www.eu-egee.org

Grid computing

Assaf Gottlieb
Tel-Aviv University

assafgot @ tau.ac.il

http://iag.iucc.ac.il

EGEE is a project funded by the European Union under contract IST-2003-508833

Outline

- ***What is a grid & e-Research***
- The EGEE III grid
- Joining the grid

What is e-Research?

- Collaborative research that is made possible by the sharing across the Internet of resources (data, instruments, computation, people's expertise...)
  - Crosses organizational boundaries
  - Often very compute intensive
  - Often very data intensive
  - Sometimes large-scale collaboration
- Began with focus in the "big sciences" hence initiatives are often badged as "e-science"

What is e-Science?

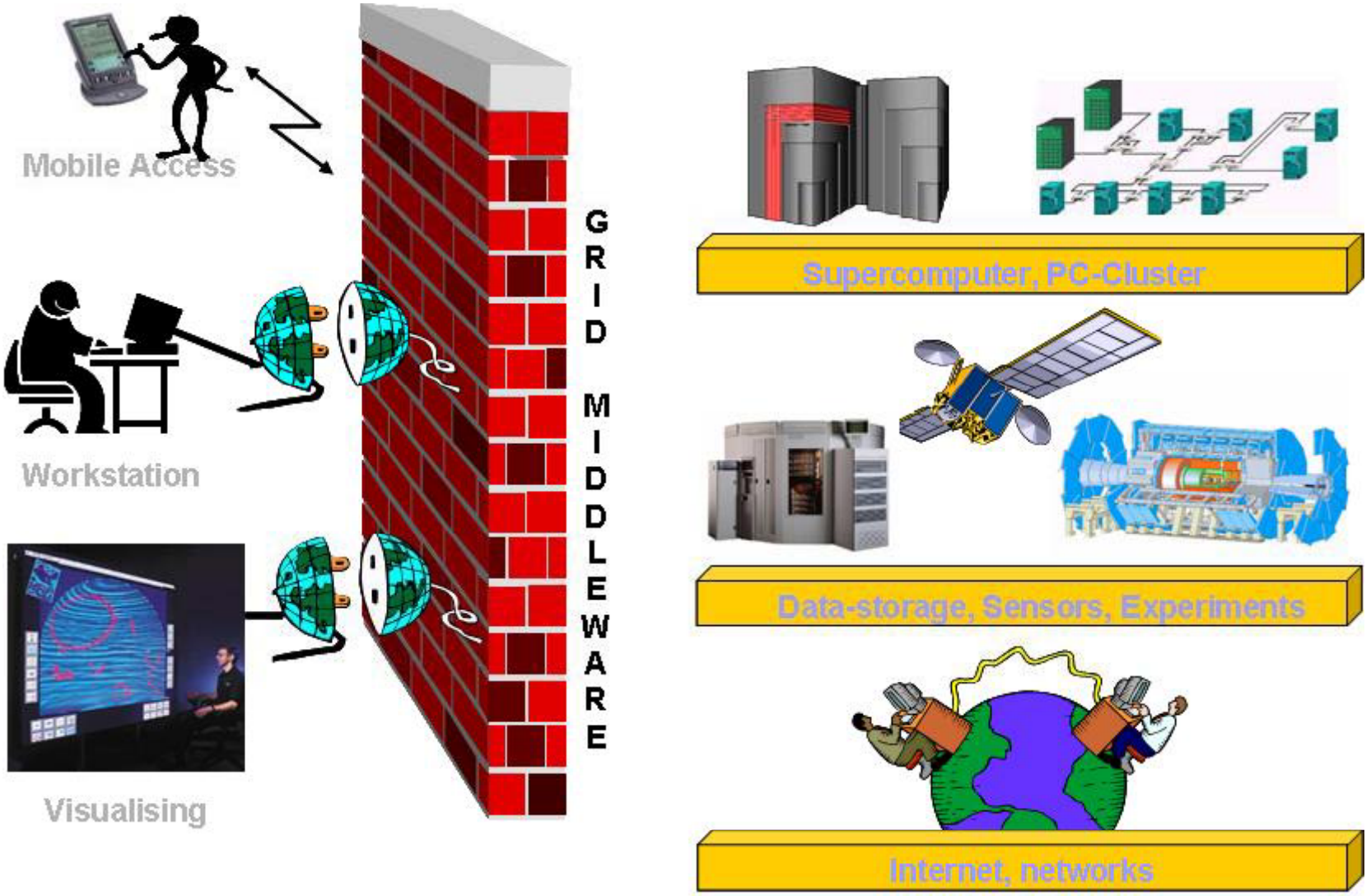
'e-Science is about global collaboration in key areas of science, and the ***next generation of infrastructure*** *that will enable it.'*

