

PIC
port d'informació científica

Increasing the value of data storage through e-Infrastructures

Prof. Manuel Delfino Reznicek
Director, Port d'Informació Científica, Barcelona

Invited talk at the Third European Conference on Research Infrastructures
Nottingham, UK, 6-7 December 2005

PIC
port d'informació científica

Outline

- PIC: A Data Storage and Transformation e-Infrastructure accessible to global users
- Use cases:
 - General Medical Image data from UDIAT/Parc Taulí Hospital
 - Neuroradiology data from IDIBAPS/Hospital Clínic
 - Turbulent Flow Simulation Results from ETSI/UPM
 - Molecular Biology Simulation Checkpoints from IBB/UAB
 - Satellite Imagery for Emergencies from UNOSAT
- Common Features
- Some drawbacks and possible solutions
- Conclusions

7 December 2005

PIC
port d'informació científica

PIC: A Data Storage and Transformation e-Infrastructure accessible to global users

6000 slot STK L5500 tape robot

29 racks

Edifici D, Campus UAB
Provided by UAB:

- 24 hour/day security and alarms
- Cooling and electricity
- Cleaning and maintenance

Rack smoke detectors interlocked to electrical panel

Potassium Carbonate fire suppression triggered by laser

200 KVA electric input to room
Merlin-Gerin UPS
Active Harmonic Compensation
500 KVA Rolls-Royce generator

7 December 2005

PIC
port d'informació científica

Use cases: Medical Image data from UDIAT/Parc Taulí Hospital

- Moving to all-digital imaging equipment
- Keep copy on DVD for legal reasons

CR - XA 2 CT 3 US 2 MR 2 NM

Estudios recibidos mensualmente

Corporació Parc Taulí

UDIAT Centre Diagnòstic

- PACS + e-Infrastructure=Access Tape Robot at PIC
 - Handle data explosion: Adding Digital Mammography doubled!
 - Clinical recall: Maintain response time, allow "remote" access
 - Make data available for research by multi-disciplinary teams

7 December 2005

PIC
port d'informació científica

Use cases: Neuroimage data from IDIBAPS/Hospital Clínic

- Neuroimaging: Magnetic Resonance of brain activity
- Data obtained in hospital environment
- Modest data volume requires substantial processing

- Make data accessible to CPU farm at PIC
- Re-process data from 30 patients in parallel
- Create a focal point for multi-disciplinary research: hospital doctors, university physicists, data handling experts

7 December 2005

PIC
port d'informació científica

Use cases: Turbulent Flow Simulation Results from ETSIA/UPM

- Reynolds number 2000 turbulent flow on Mare Nostrum
- Best calculation of this type ever
- 25 TB of output - a "gold mine" to be analyzed

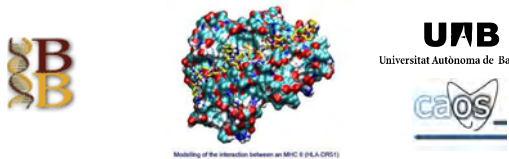
- Data stored in quasi-real time at PIC
- Fast access for Stanford University collaborators
- In process of implementing audited access to data sets

7 December 2005

PIC port d'informació científica

Use cases: Molecular Biology Simulation Checkpoints (IBB/UAB)

- All UAB desktops run simulations parasitically
- Condor pool setup by UAB Computer Science Dept.



UAB
Universitat Autònoma de Barcelona

caos

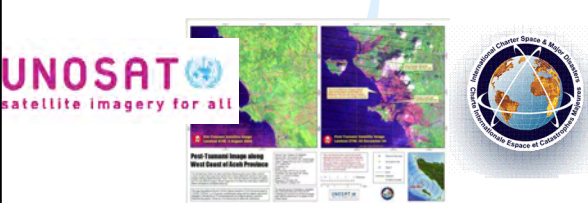
- Desktop must react instantly if user arrives
- Need safe place for checkpoint data and outputs
- One calculation can produce 250 Kfiles stored at PIC

7 December 2005 Increasing the value of data storage through e-Infrastructures/II, Delft/EP/Third European Conference on Research Infrastructures

