



TWGrid: The Grid and e-Science Global Infrastructure in Taiwan

Eric Yen and Simon C. Lin
ASGC, Taiwan

ISGC at Academia Sinica
27 Mar. 2007



Outline

- TWGrid Introduction and Status Update
- Services
- Applications
- Interoperation
- Summary



Introduction



TWGrid Introduction

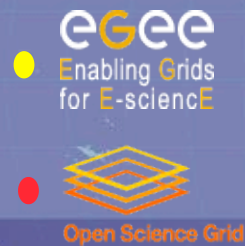
- Consortium Initiated and hosted by ASGC in 2002
- Objectives
 - Gateway to the Global e-Infrastructure & e-Science Applications
 - Providing Asia Pacific Regional Operation Services
 - Fostering e-Science Applications collaboratively in AP
 - Dissemination & Outreach
 - Taiwan Grid/e-Science portal
 - Providing the access point to the services and demonstrate the activities and achievements
 - Integration of Grid Resources of Taiwan
 - VO of general Grid applications in Taiwan





Potential Contributions to the World Wide e-Science/Grid

- Extend the global e-Science infrastructure to AP region
- Reduce the complexity of infrastructure interoperation
- Facilitate the worldwide collaboration by linking the people, data, CPU, instruments globally
- Bridge the digital divide
- Advance essential collaborations of e-Science applications
- Advance the quality of services and applications of worldwide e-Science