John Taylor
Director General of Research Councils
Office of Science and Technology, UK

Networks + Grids

- *Networks connect resources*
- *Grids enable "virtual computing"*

The Grid Metaphor

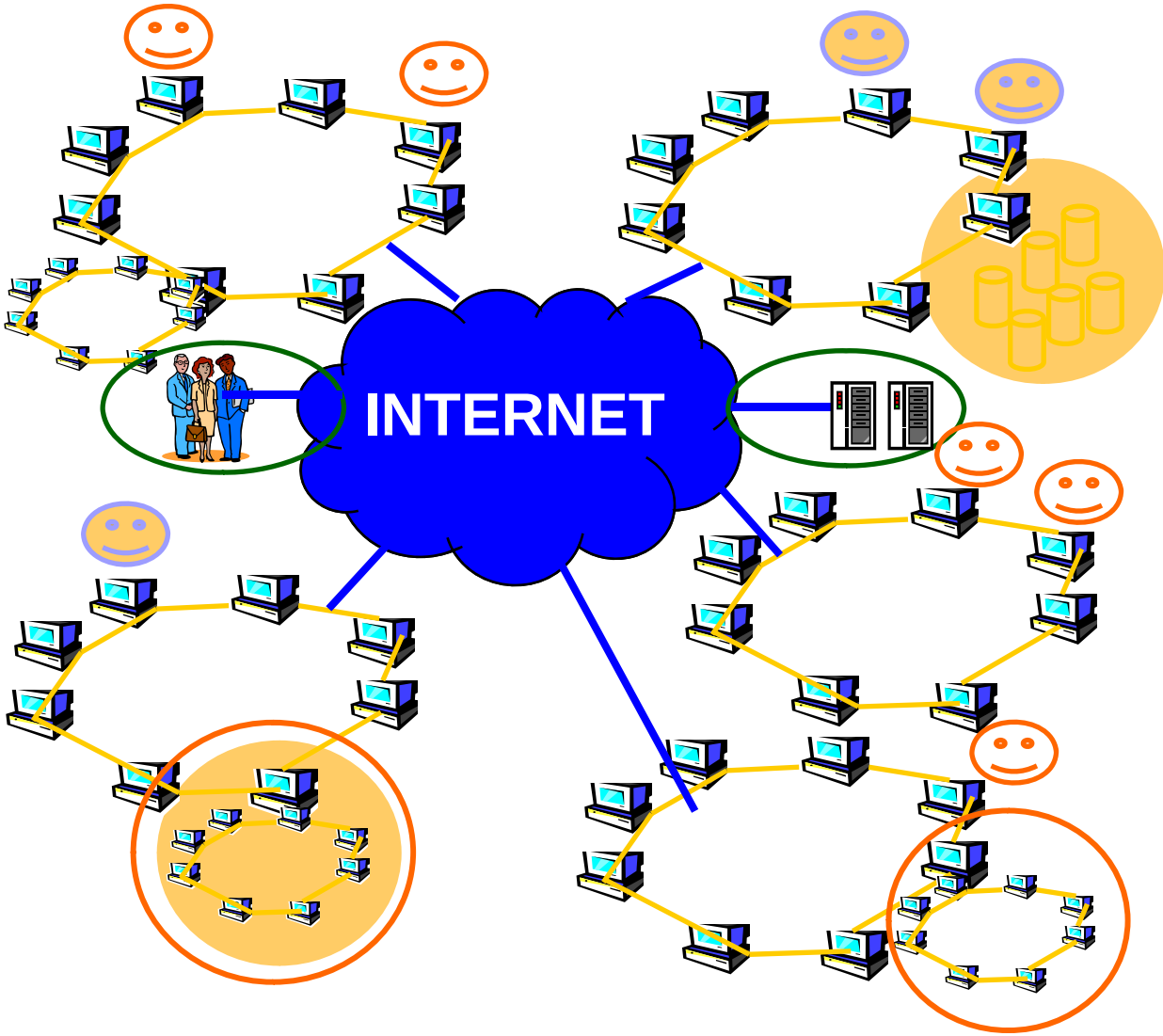


What is Grid computing?

