Accessing and Exploiting Solar Dynamics Observatory (SDO) Data in Europe

Veronique Delouille

SWWT plenary meeting, 28 June 2011

Introduction

- As from NASA policy:
  - Data from SDO mission are freely available
- Problem:
  - SDO delivers between 1000 and 10000 times more data than previous missions such as SOHO, TRACE, or STEREO

- Two limitations users run up against:
  - physically accessing the data
  - processing (exploiting) the data

- Conclusion: Some effort is needed to make the data 'freely' available in practice!
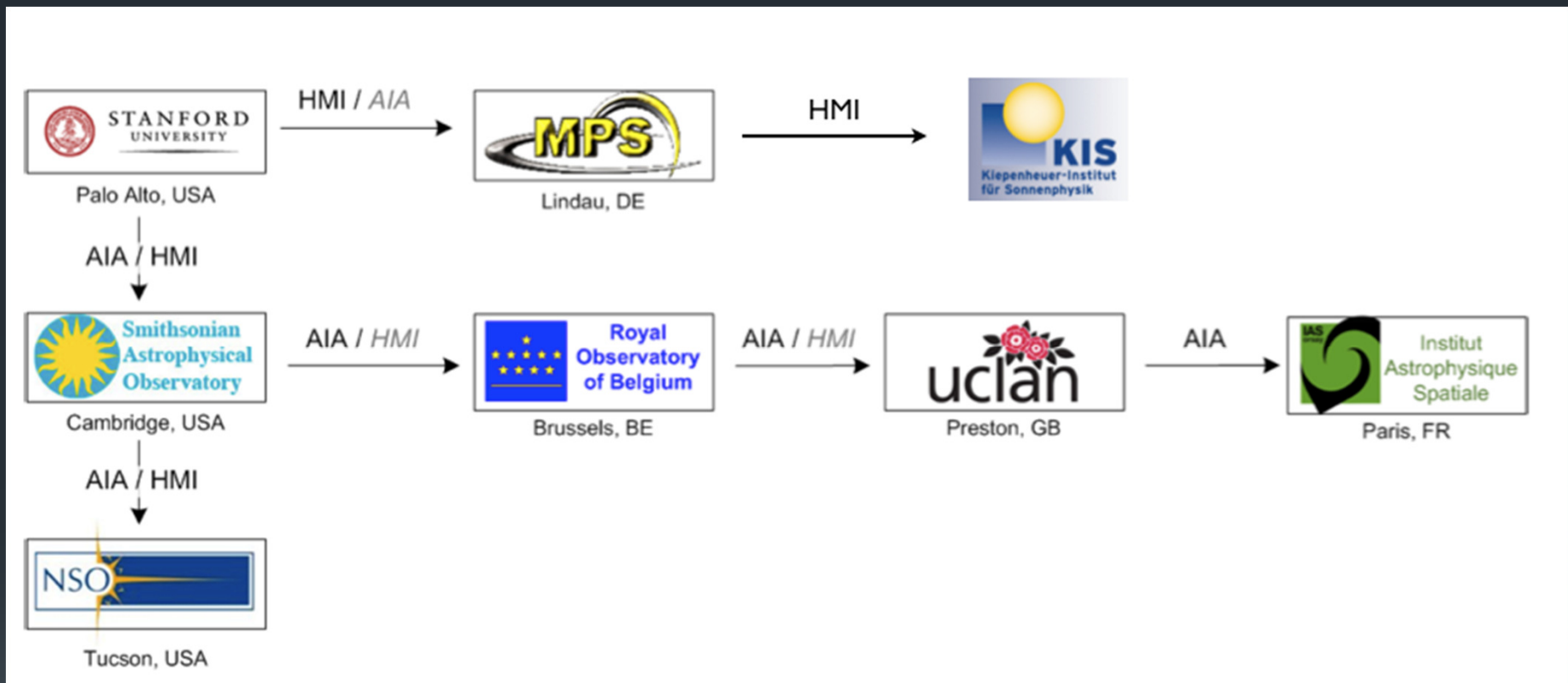
AIA

