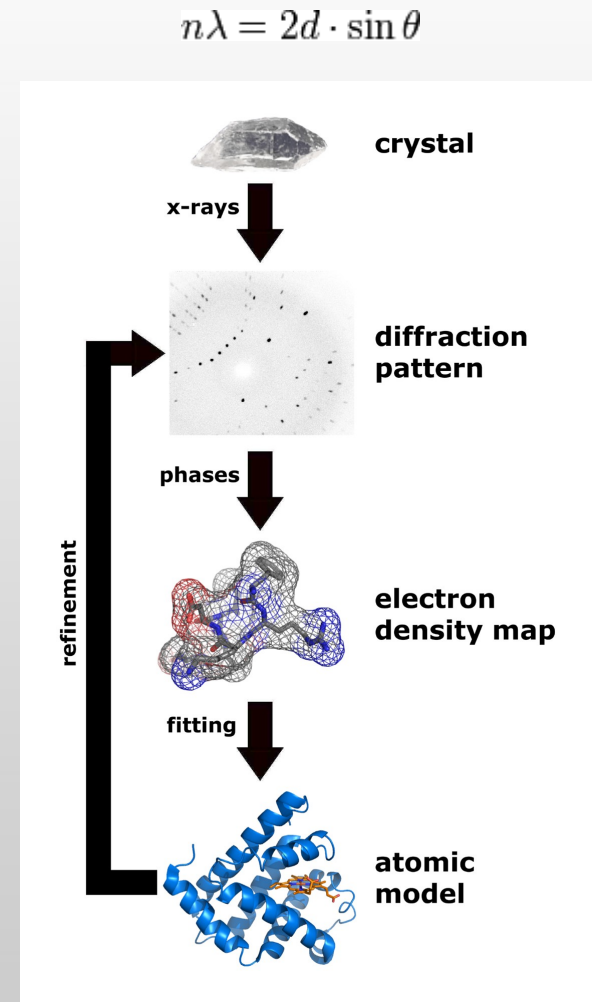
- GRID community has a HUGE experience in distributed computing it would be CRAZY to ignore it !
- GRID provides distributed solutions for :
 - Security
 - Managing sites
 - Submitting jobs
 - Managing data
 - Networking (computer and human)



Synchrotron Science and the GRID

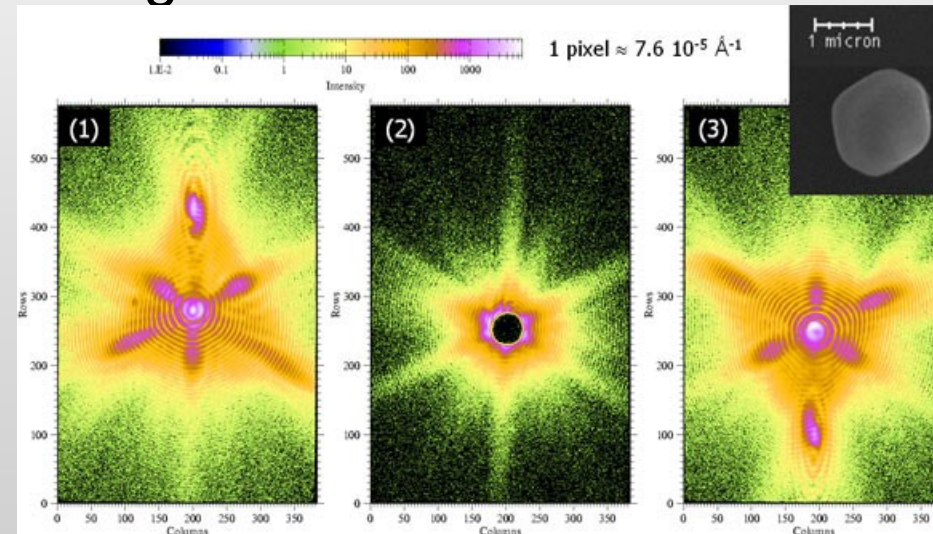
• **A BAD MATCH ?**

- Experiments are seldom simulated
- Experiments are quick and repeatable
- Data analysis is not compute bound
- Many programs are interactive
- Lots of small tasks with lots of small file I/O
- Data sets can be transported on cheap media
- No synchrotron scientists communities