- The term “Grid” has become popular!
  - Sometimes in Industry : “Grids” = clusters
    - Motivations: better use of resources; scope for commercial services
  - Also used to refer to the harvesting of donated, unused compute cycles
    - (SETI@home, Climateprediction.net)
  - These are e-Infrastructure but are not “grids” from the e-Research viewpoint!

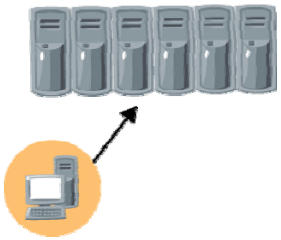
Typical grid

- Grid middleware runs on each shared resource to provide
  - Data services
  - Computation services
  - Single sign-on
- **Virtual Organisation:** People in different organisations seeking to cooperate and share resources across their organisational boundaries
- Virtual organisations negotiate with sites to agree access to resources

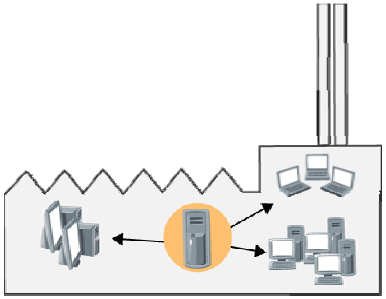


From a single PC to a Grid

Farm of PCs



Enterprise grid: Mutualization of resources in a company



Example:
Novartis

Volunteer computing: CPU cycles made available by PC owners

Examples:
Seti@home
Africa@home

Grid infrastructure: Internet + disk and storage resources + services for information management (data collection, transfer and analysis)