… 2TB/day
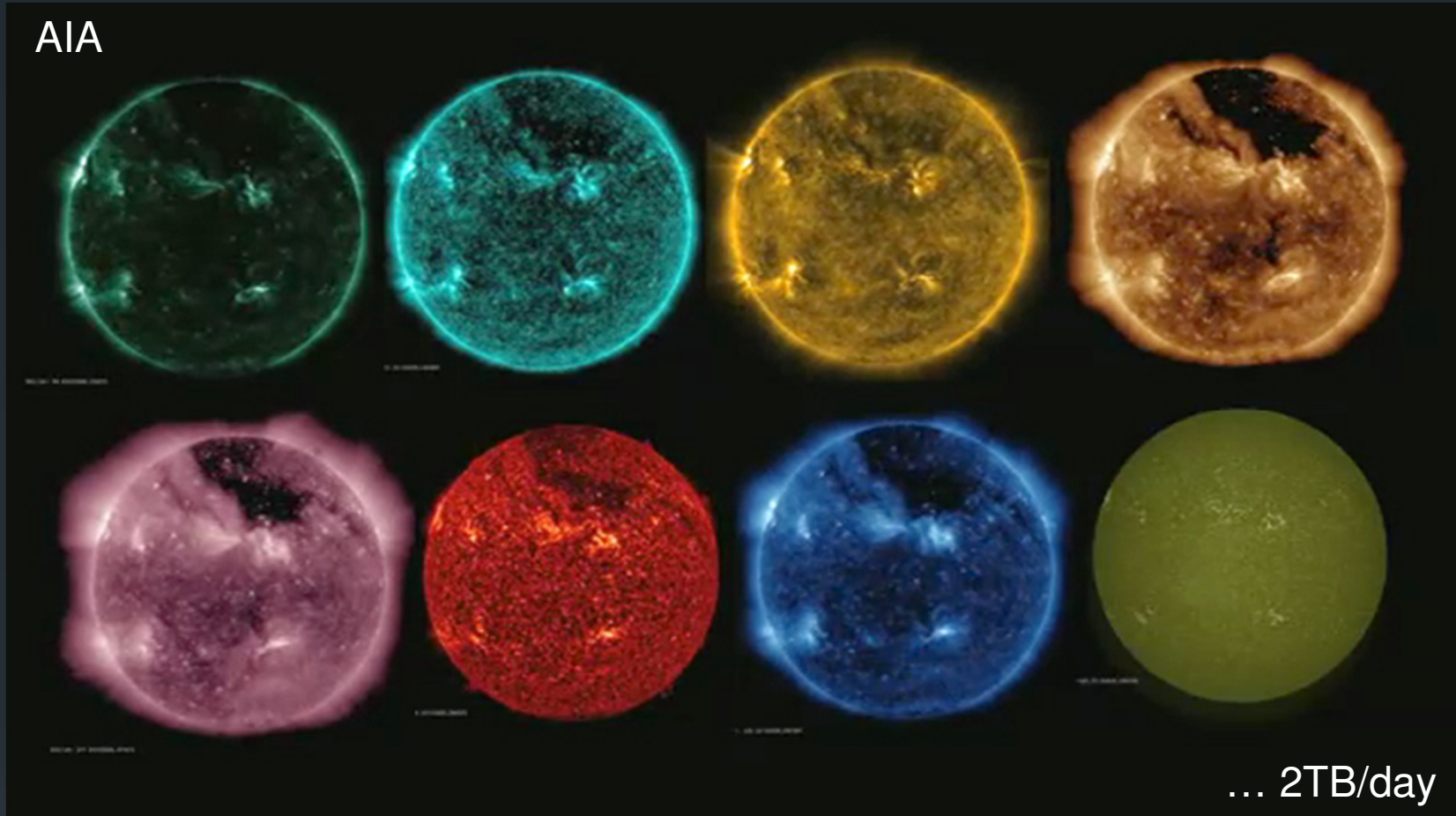
Accessing SDO data

HMI (Helioseismic and Magnetic Imager): 4k x 4k images of LOS magnetogram, intensitygram, dopplergram, and