INGRID 2010
INSTRUMENTING THE GRID
MAY 12-14, 2010
POZNAŃ, POLAND

Synergies on different ESFRI initiatives in Physics

AN INSTITUTIONAL POINT OF VIEW

ESFRI projects and Grid technology:

how far can they walk together?

CSIC
Consejo Superior de Investigaciones Científicas

Jesús Marco de Lucas, CSIC-IFCA

SUMMARY

- What is the objective of this talk?
- What are the ESFRI initiatives in **Physics**?
- What is an “Institutional Point of View”?
- What is “synergy”?

The case for ESFRI projects and Grid technology:

- Findings from e-IRG and EEF
- Who does what? A more clear scheme ?
- *Advice/Opportunities*

What is the Objective of this talk ?

- OPPORTUNITY:
  - ESFRI proposals in “maturity” state
  - ECRI 2010 Barcelona
  - Prioritization in SPAIN of ESFRI proposals
  - Institutional Support within CSIC (Spanish National Research Council)
  - “NEW” European GRID e-Infrastructure: EGI
  - EEF Forum and e-IRG report

- IDEAS (learning from experience):
  - Identify technology challenges and promote common solutions
  - From the start of the ESFRI research projects!
  - Avoid duplication of efforts (and costs)…
  - …But include “competition” (effectiveness vs. effort)
  - Define “technology” areas and identify “platforms” and “infrastructures”
  - Address “technology oriented research”

- APPLIED TO e-INFRASTRUCTURES TO SUPPORT ESFRI
  - Identify a “USE CASE” (FUSION? ASTRO?)

CSIC
Consejo Superior de Investigaciones Científicas

Research Infrastructures, WHAT?

- The term ‘research infrastructures’ refers to **facilities, resources and related services** used by the scientific community to conduct top-level research in their respective fields, ranging from social sciences to astronomy, genomics to nanotechnologies.
- Examples include singular large-scale research installations, collections, special habitats, libraries, databases, biological archives, clean rooms, integrated arrays of small research installations, high-capacity/high speed communication networks, highly distributed capacity and capability computing facilities, data infrastructure, research vessels, satellite and aircraft observation facilities, coastal observatories, telescopes, synchrotrons and accelerators, networks of computing facilities, as well as infrastructural centres of competence which provide a service for the wider research community based on an assembly of techniques and know-how.
- RIs may be ‘**single-sited**’ (a single resource at a single location), ‘**distributed**’ (a network of distributed resources), or ‘**virtual**’ (the service is provided electronically).
- These key infrastructures have not only been responsible for some of the greatest **scientific discoveries and technological developments**, but are also influential in attracting the best researchers from around the world and in building bridges between national and research communities and scientific disciplines.

CSIC
Consejo Superior de Investigaciones Científicas

Research infrastructures, WHY?

- They are a key instrument in bringing together a wide diversity of stakeholders to look for solutions to many of the problems society is facing.
- They offer unique research services to users from different countries, attract young people to science, and help to shape scientific communities.
- New knowledge and, by implication, innovation, can only emerge from high-quality and accessible Research Infrastructures.
- For example, radiation sources, data banks in genomics, observatories for environmental sciences, systems of imaging or clean rooms for the study and development of new materials or nano-electronics are at the core of research and innovation processes.
- Moreover, they help to create a new research environment in which all researchers - whether working in the context of their home institutions or in national or multinational scientific initiatives - have shared access to unique or distributed scientific facilities (including data, instruments, computing and communications), regardless of their type and location in the world.

Research Infrastructures are therefore at the centre of the knowledge triangle of research, education and innovation, producing knowledge through research, diffusing it through education, and applying it through innovation.

CSIC
Consejo Superior de Investigaciones Científicas

Research Infrastructures, WHICH?

- INTERGOBERNAMENTAL RESEARCH INFRASTRUCTURES
  Examples: **CERN, ESO, *ILL, ESRF***

- RESEARCH INFRASTRUCTURES IN EU DATABASE
  - provide resources, facilities and services essential to the scientific community
  - have high construction and operation costs compared to research costs in their respective fields
  - be open to external researchers
  - have a clear European dimension and added value
  - **78 RI in PHYSICS AND ASTROPHYSICS** (may 2010)

- **ESFRI: European Roadmap for Research Infrastructures**
  2008 ROADMAP, Physical Sciences and Engineering:
  - Astronomy & Astroparticle: E-ELT, KM3NET, CTA
  - Particle Physics (via CERN): S-LHC, ILC/CLIC, Advanced neutrino facility, Underground Lab, Super-B
  - Space Science (via ESA): Cosmic Vision paper
  - Nuclear Physics: FAIR, Spiral2
  - Engineering: PRINS (Nanostructures)
  - High Intensity Lasers: ELI

CSIC
Consejo Superior de Investigaciones Científicas

An Institutional Point of View

- **The "naive" researcher approach**
  - Excellent ideas have to be funded!
  - Consortia are the first step to setup a good project
  - Funding will be granted…discoveries will arrive next!

