Manchester Computing
Supercomputing, Visualization & eScience

Managing Supercomputing Resources at the University of Manchester

Mike Pettipher
(m.pettipher@man.ac.uk)

Date 15th May 2003
Event: Cray User Group
Venue: OSC, Columbus, Ohio

THE UNIVERSITY of MANCHESTER

Manchester Computing
Supercomputing, Visualization & eScience

The University of Manchester
and Manchester Computing

University of Manchester

- Established 1851
- One of the largest in the UK
  - Student numbers
  - Research Income
  - …
- One of the best in the UK
  - Research quality
  - Graduates getting jobs
  - Teaching

Manchester Computing users

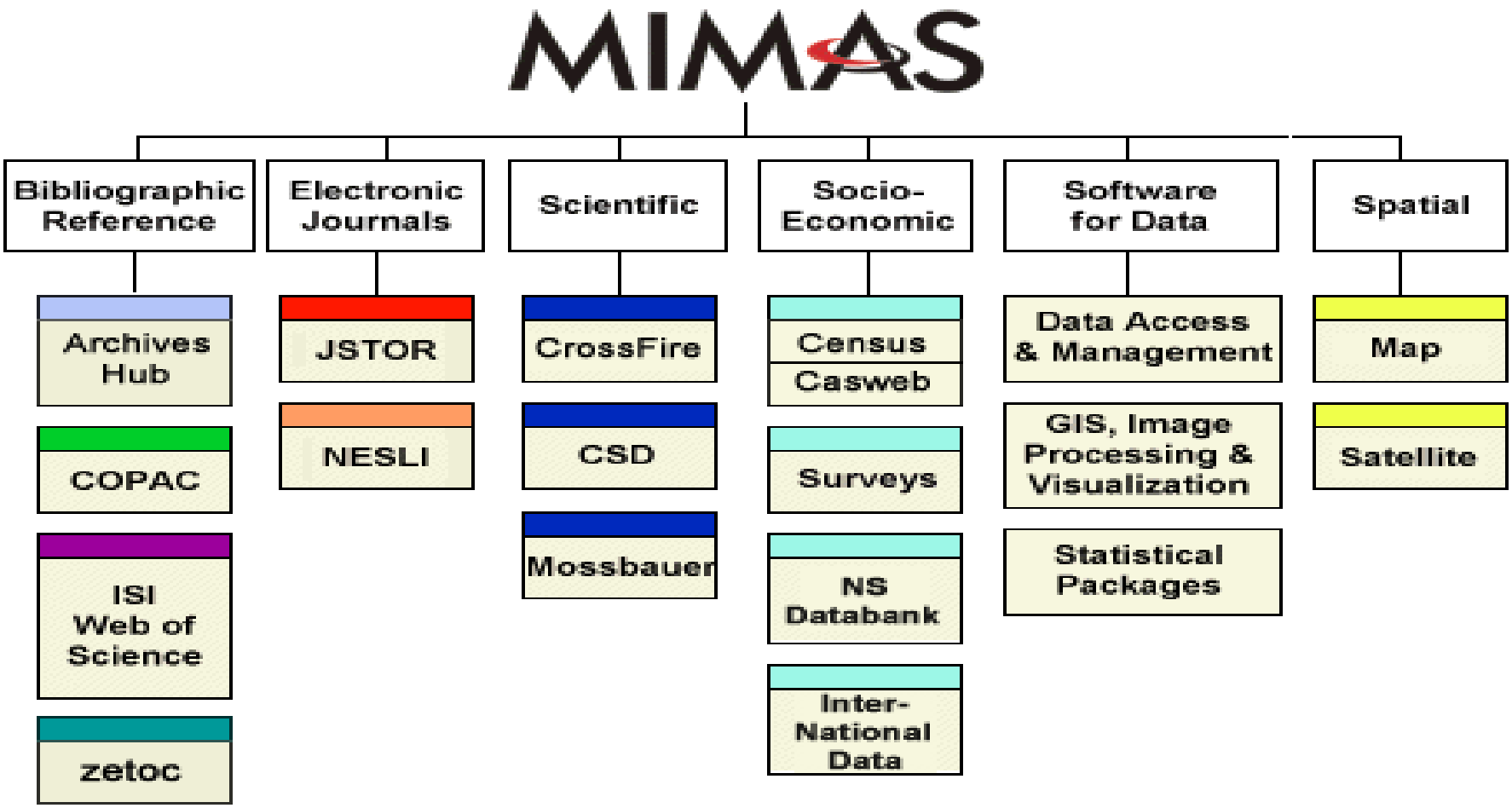
- University of Manchester
  - Administrative computing, academic computing, telecoms
- UK Academia
  - Supercomputing (CSAR)
  - Information & data services (MIMAS)
  - Major node in UK Academic Network
    - Managing agent for Net NorthWest
  - Used by HE, FE & RCs
- International
  - MIMAS
  - International AVS Centre
- Government, Commerce & Industry
  - Supercomputing
  - Networks and hosting
    - IFL
  - R&D

Supercomputing @ Manchester: A Brief History

Year	MFlops	Machine
1948	0.0007	Williams-Kilburn Baby
1972	12	CDC 7600/ICL1906A
1977	12	CDC 7600/ICL1904S
1983	192	CDC Cyber 205/Amdahl 470
1988	576	Fujitsu VP1200
1993	2,200	Fujitsu VPX 240/10
1998	0.7M	Cray T3E 1200e, 576 PE
2000	1M	T3E upgrade to 816 PE
2000	0.1M	SGI Origin2000, 128 PE
2001	~0.5M	SGI Origin3800, 512 PE

Others
Sun
Cray CS6400
Meiko (80PE)
KSR
IBM SP (146PE)
Cray EL98
Compaq
Amdahl
Fujitsu
Digital

National Datasets



Machine Room

- MIMAS
  - Sun E6500 (MIMAS), E4500, E4000 (JSTOR)
- CSAR
  - Cray T3E-1200E
  - SGI Origin 3800
  - SGI Origin 2000
  - SGI Altix (soon)
- Manchester
  - IBM RS6000/SP 146 PEs
  - Beowolf cluster 204 Pentiums
  - SGI Origin : >5

Manchester Computing – Managing Resources

- Structure
  - Administration
  - Communications, Operations and Systems (COS)
  - Information Services (IS)
  - Manchester Computing Informations System Office (MCISO)
  - Manchester Information and Associated Services (MIMAS)
  - Supercomputing, Visualization and e-Science (SVE)
- System administration managed by COS group at MC, except for CSAR – managed by CSC.
- Devolvement
  - Faculty Support Units
  - Principal Investigators
  - End users

Manchester Computing
Supercomputing, Visualization & eScience

Introduction to UK (Academic) High End Computing

UK High End Computing

- High End Computing
  - National supercomputing facilities
  - SRIF/JREI funded local HPC/cluster facilities
  - Locally/other funded institution/department HPC/cluster facilities
- Money from Research Councils UK (RCUK)
  - Managing agent is EPSRC
- Strategy and implementations formed by
  - HPC Strategy Committee
  - Technology Watch Panel
  - Private Consultants
  - National HPC Services

National HPC services

- UK has 2 national supercomputing services:

- CSAR, provided by CfS consortium:
  - University of Manchester (SVE Group)
  - Computer Sciences Corporation (CSC)
  - SGI
  - A Private Finance Initiative (PFI) procurement.