vector magnetogram every 45 sec
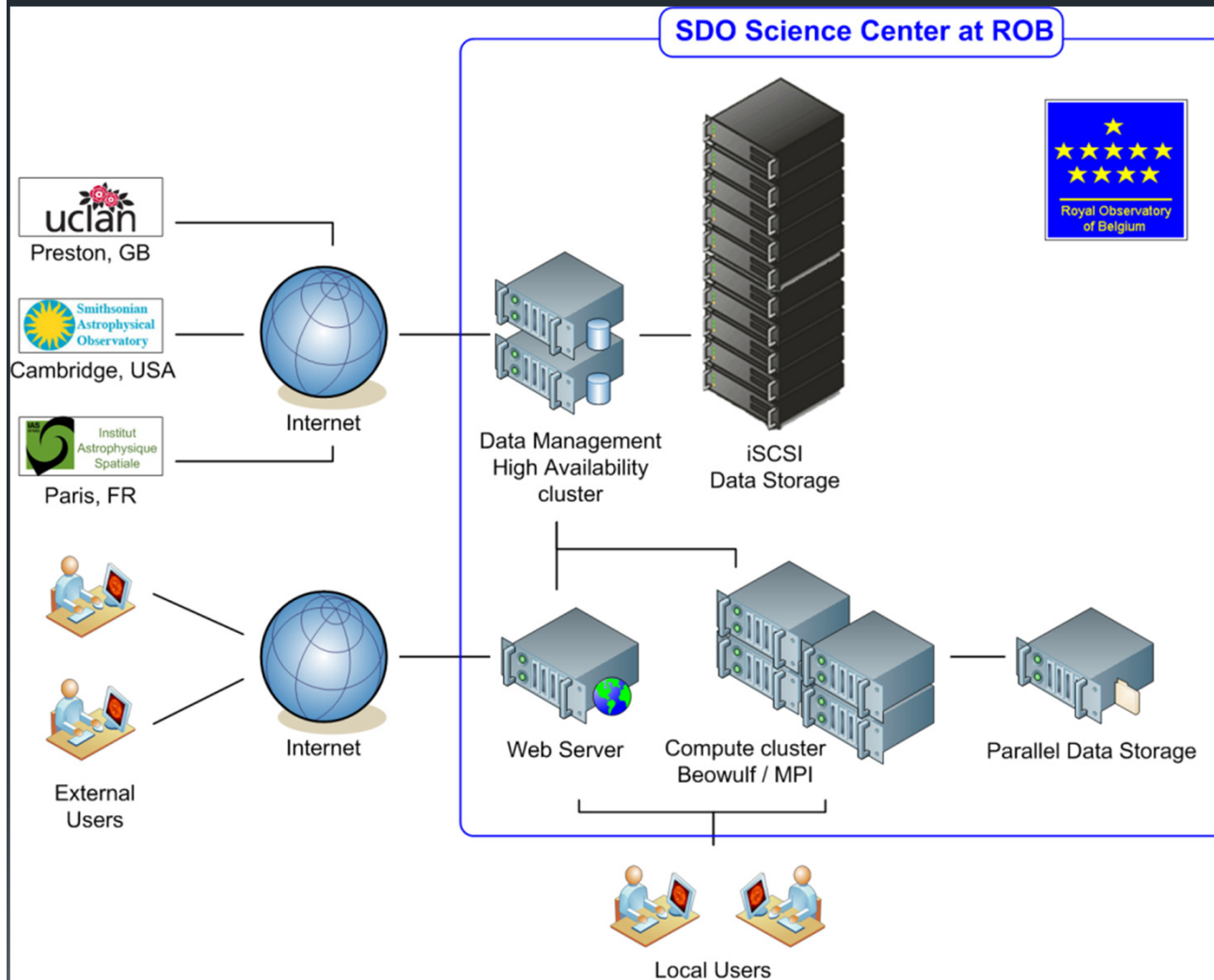
AIA (Atmospheric Imaging Assembly): 4k x 4k images in 10 wavelengths every 10s.