Synchrotron Science is Evolving

- Data sets are growing exponentially in size :
 - 3D volumes are the future in diffraction, fluorescence, ...
 - This means Tera bytes of data and tens thousands of images
- Experimental techniques are evolving :
 - Coherent diffraction is the future :
 - Data analysis is becoming more compute intensive
- User communities are growing
 - New users need help analysing their data
- Data sets need to be archived and accessible
 - The future trend is to archive data



Beginner's Mind

Zen Mind

- *Approaching the GRID with a beginner's mind while staying zen ...*



My Checklist of a Good Software

1. It should just work - we should not have to talk about it
2. Should be simple to install and use
3. Should be easy to program
4. Should be useful

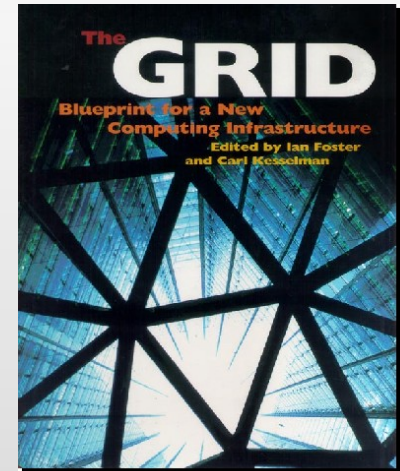
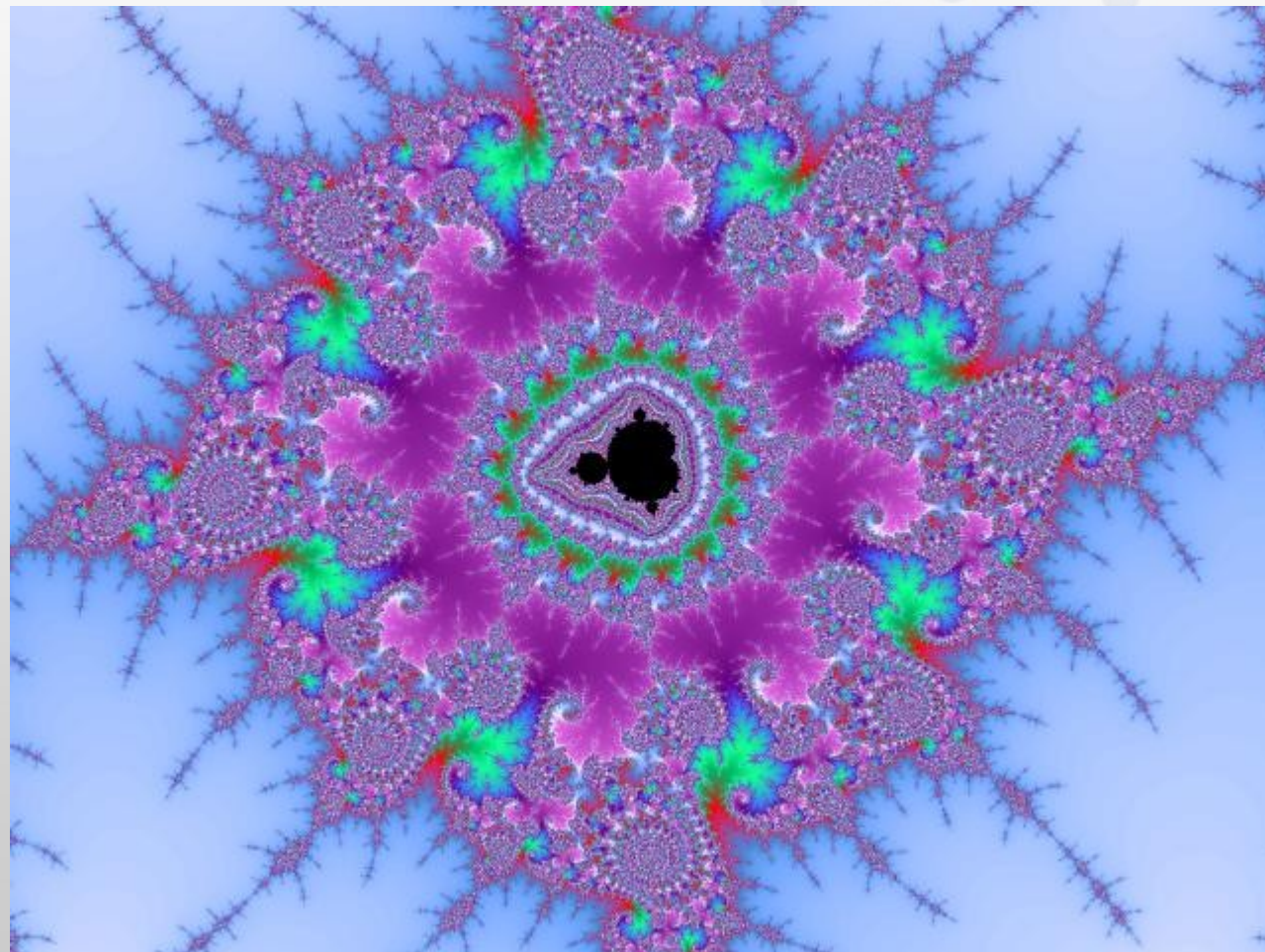
• After 3 months of GRID my feeling is :