- HPCx, provided by HPCx consortium:
  - Edinburgh Parallel Computing Centre (EPCC)
  - Daresbury Laboratory (DL)
  - IBM
  - Not a PFI

Manchester Computing
Supercomputing, Visualization & eScience

CSAR Access

THE UNIVERSITY of MANCHESTER

How it works

- Principal Investigator
  - Scientific Case
  - Justify resources
  - List of resources
- RC
  - Referees report
  - Comments from CSAR
- Allocation of Tokens

Applying for Resources

- CSAR Resource Application Form
- Class 1: Full Peer Review
- Class 2: Pump-priming
- Class 3: New application areas
- Class 4: ‘Commercial’ use

Getting Resources

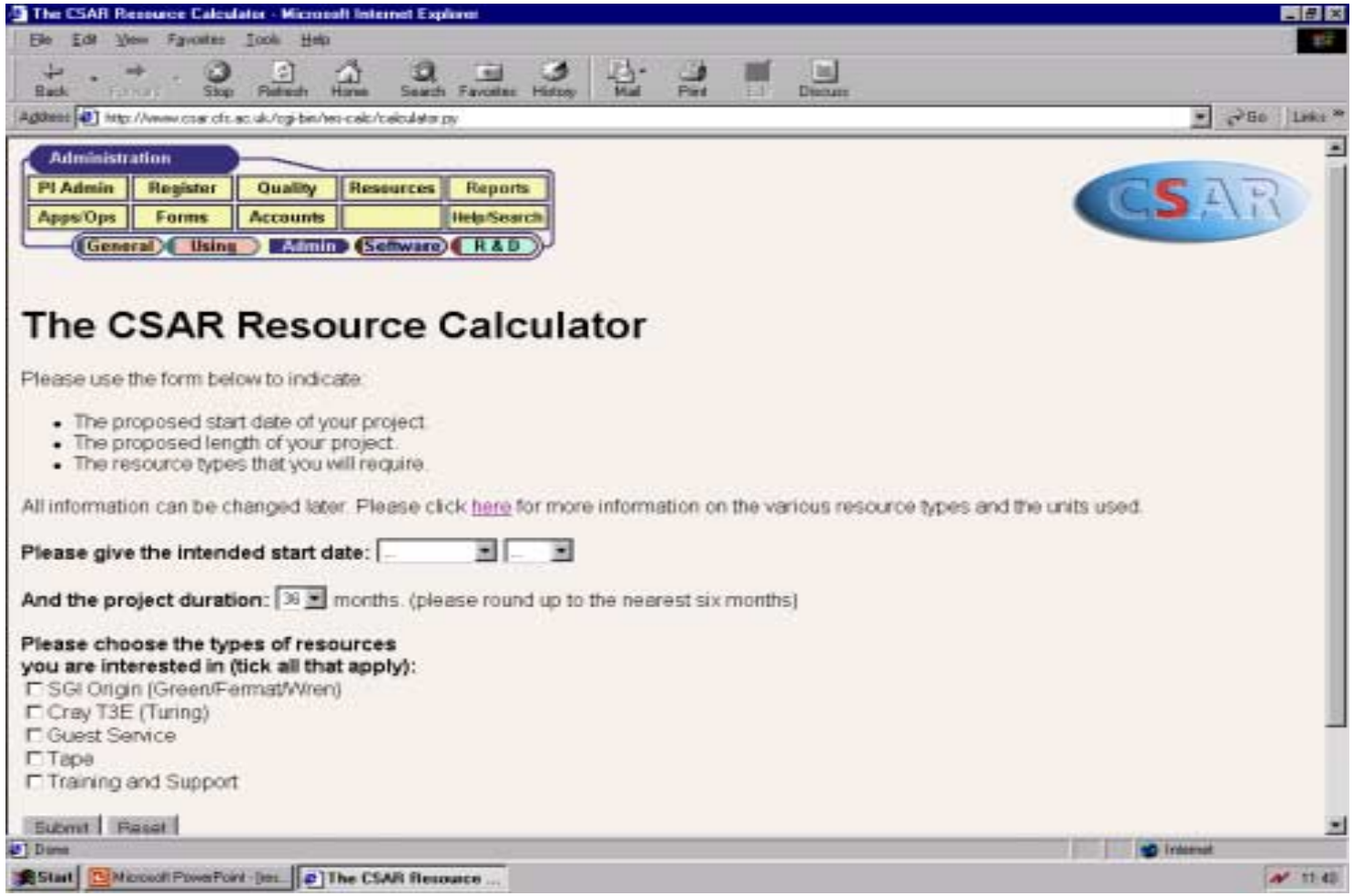
- PI applies to RCs (with advice/help from us)
- We comment upon application
- We enter resources into database (as proposed)
- We complete initial capacity plans
- RCs approve grant. PI gets tokens
- We change proposed -> live
- We perform trades and update capacity plans if necessary
- PI invited to register – we approve
- PI invites colleagues to register
- User fills in web page
- PI gets email and through web page approves user
- User gets email with username

- Every month RCs are invoiced for tokens used

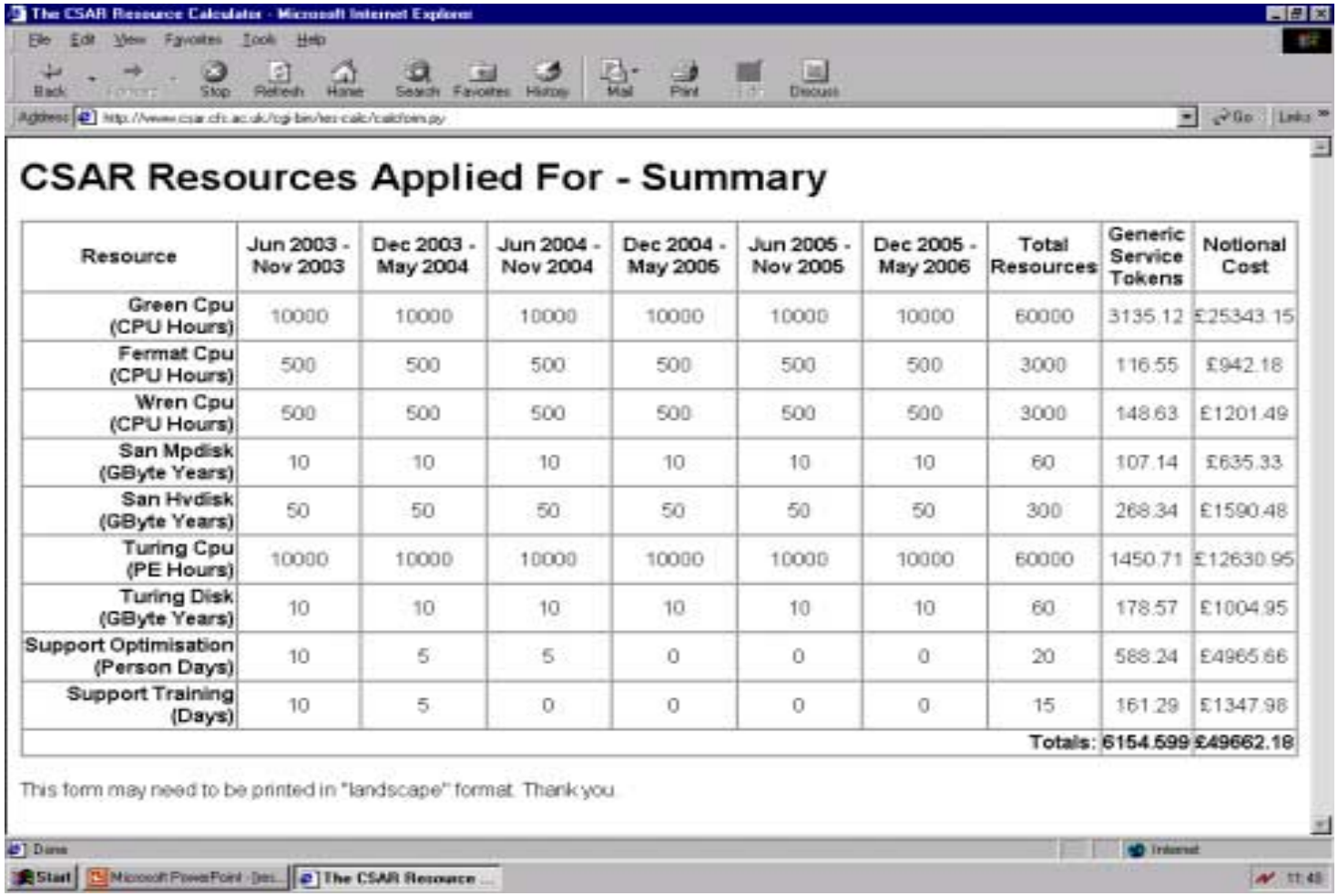
‘Purchasable’ Resources

- Turing, Fermat, Green and Wren cpu time
- Turing, SAN (Origin) disk
  - (can use either on the ‘other’ machine)
- Tape storage
  - (much cheaper than disk)
- Training Courses
- Optimisation/Application support
- Guest System resources

CSAR Resource Calculator



Resources Applied For



Manchester Computing
Supercomputing, Visualization & eScience

Registration (Resource Management) System

General Requirements

- Minimise administration
- Make it `easy' to register
- Rapid response
- Web based
- Hierarchical structure to allow delegation/devolvement. 'Approved users' (e.g. Principal Investigator) should be able to manage as many of the resources as possible.