A series of relay SDO data centre

- JSOC alone cannot provide access for the very many streams of data !
- In 2009, ROB demonstrated feasibility of sending 2TB of data per day over internet (300Mb/sec)
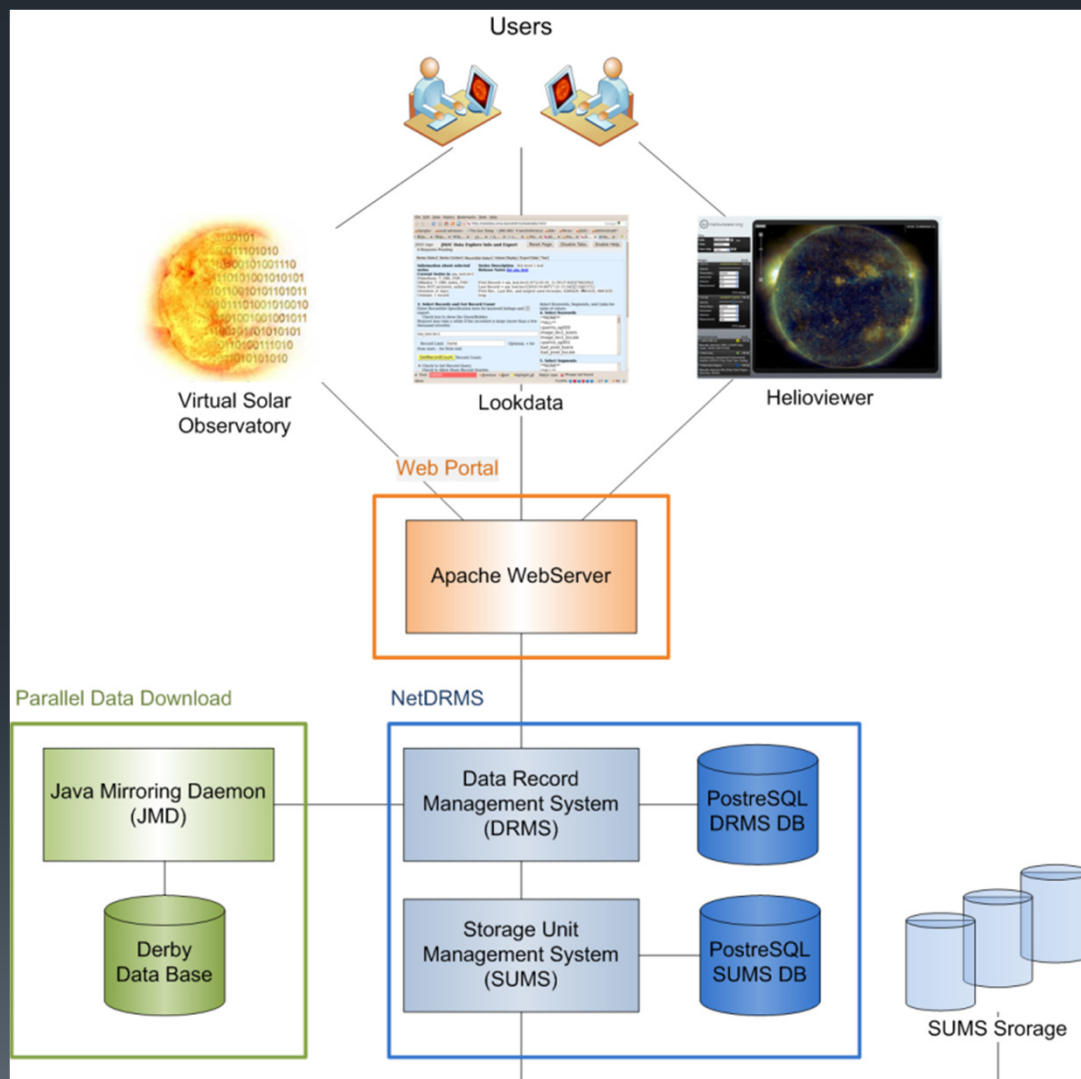
What does a data centre look like?

In graphical …

And practical term !

From a data centre to end-users

- NetDRMS responsible for managing and serving the data locally.
- The JMD downloads the data from other institutes.
- The web portal allows to retrieve the data thanks to multiple interfaces :
  - VSO
  - Python,
  - Pseudo-file system
  - lookdata
 - Local copy of helioviewer data

From a data centre to end-users

- AIA and HMI are arranged into 'series' (measurements with common characteristics).
- AIA is provided in 4k x 4k and 1k x 1k format, with different delays for availability:
  - 1Kx1K copied from JSOC in maybe an hour
  - 4Kx4K science data after several days