- **An "institutional" point of view**

(replace Institutional by European Commission, National Science Ministries, Research Council and/or Universities Authorities, Funding Agemcies and related panels, Local Authorities, etc. etc. etc. etc. etc.)

  - Researchers may have a good idea but never know the cost
  - They can lobby and also get into "political" issues
  - Close communities don't collaborate so easily (i.e astrophysics and astroparticles), they will come one after other…
  - Distant communities simply don't know each other
  - All of them underestimate technology and management risks
  - Lack provision for maintenance, exploitation, industry collaboration (except to re-inject funds)
  - And how popular/supported is this initiative?
  - Anyway, we have the power…

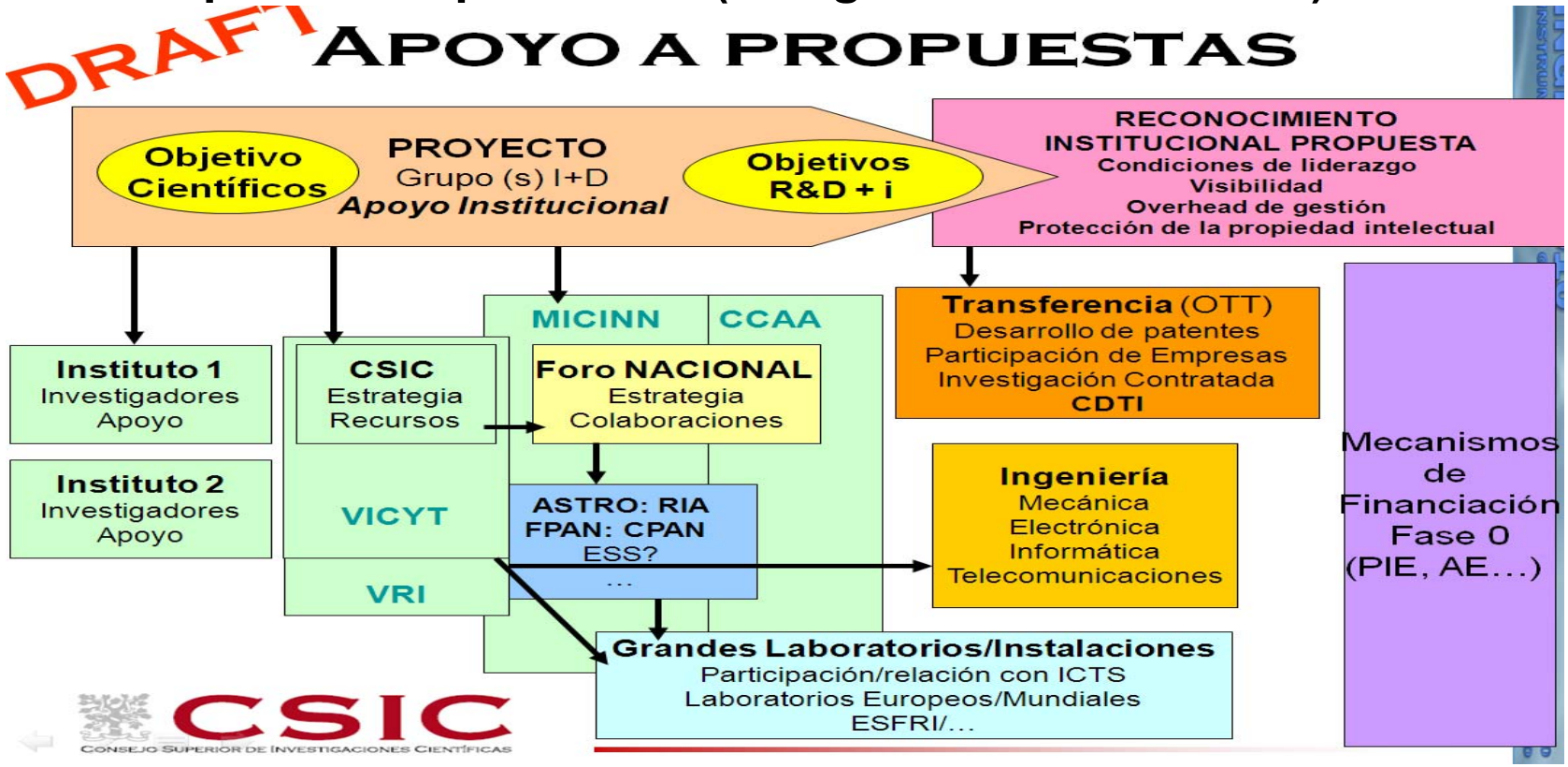
CSIC
Consejo Superior de Investigaciones Científicas