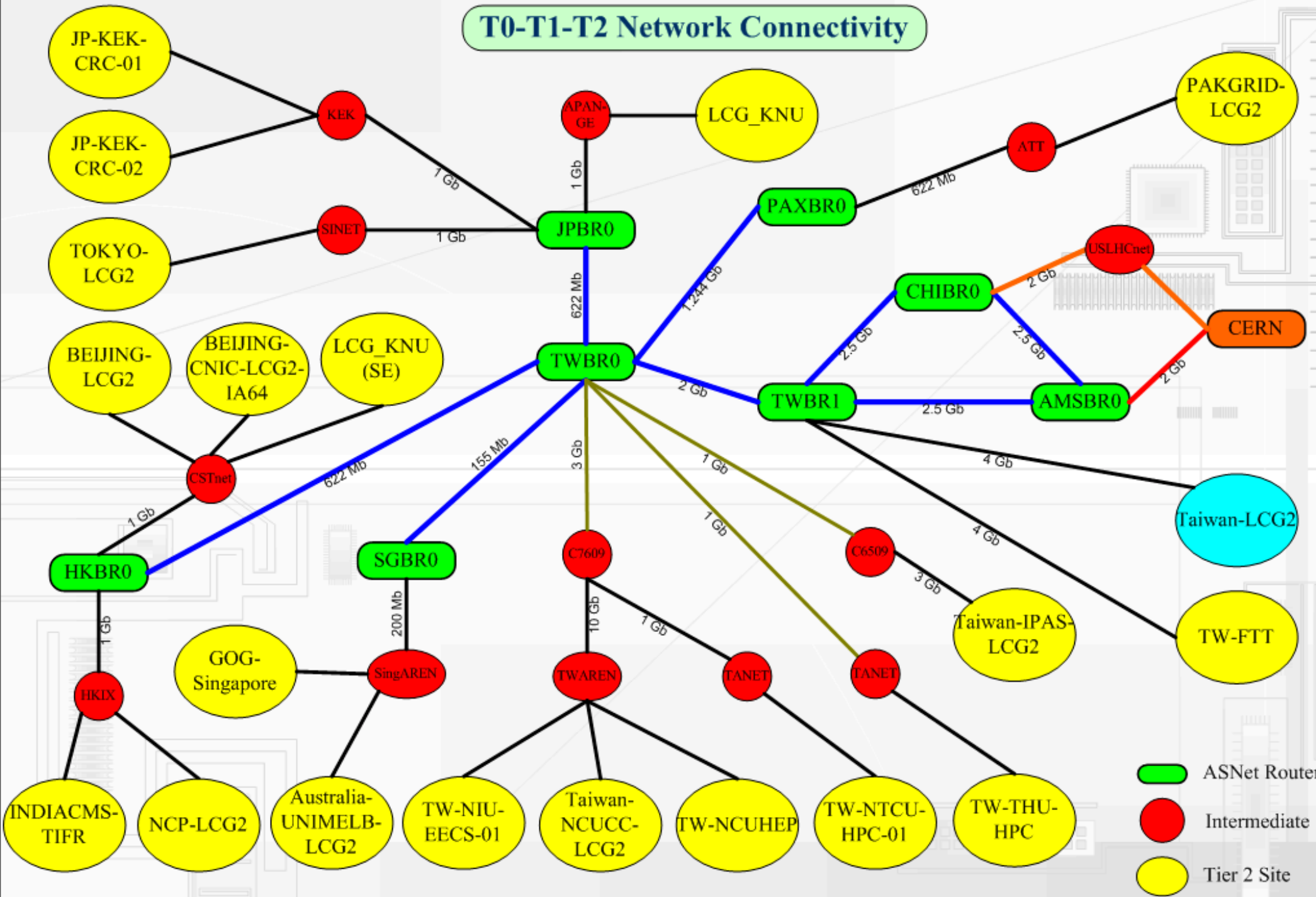




TWGrid: Fostering e-Science Applications by National and Regional Collaboration

- Infrastructure: gLite + OSG
- Status:
 - 8 production sites and 5 sites in certification process
 - 971 CPU, > 450 TB disk and 5 VOs
- Identify Core Services -- common requirements of each application domain
 - Data Management
 - Resource Discovery and Integration
 - Security
 - VO (Role-based rights management and collaboration)
 - Operation & Management
- Foster user communities, such as HEP, Digital Archives, BioMedical, Earth Science & Monitoring, Astronomy, and Humanity and Social Sciences etc.
- Buildup Application Development Framework to reduce the threshold
- Sustainable Services

T0-T1-T2 Network Connectivity



- ASNet Router
- Intermediate
- Tier 2 Site

2006/11/24



TWGrid Services

- Production CA Services: production service from July 2003
- AP CIC/ROC: 20 sites 8 countries, > 1,440 CPUs
- VO Infrastructure Support: APeSci and TWGrid
- WLCG/EGEE Site Registration and Certification
- Middleware and Operation Support
- User Support: APROC Portal (www.twgrid.org/aproc)
- MW and technology development
- Application Development
- Education and Training
- Promotion and Outreach
- Scientific Linux Mirroring and Services



Asia Pacific Regional Operations Center

- **Mission**
 - Provide deployment support facilitating Grid **expansion**
 - Maximize the **availability** of Grid services
- **Supports EGEE sites in Asia Pacific since April 2005**
 - 20 production sites in 8 countries
 - Over 1,470 CPU and 500 TB
- **Runs ASGCCA Certification Authority**
- **Middleware installation support**
- **Production resource center certification**
- **Operations Support**
 - Monitoring
 - Diagnosis and troubleshooting
 - Problem tracking
 - Security

Normalised CPU time [units 1K.SizK.Hours] by SITE and VO

SITE	alice	apdg	atlas	belle	blomed	cms	dteam	g4med	lhcb	ops	ppj	twgrid	Total	%
Australia-UNIMELB-LCG2	0	0	44,302	0	17,586	0	14	0	0	20	0	0	61,922	3.78%
GOG-Singapore	0	0	5,178	0	4,798	1,337	28	0	178	6	0	0	11,525	0.70%
HK-HKU-CC-01	0	0	0	0	0	0	6	0	0	7	0	0	13	0.00%
IN-DAE-VECC-01	3,885	0	0	0	0	0	2	0	0	3	0	0	3,890	0.24%
INDIACMS-TIFR	0	0	0	0	0	53,884	73	0	0	5	0	0	53,962	3.30%
JP-KEK-CRC-01	0	2	0	10,628	0	0	56	5	0	17	0	0	10,708	0.65%
JP-KEK-CRC-02	0	0	737	33,959	0	0	11	0	0	13	0	0	34,720	2.12%
KR-KISTI-GCRT-01	411	0	0	0	0	0	0	0	0	0	0	0	411	0.03%
LCG_KNU	0	0	0	0	0	6,861	1	0	550	2	0	0	7,414	0.45%
NCP-LCG2	1	0	5,960	0	0	12,081	23	0	3,016	4	0	0	21,085	1.29%
PAKGRID-LCG2	1	0	1,084	0	13,198	4,134	10	0	1,127	11	0	0	19,565	1.20%
Taiwan-IPAS-LCG2	0	0	25,474	0	0	0	0	0	0	0	0	0	25,474	1.56%
Taiwan-LCG2	0	0	432,019	2,705	45,328	594,219	19	0	0	21	0	45,553	1,119,864	68.45%
Taiwan-NCUCC-LCG2	0	0	3,141	0	0	5,443	2	0	1,073	2	0	29	9,690	0.59%
TOKYO-LCG2	0	0	205,750	0	0	0	790	0	0	12	0	0	206,552	12.62%
TW-FTT	0	0	12,941	0	0	28,679	7	0	0	0	0	0	41,627	2.54%
TW-NCUHEP	0	0	0	0	0	6,368	7	0	0	0	0	0	6,375	0.39%
TW-NIU-EECS-01	0	0	0	0	0	0	9	0	0	17	0	1,238	1,264	0.08%
TW-NTCU-HPC-01	0	0	0	0	0	0	8	0	0	10	0	0	18	0.00%
TW-THU-HPC	0	0	0	0	0	0	8	0	0	7	0	0	15	0.00%
Total	4,298	2	736,586	47,292	80,910	713,006	1,074	5	5,944	157	0	46,820	1,636,094	
Percentage	0.26%	0.00%	45.02%	2.89%	4.95%	43.58%	0.07%	0.00%	0.36%	0.01%	0.00%	2.86%		