CSAR Requirements

- Self Registration.
- Allocate any resource.
- Resource trading.
- Facility for sub-project management.
- Individual user allocations.
- Usage reporting.
- Capacity Planning.
- Principal Investigator is primary administrative level. Must be able to:
  - Authenticate users.
  - Sub-allocate resources within group or consortium.
  - Change users allocations and disk quotas.

Local HPC Requirements

- Subset of CSAR facilities:
  - Similar organisation using PIs, but registration managed centrally.

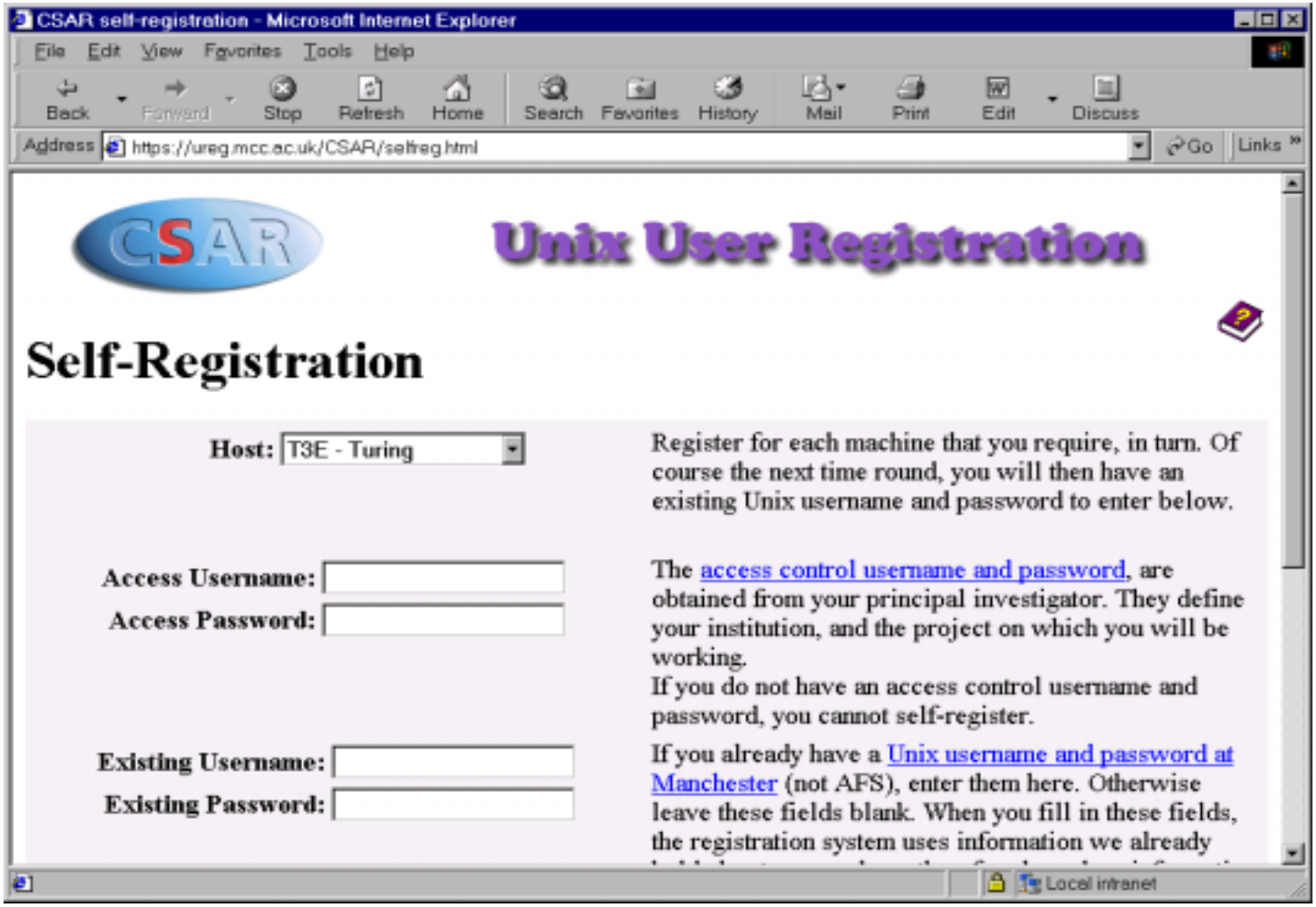
MIMAS Requirements

- Registration of the MIMAS service users (~7500 users).
  - Central registration.
  - Resource management (changing passwords, disk quota etc.) devolved to site representatives
- Crossfire service:
  - Self registration for users (~13000 users).
  - Authentication by Athens service.
- Some requirements shared with CSAR; others specific to MIMAS.