PIC port d'informació científica

Use cases: Satellite Imagery for Emergencies from UNOSAT

- International Charter provides access to images
- Need flexibility in finding storage and processing



UNOSAT satellite imagery for all

- Implementing over e-Infrastructure can enhance:
 - Wider reach to get CPU and storage donations
 - Speculative caching of data: Keep private until need arises
 - Bandwidth donations

7 December 2005 Increasing the value of data storage through e-Infrastructures/II, Delft/EP/Third European Conference on Research Infrastructures

PIC port d'informació científica

Common Features

- Multiple institutions (legal entities) involved
- "Owners" of the project and its data are not particularly interested in running their own data facility
- Data storage has a "quasi-online" component
- Data access load has large variations
- Serving many projects in an infrastructured manner can create economies of scale
- Benefits arise when scientists from multiple institutions can transparently use the various resources as a "Virtual Organization"
- But...these are small V.O.s and there will be thousands of them

7 December 2005 Increasing the value of data storage through e-Infrastructures/II, Delft/EP/Third European Conference on Research Infrastructures

PIC port d'informació científica

Some drawbacks and possible solutions

- The "who pays" syndrome:
 - Scientist gets funding, but the rules of many funding agencies make it complicated to reimburse the storage provider
 - Pressure: "spend the money at home" (is "home" well defined?)
 - Pressure: "why should my institute pay to support data processing for researchers at some other institute" (even though they are crucial for the best quality research!)
 - Worries that funding the "Storage e-Infrastructure" directly will result in misuse or uncontrolled usage
- Can be circumvented with Collaboration Agreements
- But signing a CA that involves money can be a lengthy and bureaucratic procedure – not the best for many small V.O.s
- Another solution: Fund the Storage e-Infrastructure directly to give "Storage Grants" to projects
- Is experience from supercomputing or networks applicable?

7 December 2005 Increasing the value of data storage through e-Infrastructures/II, Delft/EP/Third European Conference on Research Infrastructures

PIC port d'informació científica

Conclusions

- Storing data on e-Infrastructures can increase its value
 - Usage of e-Infrastructures can be done preserving the beneficial part of Outsourcing
 - Future opportunity for industry to provide advanced services
- Need to gain more experience without being blocked by the "who pays" syndrome
 - Stimulate more examples under realistic conditions
 - Fund a network of large and small Storage e-Infrastructures connected between them and to the users via Grid and Communications e-Infrastructures (e.g. EGEE and GEANT-2)
 - Involve industry in funding, building and operating these Storage e-Infrastructure Centers
 - PIC has made a modest contribution to show the way
- Win-Win situation: Better research in fields with data, better development and innovation in e-Infrastructures

7 December 2005 Increasing the value of data storage through e-Infrastructures/II, Delft/EP/Third European Conference on Research Infrastructures

PIC port d'informació científica

Acknowledgments

- Dr. Joan Guanyabens and Dr. Melcior Sentis, UDIAT
- Dr. Beatriz Gómez and Ms. Gemma Monte, IDIBAPS
- Prof. Javier Jiménez and Dr. Sergio Hoyas, ETSI/UPM
- Prof. Xavier Daura, IBB/UAB
- Mr. Alain Retiere, UNOSAT/UNITAR
- RedIRIS (Spanish NREN) and Anella Científica (Catalan RREN)
- PIC is maintained through a collaboration agreement of
 - Generalitat de Catalunya Departament d'Universitats, Recerca i Societat de la Informació
 - MINISTERIO DE EDUCACION Y CIENCIA
 - Cimat Centre de Investigacions Energètiques, Bioquímiques i Tecnològiques
 - UAB Universitat Autònoma de Barcelona
 - IFAE Institut de Física d'Altes Energies
- PIC is the coordinator of the SouthWest Europe Federation of
 - EGEE Enabling Grids for e-science
 - funded by contract INFISO 508833
 - Information Society and Media
 - European Union

7 December 2005 Increasing the value of data storage through e-Infrastructures/II, Delft/EP/Third European Conference on Research Infrastructures