Example:
EGEE

Outline

- What is a grid & e-Research
- ***The EGEE III grid***
- Joining the grid

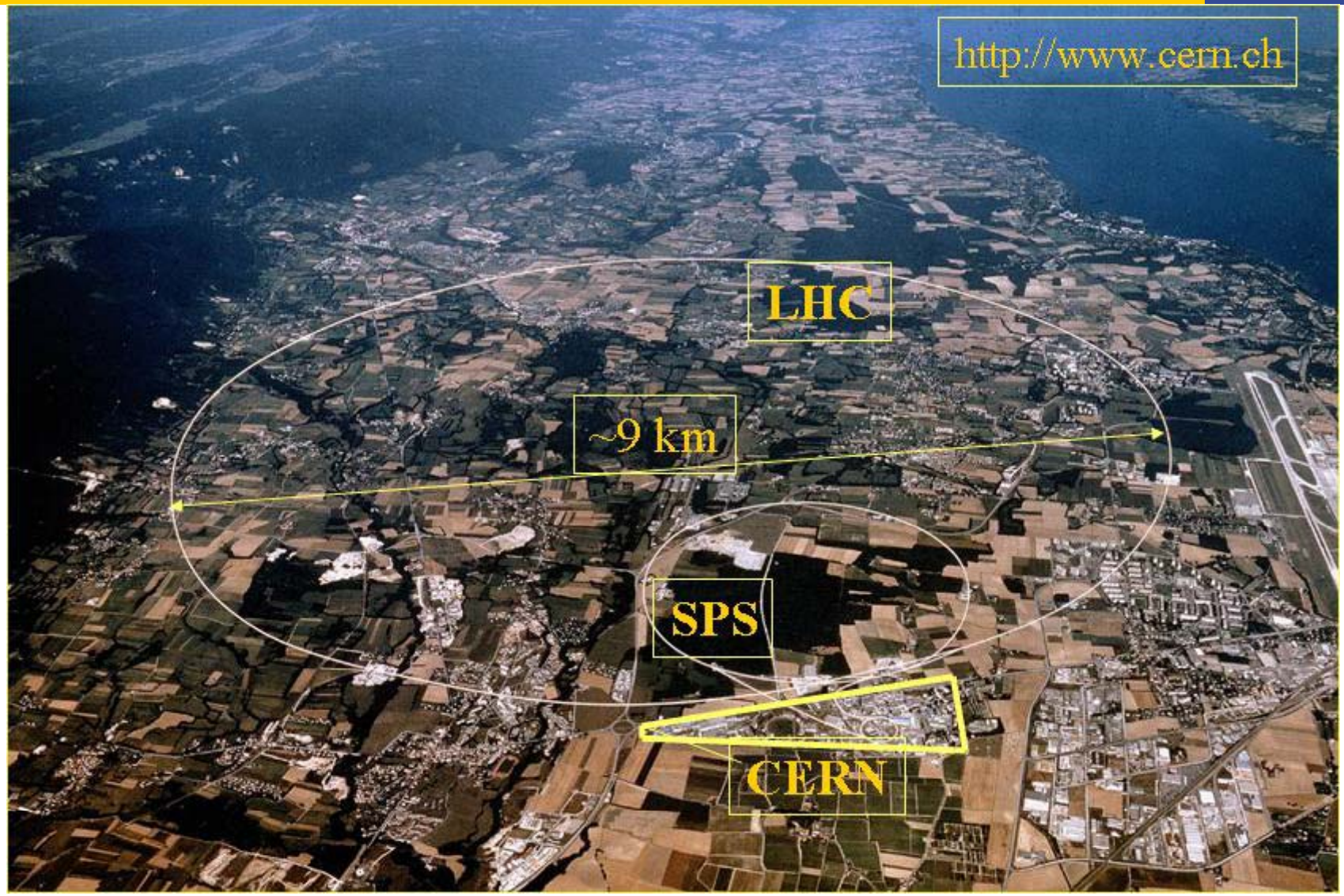
EGEE is …

EU-funded project that has established the largest multi-VO production grid in the world