- GRID works - we submit jobs to DESY
- GRID is not simple nor easy to install
- We do not know if it is useful for us yet



Ian Foster's + Carl Kesselman (in 1998)

- *“Grid computing is like a fractal extension of local computing”*



Bible

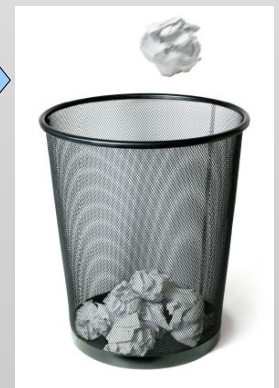
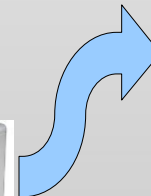


Middleware + Status - Session

- Middleware is the software provided by the GRID to run everything
- We plan to evaluate :
 - **gLite** and **globus** (for managing the grid)
 - **dCache** and **DPM** (for disk storage)
 - **Torque** and **Condor** (for batch jobs)
- Should we be looking at others e.g. ?
 - **Swift** (designed to submit millions of jobs fast)
 - **Taverna** (workflow tool)
 - ... you tell us ...

Data Session

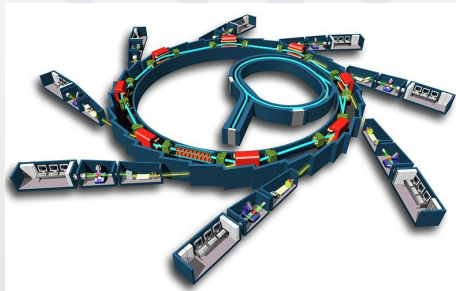
- Data is our raw product – we produce Tera bytes of it (Peta bytes soon)
- Most of our users carry their data around on removable hard disks or delete them
- How can we use the GRID to better manage our data ?