An Institutional OK?

- **Spanish Ministry of Science Strategy for participation in Scientific Infrastructures**
  - Large volume describing in detail all collaborations
  - Prioritization *(from very high to low)*:

E-ELT	FAIR	SKA	CTA	SPIRAL-2	KM3NET	ELI	PRINS

- **An operational procedure (being discussed in CSIC):**



συνεργία

- **Pronunciation: \ˈsi-nər-jē**
- **Function: noun**
- **Inflected Form(s): plural syn·er·gies**
- **Etymology: Latin synergia,**
  from Greek synergos: working together
- **Date: 1660**
- **1 : synergism; broadly : combined action or operation**
  2 : a mutually advantageous conjunction or compatibility of distinct business participants or elements (as resources or efforts)

- **EXAMPLE (false):**
  - **CTA and E-ELT could share the design of a robotic alignment system**
- **EXAMPLE (true?):**
  - **CTA and E-ELT could share the basic design of a long term persistent Virtual Observatory**

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

ESFRI , e-IRG, EEF

- ESFRI= European Strategy Forum on Research Infrastructures

The e-infrastructure aspect also needs particular attention, because these facilities will create such large amounts of data that new creative solutions for their transmission, analysis and preservation are needed, irrespective of the field of science

- e-IRG= e-Infrastructure Reflection Group

The efficient management of scientific data is one of the key challenges to be faced during the next years. ESFRI, together with e-IRG, will analyze this challenge and make recommendations. The best practices developed under the current ESFRI and e-IRG related projects will serve as important inputs to this reflexion work.

See e-IRG Report on Data Management Dec 2009

- EEF= European e-Infrastructures Forum

Forum for the Pan-European e-Infrastructures providers in the area of High Performance Computing, Networking, Secure Data Storage, and Services and the European Grid Infrastructure

- Report: ESFRI project requirements for PAN-European e-infrastructure resources and facilities

(28th April 2010, http://www.einfrastructure-forum.eu)

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Some Findings after Surveys...

- **EEF Report**

Conclusions from the Initial Requirements Analysis:

  - Single sign-on
  - **Virtual Organizations**
  - Persistent Storage
  - User Support
  - Training and Consultancy
  - Web-service interfaces
  - Workflows
  - Global Scope
  - Integration with cloud systems and volunteer desktop systems

- **e-IRG Report on Data Management**
  - METADATA
  - QUALITY
  - INTEROPERABILITY

CSIC
Consejo Superior de Investigaciones Científicas

MY INITIAL COMMENTS

- **Who really cares about e-Science and e-Infrastructure?**
  we started by a very challenging initiative, LHC, but do all ESFRI research communities need sophisticated technology? On the other hand, everybody happy to copy data at 10Gb/s!
- **What about legacy? (communities and applications)**
  again, experience with LHC was singular (and funding was high!)

- Single sign-on

CERTIFICATES: FIND A SIMPLER WAY (electronic ID?)

- Persistent Storage :

PERSISTENCY ALSO MEANS METHODS!

- User Support :

IMPLEMENT *A LA* ITIL

- Workflows:

GRAPHIC TOOL (a la Labview) needed, do we have it ready?

- Integration with cloud systems:

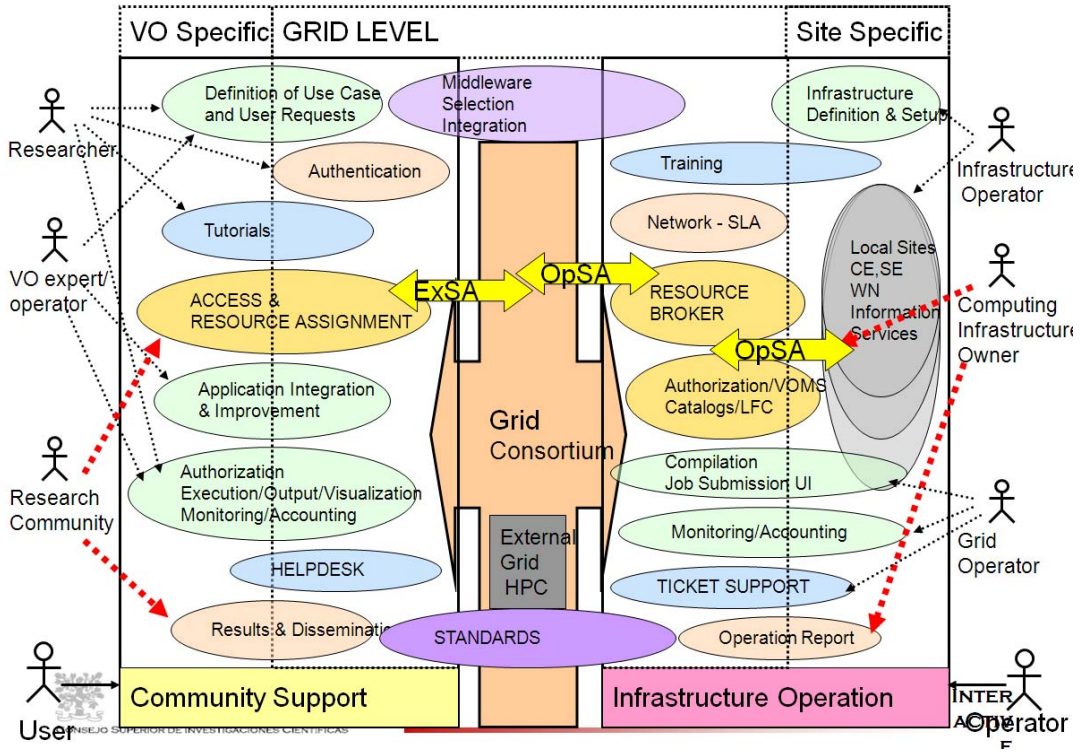
DEFINE AND PUBLISH GRID SERVICES

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

WHO DOES WHAT IN THE e-INFRASTRUCTURE ECOSYSTEM?

Are new communities to setup and operate new e-Infrastructure?

- DEFINITELY NOT FOR NETWORKING! (except remote access)
- NOT IN MOST CASES FOR SUPERCOMPUTING
- WHAT ABOUT GRID?



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Synergy Opportunities

ON e-INFRASTRUCTURE OPERATION:

- **ONLY ONE INFRASTRUCTURE, SHARED AND UNIFORM:**
  - Well supported Open Source SO: Scientific Linux
  - Virtual machines custom for each VO/community
  - Open Source queue system with support for OpenMPI parallel jobs: SGE
  - Distributed File System: GPFS/Lustre
  - LIMIT SITE TO >500 cores, >100 Tb
- **REDUCE GRID MIDDLEWARE OPTIONS (but keep gLite!)**
  - Example 1: ONE Stable version on SL5 and 64 bit
  - Example 2: ONE BASIC Storage Element: STORM (?)
- **DAILY, PROACTIVE, FOLLOW-UP OF EXECUTION BY VO (CHALLENGE-LIKE, THIS COSTS MONEY!)**
  - Both monitoring and accounting
- **INTEGRATE ADVANCED SERVICES *A LA* CLOUD**
  - Example: DB instances, large file collections, etc.
- **ITIL IMPLEMENTATION**
  - Management Procedures extended to GRID environment

Synergy Opportunities

ON COMMUNITY SUPPORT:

- IMPLEMENT BETTER AND SIMPLER GATEWAYS
  - Agree on tool for user interface
  - Agree on libraries, integrate Graphics
  - Agree on basic script language (like Python?)
  - Agree on WEB SERVICES
- BETTER USER INTERFACE INCLUDING SINGLE-SIGN ON (CERTIFICATE) AND VO ROLES MANAGEMENT TOOLS
  - Why not KERBEROS?
  - Why not a “REAL” account if we do not have >100.000 users
  - Why not an AFS domain per institution?
- DATA MANAGEMENT TOOLS:
  - Basic DB instances available
- DAILY, PROACTIVE, FOLLOW-UP OF EXECUTION
  - Challenge mode, this costs money!

FOR FINAL USERS:

- ACCESS TO A SINGLE AND “AGIL” PORTAL
  - To record all activities
  - To direct to best resource (supercomputer, grid, etc.)

CSIC
Consejo Superior de Investigaciones Científicas