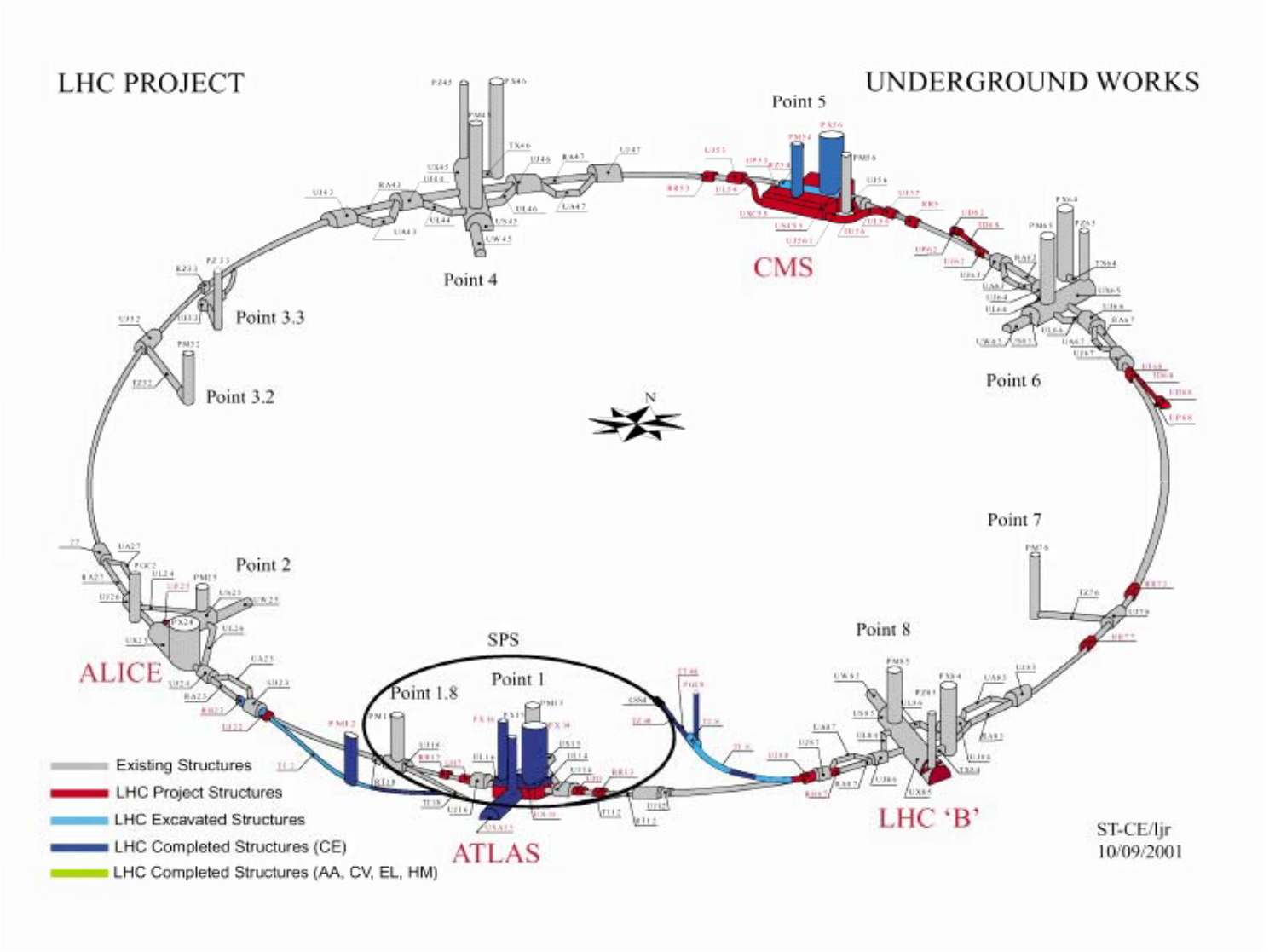
Project leader : CERN

- **245 sites**
- **45 countries**
- **>60 VOs**
- **51K CPUs**
- **~600PB Storage**
- **Avg 30K jobs/day**
- **Data transfers >1.5 GB/s**

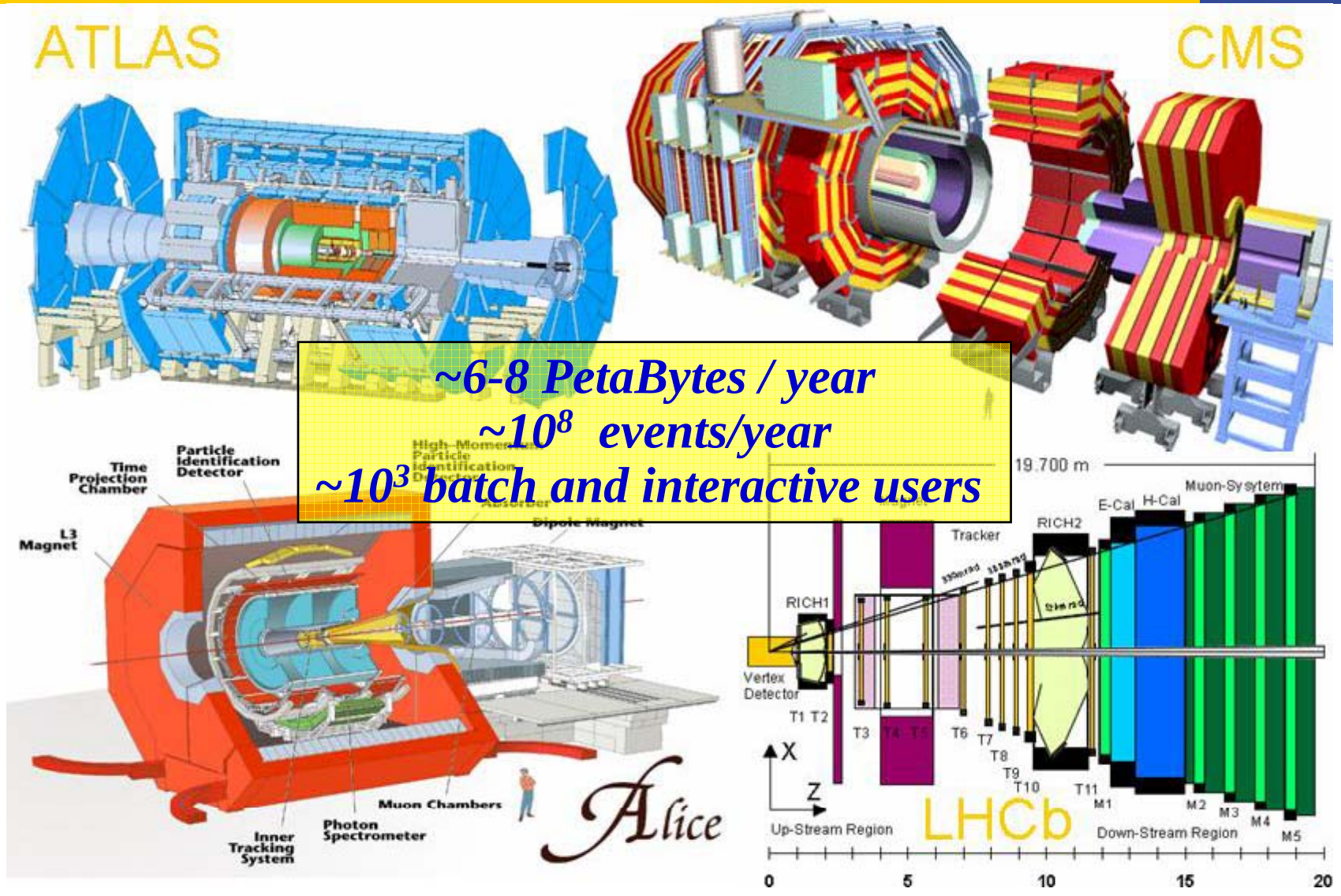
The Large Hadron Collider (LHC)