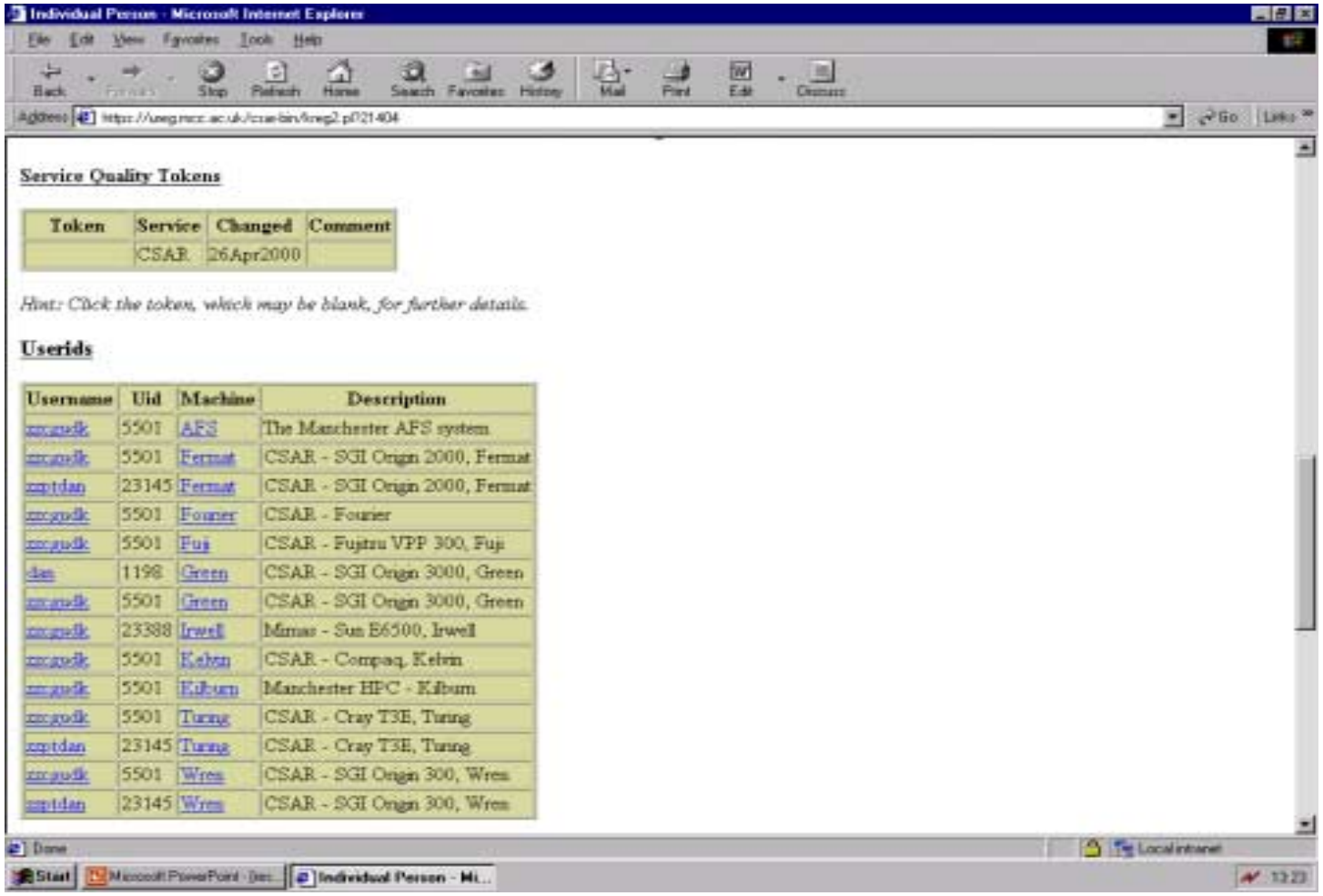
Other Requirements

- Logging of all transactions.
- Live and test sites to allow continued development.
- On-line help.
- Compatiblity with Grid activities.
- Integration with query management system.
- Feedback from user community.