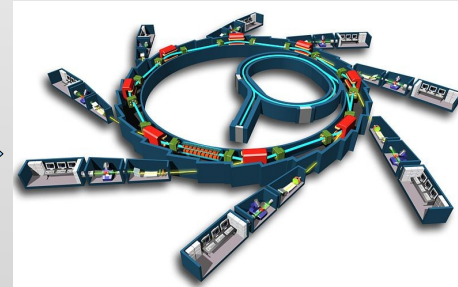
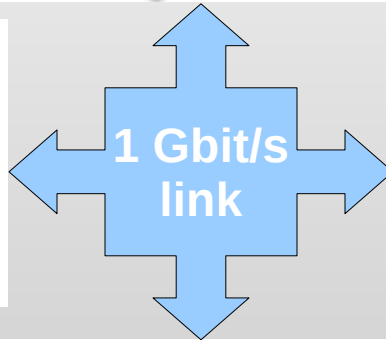
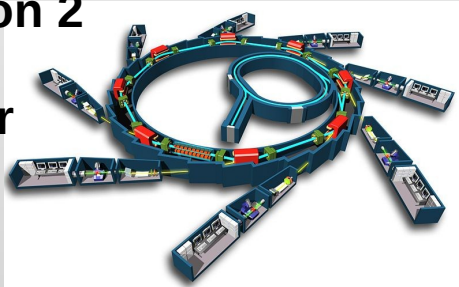
Data Grid – Tier 0 and Tier 1 data centers

- Proposal - replicate data from synchrotrons at other synchrotrons in the Synchrotron Data Grid :***

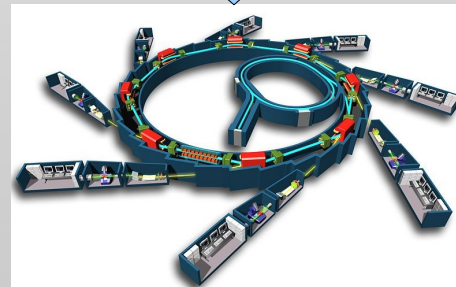
**Synchrotron 1
T0 and T1
data center**



**Synchrotron 2
T0 and T1
data center**



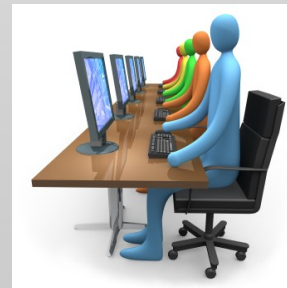
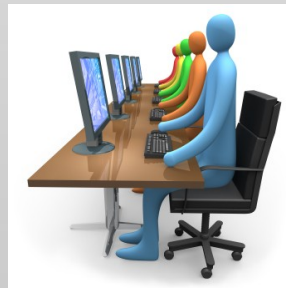
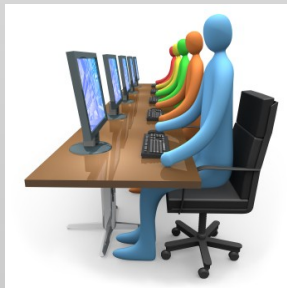
**Synchrotron 4
T0 and T1
data center**



**Synchrotron 3
T0 and T1
data center**

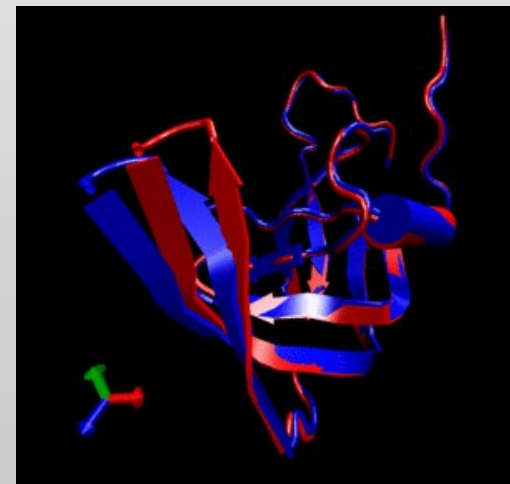
User Management Session

- Synchrotrons have on average thousands of users every year
- They need to be registered in the user database and have access to the computer resources
- GRID has setup certificate-based authentication system
- How can the GRID help manage users ?