The LHC Experiments

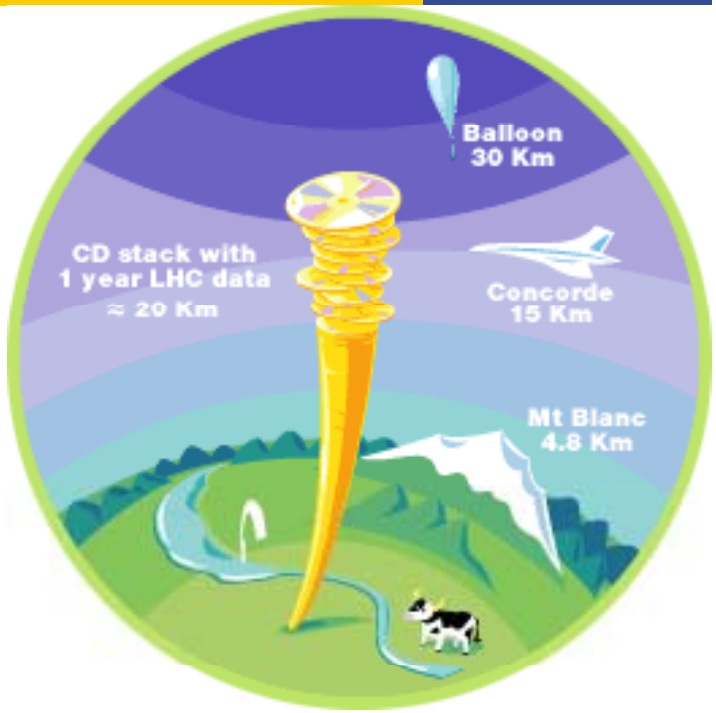


The LHC Experiments



CERN: Data intensive science in a large international facility

- The Large Hadron Collider (LHC)
  - The most powerful instrument ever built to investigate elementary particles physics
- Data Challenge:
  - 10 Petabytes/year of data !!!
  - 20 million CDs each year!
- Simulation, reconstruction, analysis:
  - LHC data handling requires computing power equivalent to ~100,000 of today's fastest PC processors!

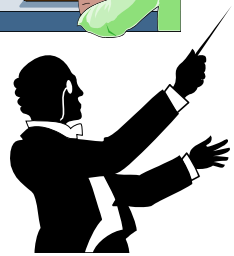


More Grid Applications

- Medical/Healthcare *(imaging, diagnosis and treatment)*
- Bioinformatics *(study of the human genome and proteome to understand genetic diseases)*
- Nanotechnology *(design of new materials from the molecular scale)*
- Engineering *(design optimization, simulation, failure analysis and remote Instrument access and control)*
- Natural Resources and the Environment *(weather forecasting, earth observation, modeling and prediction of complex systems)*

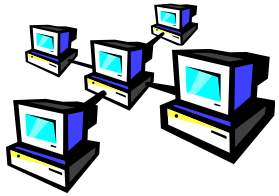
Main components

Access service How users logon to a Grid

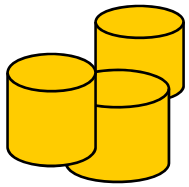


Resource Broker (RB): Service that matches the user's requirements with the available resources on a Grid

Information System: Characteristics and status of resources



Computing Element (CE): A batch queue on a site's computers where the user's job is executed