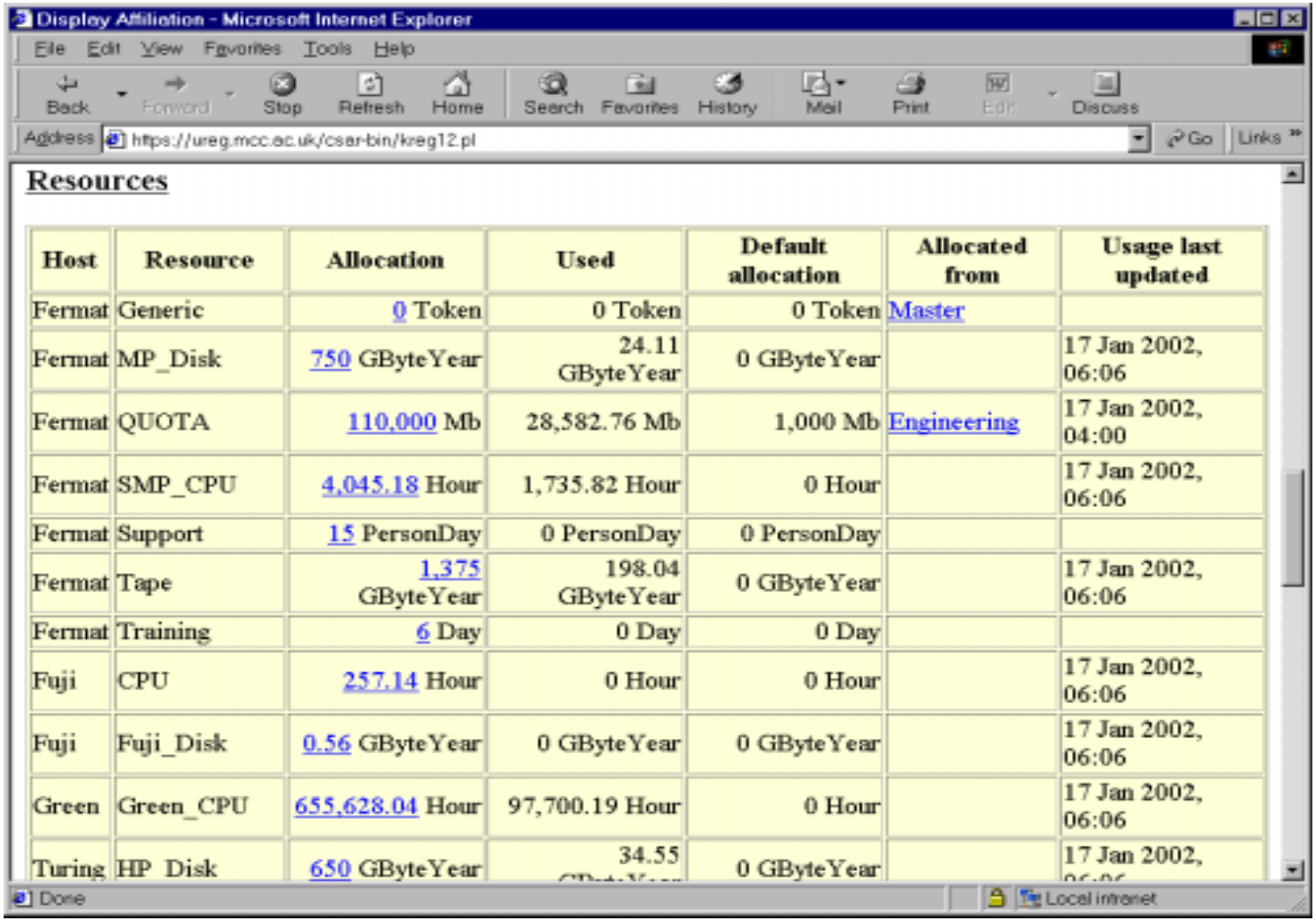
Self Registration



Single person, multiple usernames



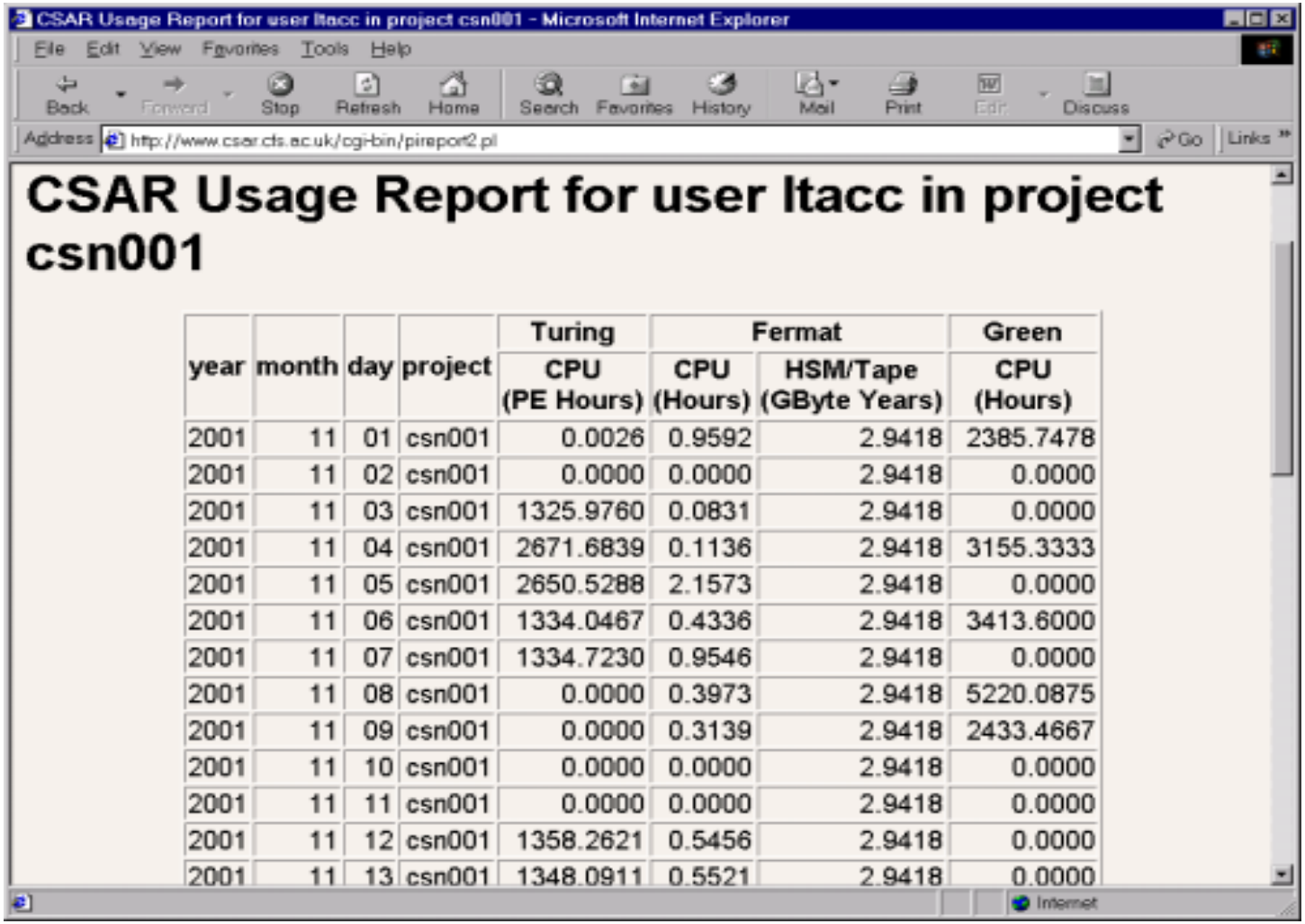
Project Resources - PIs



Resources

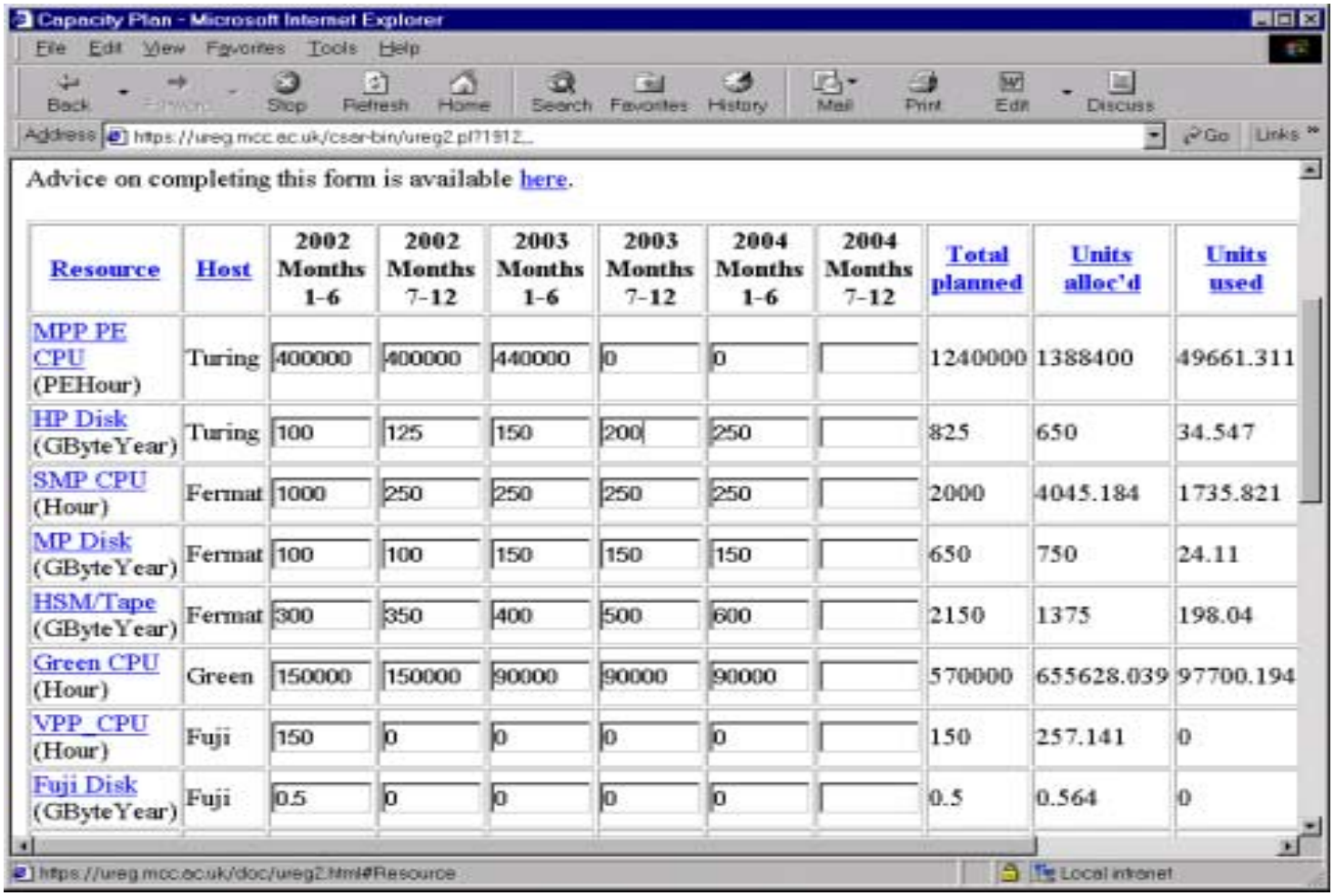
Host	Resource	Allocation	Used	Default allocation	Allocated from	Usage last updated
Fermat	Generic	0 Token	0 Token	0 Token	Master	
Fermat	MP_Disk	750 GByteYear	24.11 GByteYear	0 GByteYear		17 Jan 2002, 06:06
Fermat	QUOTA	110,000 Mb	28,582.76 Mb	1,000 Mb	Engineering	17 Jan 2002, 04:00
Fermat	SMP_CPU	4,045.18 Hour	1,735.82 Hour	0 Hour		17 Jan 2002, 06:06
Fermat	Support	15 PersonDay	0 PersonDay	0 PersonDay		
Fermat	Tape	1,375 GByteYear	198.04 GByteYear	0 GByteYear		17 Jan 2002, 06:06
Fermat	Training	6 Day	0 Day	0 Day		
Fuji	CPU	257.14 Hour	0 Hour	0 Hour		17 Jan 2002, 06:06
Fuji	Fuji_Disk	0.56 GByteYear	0 GByteYear	0 GByteYear		17 Jan 2002, 06:06
Green	Green_CPU	655,628.04 Hour	97,700.19 Hour	0 Hour		17 Jan 2002, 06:06
Turing	HP_Disk	650 GByteYear	34.55	0 GByteYear		17 Jan 2002,

Usage Report - daily



year	month	day	project	Turing CPU (PE Hours)	Fermat CPU (Hours)	Fermat HSM/Tape (GByte Years)	Green CPU (Hours)
2001	11	01	csn001	0.0026	0.9592	2.9418	2385.7478
2001	11	02	csn001	0.0000	0.0000	2.9418	0.0000
2001	11	03	csn001	1325.9760	0.0831	2.9418	0.0000
2001	11	04	csn001	2671.6839	0.1136	2.9418	3155.3333
2001	11	05	csn001	2650.5288	2.1573	2.9418	0.0000
2001	11	06	csn001	1334.0467	0.4336	2.9418	3413.6000
2001	11	07	csn001	1334.7230	0.9546	2.9418	0.0000
2001	11	08	csn001	0.0000	0.3973	2.9418	5220.0875
2001	11	09	csn001	0.0000	0.3139	2.9418	2433.4667
2001	11	10	csn001	0.0000	0.0000	2.9418	0.0000
2001	11	11	csn001	0.0000	0.0000	2.9418	0.0000
2001	11	12	csn001	1358.2621	0.5456	2.9418	0.0000
2001	11	13	csn001	1348.0911	0.5521	2.9418	0.0000