Applications Session

- Synchrotron Science data analysis can be described by “**Many Task Computing**” i.e. multiple distinct activities, coupled via (for example) file system operations or message passing
- We plan to test the main classes of applications e.g. file I/O-intensive, computation-intensive, data-intensive
- Which applications are best suited testing the GRID for Synchrotron Science ?
 - Protein crystallography e.g. elnemo, molrep, phaser, ccp4 suite
 - 3D diffraction e.g. grain mapping
 - Tomography



Use Cases

- Use Cases are what we need to define most urgently - we need your help !
- Some Use Cases :
 - **Case 1** : *Export* a data set of 1 Tera byte to user's home institute
 - **Case 2** : Submit typical image *calibration* jobs on many files
 - **Case 3** : Submit typical *processing-bound* jobs on many files
 - **Case 4** : Submit typical *data-bound* jobs on many files
 - **Case 5** : Submit typical *simulation* jobs

Why this workshop is important for us ?

- The ESRF Upgrade Program concerns the next 10 years
- WP11 is the evaluation of the GRID during the preparation of the Upgrade Program. The outcome will determine how to integrate GRID in the ESRF computing landscape or not
- This workshop will help us choose the right direction during the evaluation
- To GRID or not to GRID ...



thank you merci beaucoup



Program – Monday 8 December

- 14:00 **Welcome** - Andy Götz (ESRF)

- **Session 1 - Middleware + Status**
 - 14:15 *Introduction to Grid and Grid @ DESY* - Andreas Gellrich (DESY)
 - 14:45 *gLite middleware* - Ariel Garcia (FZK)
 - 15:15 *globus toolkit + Grid for Synchrotrons* - Gregor Laszewski (Rochester Institute of Technology)
 - 15:45 **coffee**
 - 16:05 *Overview of EGEE* - Rolf Rumler (IN2P3)
 - 16:25 *Grid @ ESRF* - Clemens Koerdt + Fernando Calvelo (ESRF)
 - 16:45 *Grid @ APS* - Dan Fraser (ANL)
 - 17:15 *Grid @ CMS + PSI* - Derek Feichtinger (PSI)
 - 17:35 *Grid @ ELETTRA* - Alessio Curri + George Kourousias (ELETTRA)

- **Discussion - Middleware + Status**

- 19:30 **Dinner at the Restaurant**  [Fantin Latour](#)

Program – Tuesday 9 December

■ Session 2 - User Management

- 08:30 *User Management Requirements* - Dominique Porte (ESRF)
- 08:45 *Single-sign-on and ShibGrid* - Bill Pulford (DLS)

■ 09:00 Discussion - User Management

■ Session 3 - Data Grid

- 09:30 *The Data Grid* - Jos von Wezel (FZK)
- 10:00 *ESRF Data Requirements* - Rudolf Dimper (ESRF)
- 10:30 **coffee**
- 10:50 *eScience Data Grid for DLS and ISIS* - Brian Mathews (STFC)

■ 11:10 Discussion - Data Grid

■ 12:00 Lunch at the ESRF guesthouse

■ Session 4 - Applications

- 13:30 *g-Eclipse - a Grid Graphical User Interface* - Harald Kornmayer (NEC Laboratories Europe)
- 14:00 *3DXRD Grain Mapping* - Carsten Gundlach (ESRF)
- 14:20 *Protein Crystallography Needs* - Frank Schluenzen (DESY)
- 14:40 *Computational Challenges at EMBL-Grenoble* - Sandor Brockhauser (EMBL)
- 15:00 **coffee**
- 15:20 *Tomography data processing and management* - Peter Cloetens (ESRF)
- 15:40 *Simulation processing requirements* - Alessandro Mirone (ESRF)

■ Discussion - Applications