Storage Element (SE): provides (large-scale) storage for files

Outline

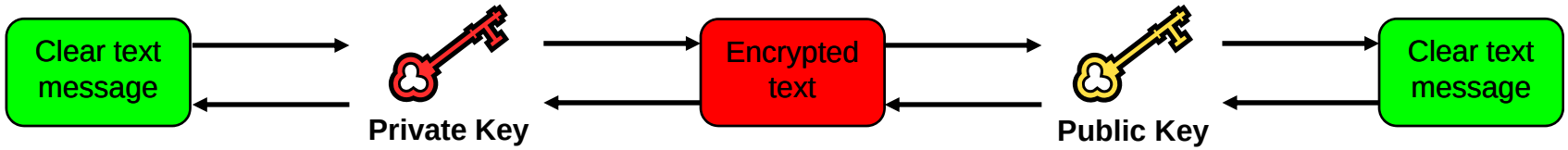
- What is a grid & e-Research
- The EGEE III grid
- ***Joining the grid***
  - ***Certificates***
  - ***Virtual Organizations***

Basic security concepts

- **Authentication**
  - Verify the identity of the peer
- **Authorization**
  - Map an entity to some set of privileges
- **Confidentiality**
  - Encrypt the message so that only the recipient can understand it
- **Integrity**
  - Ensure that the message has not be altered in the transmission
- **Non-repudiation**
  - Impossibility of denying the authenticity of a digital signature
- **Accounting**
  - What did you do, when did you do it and where did you do it from?

Public Key Infrastructure

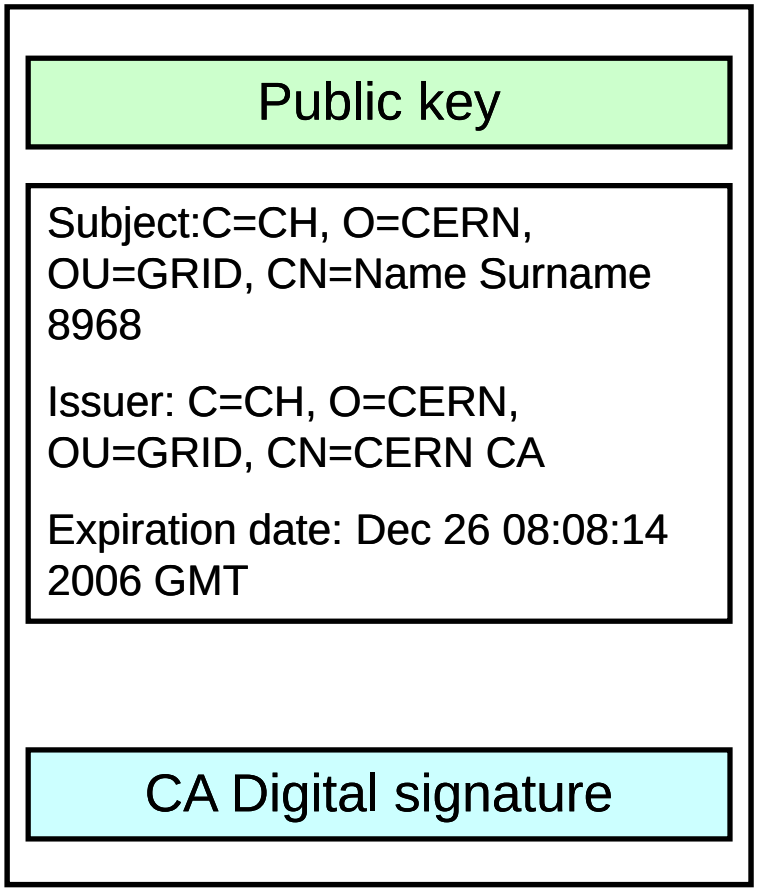
- Provides authentication, integrity, confidentiality, non-repudiation
- Asymmetric encryption



- Digital signatures
  - A hash derived from the message and encrypted with the signer's private key
  - Signature checked decrypting with the signer's public key
- Allows key exchange in an insecure medium using a trust model
  - Keys trusted only if signed by a trusted third party (Certification Authority)
  - A CA certifies that a key belongs to a given principal
- Certificate
  - Public key + information about the principal + CA signature
  - X.509 format most used
- PKI used by SSL, PGP, GSI, WS security, S/MIME, etc.