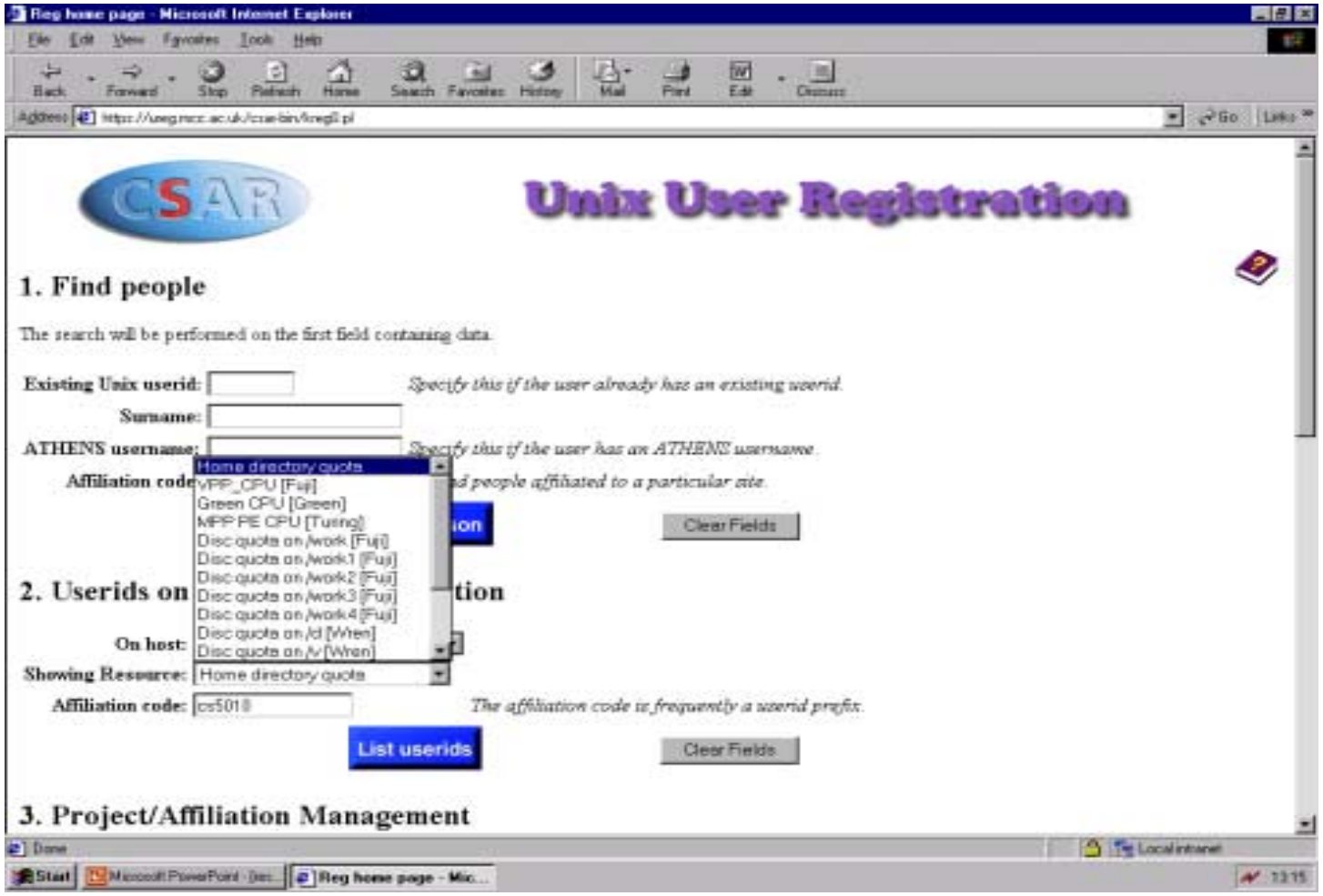
Capacity Plans



Resource	Host	2002 Months 1-6	2002 Months 7-12	2003 Months 1-6	2003 Months 7-12	2004 Months 1-6	2004 Months 7-12	Total planned	Units alloc'd	Units used
MPP PE CPU (PEHour)	Turing	400000	400000	440000	0	0		1240000	1388400	49661.311
HP Disk (GByteYear)	Turing	100	125	150	200	250		825	650	34.547
SMP CPU (Hour)	Fermat	1000	250	250	250	250		2000	4045.184	1735.821
MP Disk (GByteYear)	Fermat	100	100	150	150	150		650	750	24.11
HSM/Tape (GByteYear)	Fermat	300	350	400	500	600		2150	1375	198.04
Green CPU (Hour)	Green	150000	150000	90000	90000	90000		570000	655628.039	97700.194
VPP_CPU (Hour)	Fuji	150	0	0	0	0		150	257.141	0
Fuji Disk (GByteYear)	Fuji	0.5	0	0	0	0		0.5	0.564	0