- A variety of tools were developed at ROB, following the user's requirements:
  - Simple, (but slow) access via VSO: needs IDL with solar soft up to date
  - Get a directory with fits files: Pseudo file system
  - Access via python
  - Very fast (but complex) access via NetDRMS

http://wisSDOm.oma.be

Example: get data with VSO

- From within IDL: vso_search : to query metadata
- vso_get with keyword site='rob' or site='uclan' to get data from remote sites
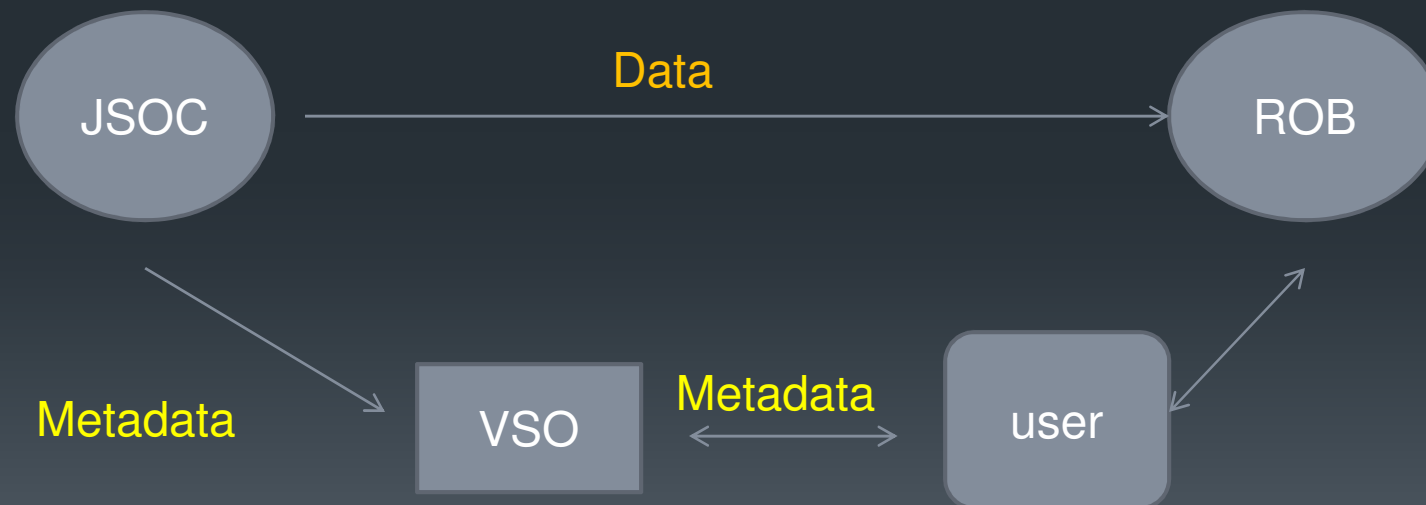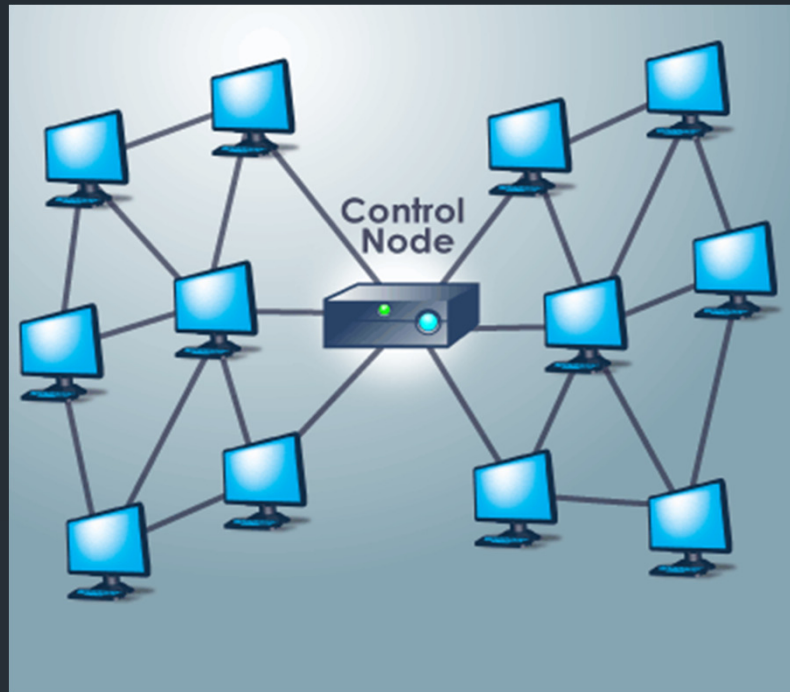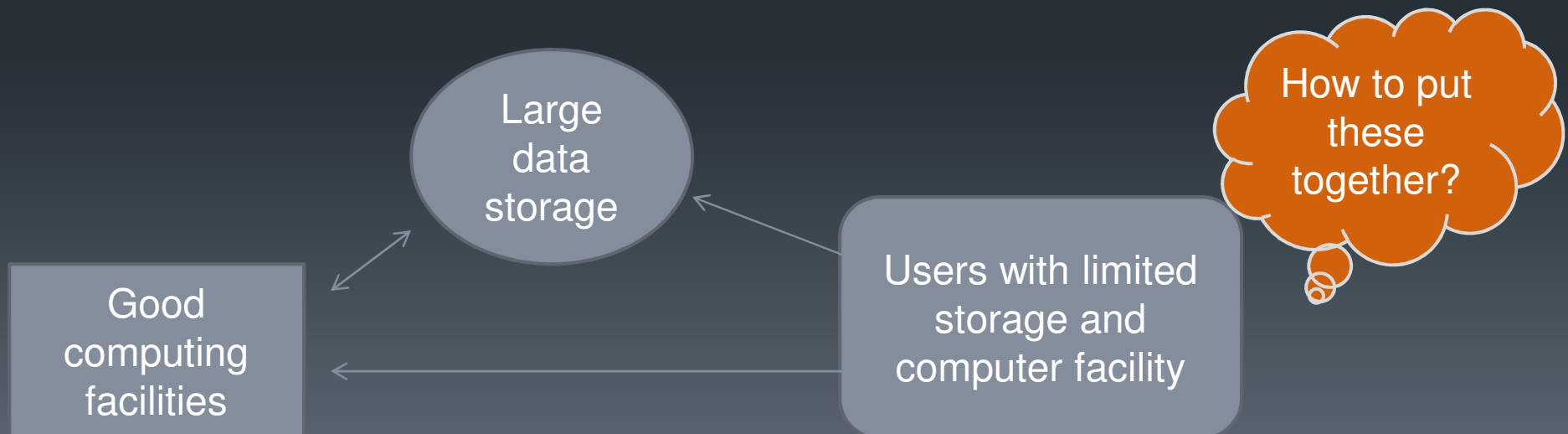
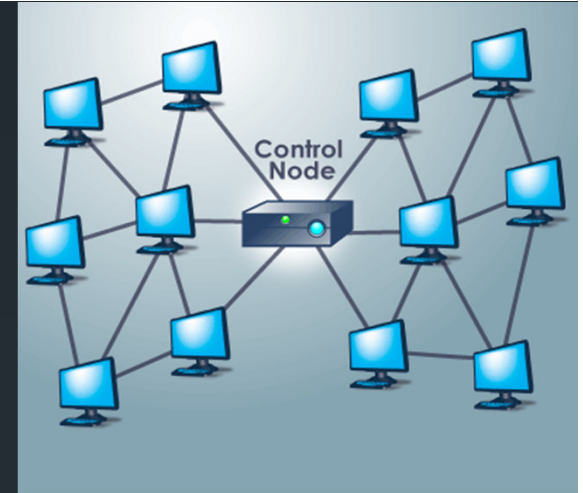# Exploiting SDO data