X.509: content of the Certificate

- An X.509 Certificate contains:
  - owner's public key;
  - identity of the owner;
  - info on the CA;
  - time of validity;
  - digital signature of the CA



User Responsibilities

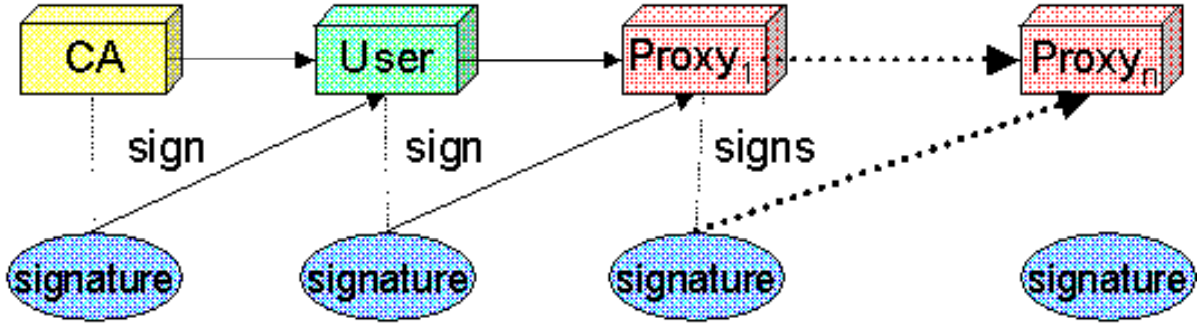
- Keep your private key secure.
- Do not loan your certificate to anyone.
- Report to your local/regional contact if your certificate has been compromised.
- Do not launch a delegation service for longer than your current task needs.

If your certificate or delegated service is used by someone other than you, it cannot be proven that it was not you.

IT IS YOUR PASSPORT AND CREDIT CARD

Globus Grid Security Infrastructure (GSI)

- *de facto* standard for Grid middleware
- Based on PKI
- Implements some important features
  - Single sign-on: no need to give one's password every time
  - Delegation: a service can act on behalf of a person
  - Mutual authentication: both sides must authenticate to the other
- Introduces proxy certificates
  - Short-lived certificates including their private key and signed with the user's certificate



Virtual Organizations and authorization

- Users MUST belong to a Virtual Organization
  - Sets of users belonging to a collaboration
  - Each VO user has the same access privileges to Grid resources
  - List of supported VOs:
    - https://lcg-registrar.cern.ch/virtual_organization.html

- VOs maintain a list of their members
  - Sites decide which VOs to accept
  - A list of supported VOs can be found here:
    - *https://lcg-registrar.cern.ch/virtual_organization.html*

Request a certificate and join a VO

- **Instructions - http://iag.iucc.ac.il/workshop/JoinGrid.htm**

Basic Certificate Request

Please enter your data in the following form.

Certificate Data

E-Mail	my.email@myserver.com
Name	Name LastName
Institution	TAU
alternative email	my.email@myserver.com

User Data

Name (first and Last name)	Name LastName
Email	my.email@myserver.com
Department	My Department
Telephone	My Phone number
Level Of Assurance chose the LOA you would like to be authenticated against.	Medium
Role	User
Registration Authority chose the RA where you will be authenticated.	Tel Aviv University
PIN [used to verify the certification request, min 10 chars (please write it down for later usage)]	**********
Re-type your PIN for confirmation	**********
Choose a keysize	1024

Continue

DN:	/C=IL/O=IUCC/OU=TAU/CN=Assaf Gottlieb
CA:	/C=IL/O=IUCC/CN=IUCC/Email=ca@mail.iucc.ac.il
CA URI:	http://iuccca.iucc.ac.il/pub/crl/cacrl.crl
Family Name:	Gottlieb
Given Name:	Assaf
Institute:	Tel Aviv University
Phone Number:	97236408337
Email:	assafgot@post.tau.ac.il
comment:	

I have read and agree to the VO's Usage Rules

I DO NOT agree to the VO's Usage Rules

Summary

- EGEE hold the largest production grid in the world to-date.
- Academic users can join a VO without cost.

- In order to use the grid a user must have
  - A valid certificate, given by the CA
  - Join a VO.
  - Have access to a grid user interface machine

- Each action on the grid requires a valid Proxy, generated from your certificate.