Additional Reporting – Select Turing

Select Home Directory Quota



Individual quotas on Turing

Userids on a host - Microsoft Internet Explorer

Unix User Registration

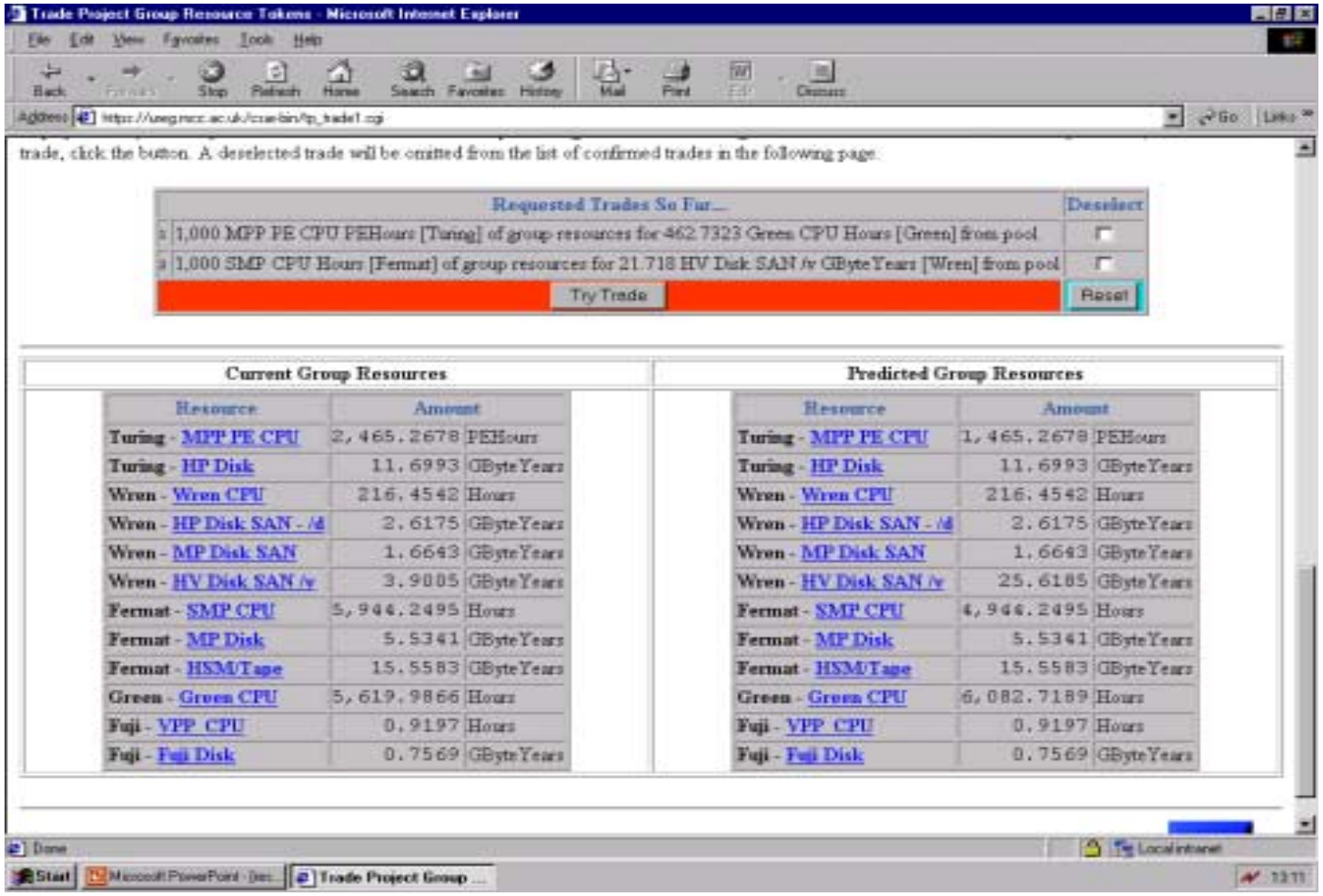
Userids on Turing

cs5010 Apps / Ops

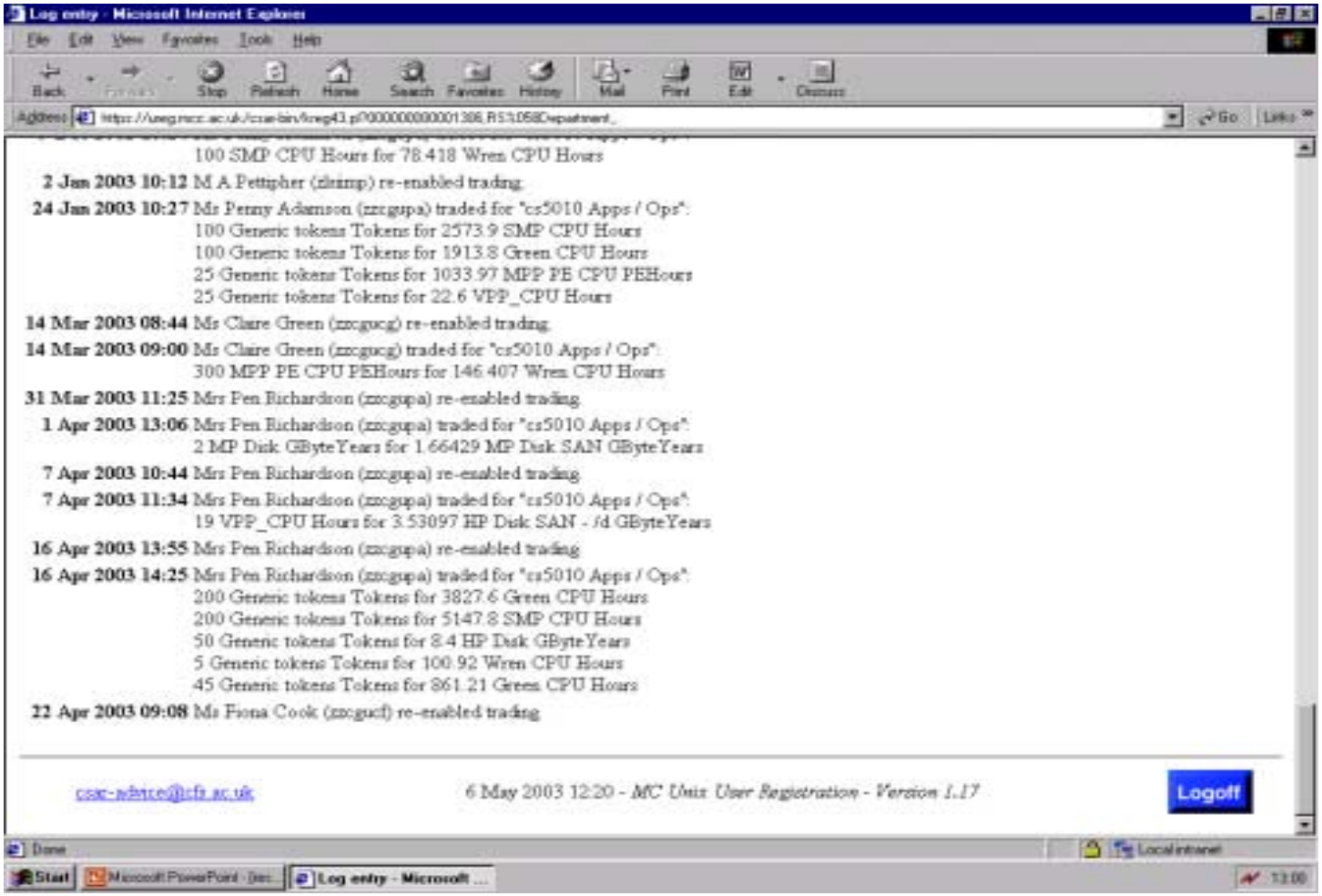
Resource	Allocated		Used		Available for sub-allocation		Available for sub-allocation to users		Usage last updated
MPP PE CPU	26,406.97	PEHour	23,941.7	PEHour	-466.8	PEHour	8,806.97	PEHour	6 May 2003, 06:04
HP Disk	78.3	GByte Year	66.6	GByte Year	10.03	GByte Year	38.3	GByte Year	6 May 2003, 06:04
Disc quota on /ehome	14,000	Mb	3,716.34	Mb	83.95	Mb	83.95	Mb	6 May 2003, 07:00

	Created	Person	Userid	Name	Disc quota on /ehome (Mb)	Usage (Mb)
	15 Apr 2002	Mr Zoe Chaplin	zregzzc	Zoe Chaplin	1,000	994.53
	24 May 2002	Ms Fiona Cook	zregzcf	Cook Fiona	5	0.26
	26 Mar 2001	Mr Mike Daw	zregzmd	Mr Mike Daw	50	30.9
	7 Feb 2002	Mr Rupert Ford	zregzrc	Mr Rupert Ford	50	40.98
	23 Feb 1999	Dr Martyn Foster	mta_poo	Martyn Foster	499	492.08
	28 Feb 2001	Dr Jonathan Gibson	zregzjg	Jonathan Gibson	200	27.33
	18 Jul 2001	Claire Green	zregzcg	Ms Claire Green	50	0.51
	30 Oct 2002	Claire Green-Deletedusers	zregzdel	Claire Green-Deletedusers	1,500	27.34
★★★★★	2 Feb 2001	Miss Kaukab Jaffri	zregzkf	Kaukab Jaffri	1	0.08
	15 Apr 1999		zregzop	Kaukab Jaffri	1	0.08