Computing facilities within ROB

- Data servers
  - 208TB redundant storage array
- Compute cluster
  - Processor X86_64 with Linux OS
  - GPU (CUDA programing language)
- High speed disk array
  - 12TB, several times faster than a normal disk, closely linked to data
- Condor software for effective management and exploitation of all available computing resources

- This is for internal use. What about external users?

Data intensive science

- Processing even only a few hours of SDO exceeds capacity of a single machine
- Numerical models also requires groups of computer (clusters)
- Some data centre may have more data, or more computing power
- Users will want to use part of these resources

Large data storage

Good computing facilities

Users with limited storage and computer facility

How to put these together?

Grid computing requires:

- A control node, for administrative duties
- A network of computers running special grid computing network software. These act as:
  - point of interface for the user
  - resources the system will tap into for different applications.
- A collection of computer software, called middleware, to allow different computers to run a process or application across the entire network of machines in a secure way

- Preliminary tests for including part of our SDO data into the B-grid system were successful

The way forward…

- Development of collaborative environment at European level.
  - This needs some trans-national resources (FP?)
- The European SDO data centres put together can provide a complete archive of SDO data
- Putting together computing resources
  - allow research accross Europe to have a wider ('free') access to those data
  - solve complex numerical simulation

Thank you for your attention
Questions welcome!