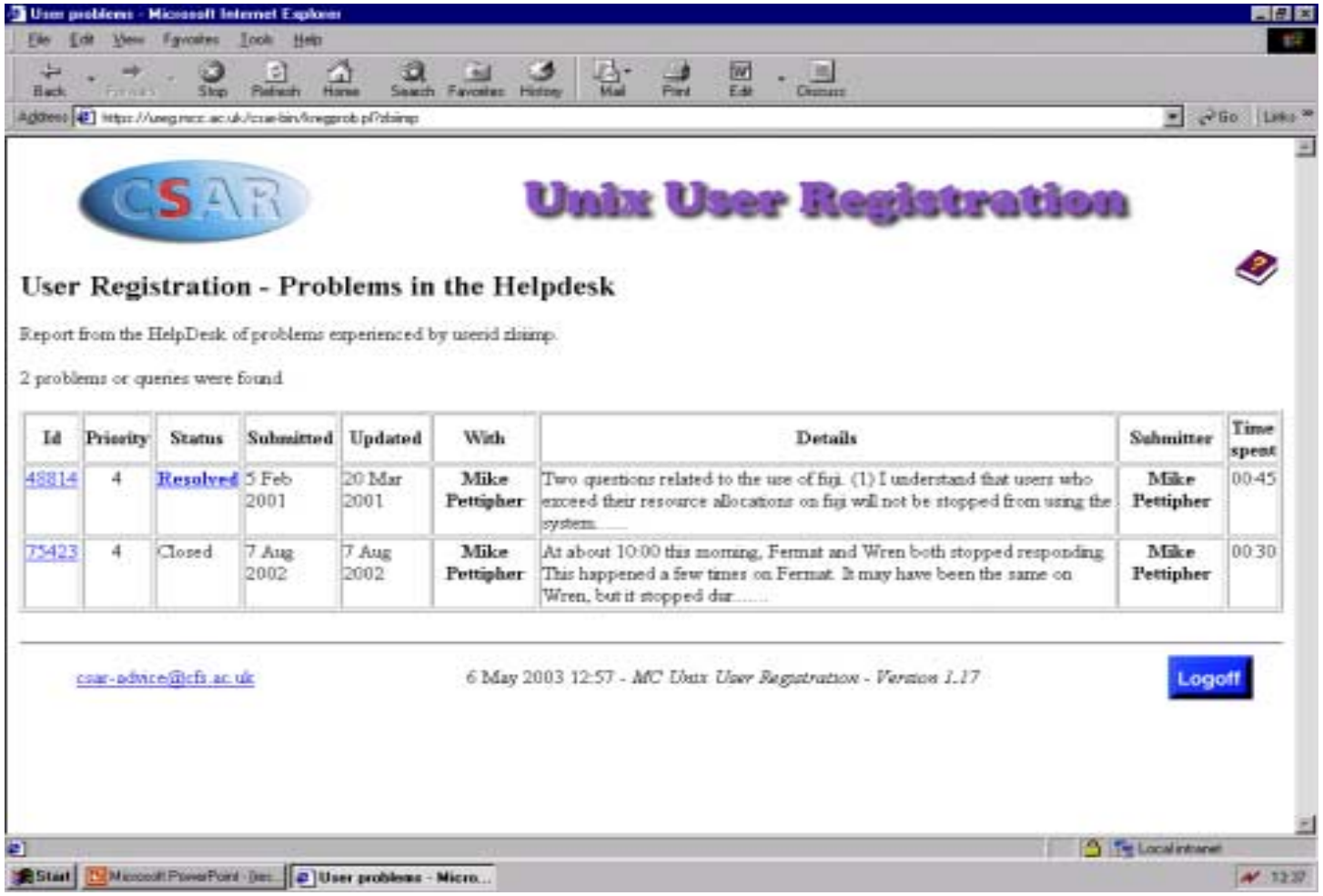
Trading Resources



Transaction logging



Managing Queries



Structure

THE UNIVERSITY of MANCHESTER

Structure of Registration System

- Helpdesk service used for all queries managed by MC, based on Action Remedy System (ARS)
- Underlying database package – Oracle.
- Registration system written in Perl in and built as a layer on top of ARS.
- Executes commands on different systems. Feedback to ensure integrity between database and actual systems.
- Person based – single entry in database for person with (potentially) multiple usernames.
- 4 components:
  - People, affiliation, systems, resources

Manchester Computing
Supercomputing, Visualization & eScience

Future Developments

THE UNIVERSITY of MANCHESTER

Future Developments

- New version under development.
- General enhancements:
  - Structural:
    - Host machines <-> programs <-> OO Library <-> Database access module <-> database
  - Independence from underlying software – ARS and Oracle.
  - Improve performance for transactions (recent move to multithreaded version of Oracle gave substantial improvements.)
  - Role association, as well as personal.
- Merger of the University of Manchester and UMIST (University of Manchester Institute of Science and Technology) – project Unity.
- Specific Grid-motivated enhancements.

Project Unity

- University of Manchester – one of the largest universities in UK.
  - University registration scheme for all students and staff.
  - Unix user registration scheme for a variety of services managed by Manchester Computing.
- UMIST
  - University registration scheme for all students and staff.
  - Another registration scheme for specific services managed by local computing support unit.
- A new registration scheme, based on the MC unix user registration system, is being developed. This will provide the capability to manage all of the separate schemes mentioned above.

UK Grid Environment

- 2002/2005 – significant funding for e-Science projects and centres.
- How can real science exploit the grid?
- e-Science projects:
  - Industrial projects
  - Pilot projects
  - Interdisciplinary Research projects
  - Demonstrator projects
  - International projects

UK e-Science Centres

UK Grid Resources

- Most e-Science centres have committed resources for grid use.
- The National HPC providers – CSAR and HPCX are both committed to supporting activities on the grid, and will make some national resources available.
  - Globus and UNICORE are installed on the CSAR systems.
- Most e-Science pilot projects have been awarded significant resources on the national HPC services (Cray T3E, SGI Origin, IBM, SGI Altix), and have access to other local and e-Science facilities.

Grid Registration – The Challenges

- Potentially thousands rather than hundreds of HPC users, and tens of thousands social science and library users.
- Centres in general do not own resources being made available to e-Science programme, nor do they necessarily have direct administrative control of them.
- Authentication, Authorisation and Accounting are potential administrative nightmares…

Authentication, Authorisation, ...

- A Globus-based Grid user requires, for each system to be used:
  - A username on the system
  - A personal certificate (X 509), issued (e.g) by the UK e-Science certificate authority (operated by the UK Grid Support Centre).
  - An entry in the grid-mapfile (to map the certificate name to the username).
- Typically entries in the grid-mapfile are managed by the system administrator (by default, the grid-mapfile is owned by root). Need to ensure that the user making the request is the same person as owns the private key of the certificate. Note that a user may legitimately possess multiple certificates, issued by different certification authorities.
- Current, informal procedures where grid-mapfile maintainer knows users personally cannot scale for projected numbers of Grid users.
- Similar model exists in UNICORE, with the UNICORE User Data Base (UUDB) replacing the grid-mapfile.
- …leads to complicated and excessive administration.

So ...

- As we delegate username creation (to the PI), it makes sense also to similarly delegate control of the grid-mapfile or UUDB entries.
- Can also add digital certificate authentication, as an alternative means to access the registration system (user imports X 509 certificate into browser).
- Can associate one (or more) certificates with the person that owns the accounts.

Accounting

- Most sites providing resources on the Grid need to know to whom the usage should be charged, which means knowing at least the user and the amount of each resource used.
- In the longer term, we need:
  - A logging infrastructure that records resource usage.
  - A commonly understood means of describing charges.
  - A mechanism to negotiate charges between the consumer and the service provider.
  - A secure payment mechanism.
- We are pursuing these issues through the UK e-Science project ‘A Market for Computational Services’.

UoM/CSAR Resource Management

- The University of Manchester Unix User Registration System (UURS) already contains many of the features required:
  - Delegate authority/administration as far as possible down the tree – to the Principal Investigator and the end user, thus reducing the burden on system administrators.
  - A single personal authentication, for use with many systems.
  - A charging mechanism for all resources, with the ability to trade resources as required.
  - Capacity planning to address future requirements.
  - Modularity to maximise flexibility.

Summary

- University of Manchester Unix User Registration System handles wide range of resource management activities.
- A major objective has been to devolve administrative tasks as much as possible, to reduce the burden on system administrators and to give the users more control.
- New version under development will provide more flexibility and be capable of supporting a wider range of services both for local and national users of systems based at Manchester, and hopefully also in the Grid environment.
- We are open to suggestions …

Manchester Computing
Supercomputing, Visualization & eScience

SVE @ Manchester Computing

World Leading Supercomputing
Service, Support and Research

Bringing Science and Supercomputers Together

www.man.ac.uk/sve
sve@man.ac.uk