Site Deployment Services

- Deployment consulting
 - Directing to important references
 - Tutorial DVDs (Chinese)
 - Site architecture design
 - Hardware requirements
- Middleware installation support
 - Configuration
 - Troubleshooting
- Site certification
 - Functionality testing
 - Official EGEE infrastructure registration



Operations Support Services

- Operations Support
 - Monitoring
 - Diagnosis and troubleshooting
 - Problem tracking via OTRS ticketing system
- M/W release deployment support
 - Pre-Production site operations
 - Certification testbed
 - Supplementary release notes
- Security Coordination
 - Security release announcement, instructions and follow-up
- Documentation: APROC Portal and wiki
 - <http://www.twgrid.org/aproc>
 - <http://list.grid.sinica.edu.tw/apwiki>
 - **Troubleshooting Guides (New)**
- Site communication and support channels
 - Phone, Email, OTRS Ticketing System
 - Monthly meeting with AsiaPacific sites over VRVS

Asia Pacific ROC
http://www.twgrid.org/aproc

Welcome min tsai (minhong.tsai@gmail.com) 11/19/2005 18:21:43

Home Products Support Online-Support Contact

Logout New Ticket MyTickets CompanyTickets Search FAQ-Area Preferences

Zoom Ticket#: 2005060910000025 Age: 162 days 21 hours
Created: 06/09/2005 20:44:43

State: closed successful
Priority: 3 normal
Queue: CIC/ROC
CustomerID: minhong.tsai@gmail.com

From: min tsai <minhong.tsai@gmail.com>
To: CIC/ROC
Subject: I need some help
Attachment:

problem details...

Subject: I need some help
Text: problem details...

[OTRS] Min-Hong Tsai (min17502@gate.sinica.edu.tw) Sat Nov 19 18:20:42 2005

Logout QueueView Phone-Ticket Email-Ticket Search Preferences Customer Bulk-Action Calendar FileManager WebMail FAQ Admin New message (0) Locked Tickets (1)

[Queue: CIC/ROC]

Tickets shown: 1-5 - Page: 1 - Tickets available: 5 - All tickets: 6
Queues: My Queues (0) - CIC/ROC (12) - CIC/ROC - closed (158) - HPC - closed (39) - PRAGMA (2) - PRAGMA - closed (13) - CASTOR (4) - Ttic1 (3)

[Ticket#: 2005101810000036] [Age: 32 days 4 hours]

Lock - Zoom - History - Priority - Note - Close Created: 11/15/2005 11:55:00

From: APROC Trouble Ticketing System <trs@roc.grid.sinica.edu.tw>
To: Nirranjan <nirranjan@tfr.res.in>
Cc: ASGC-ROC <roc@lists.grid.sinica.edu.tw>
Subject: Re: [Ticket#2005101810000036] [RoC] Re: fail to cr file into your classic SE

Dear Nirranjan,

we also report this prob to CIC, as we can solve this during the past three weeks.

but would like to proceed if you can help switching off the wn030 from DNS.

Thanks.

BR,
J

Thank you for your request.
APROC Trouble Ticketing System <trs@roc.grid.sinica.edu.tw> wrote:

> Dear Nirranjan,
>
> sorry for the late,
[...]

State: open
Priority: 3 normal
Queue: CIC/ROC
CustomerID:
Escalation in: none

Compose Answer (email):
• empty answer

Contact customer (phone):
• phone call

Change queue:
CIC/ROC [Move]



Application Startup

- Initial startup: APESCI VO
 - Provided for new communities to test and develop Grid applications
 - Acts as incubator VO for fast access to Grid resources
 - Centralized services already running
 - Resource Broker, LFC and VOMS services
- Next step: Production VO
 - Discuss with NA4 to join existing VO and collaborate
 - Create a new VO
 - APROC can also help host LFC and VOMS for the new VO



ASGCCA



- Production service since July 2003
 - Member of EUGridPMA and APGridPMA
 - LCG/EGEE users in Asia Pacific without local production CA
 - **AU, China, KEK, Korea, Singapore, India, Pakistan, Malaysia**
- Recent Activities
 - Tickets automatically generated for service request tracking
 - FAQ section added to <http://ca.grid.sinica.edu.tw> to answer common user issues
 - Updated CPCPS defining RA structure
- Registration Authority
 - Permanent staff of organization within LCG/EGEE collaborat
 - Responsibilities
 - Verification of user identification
 - Face-to-face interviews
 - Official ID verification
 - Assist users with certificate registration
 - Archive RA activities for auditing
 - Request revocation





Dissemination & Outreach

- International Symposium on Grid Computing from 2002
- TWGRID Web Portal
- Grid Tutorial, Workshop & User Training: > 750 participants in past 10 events
- Publication
- Grid Café / Chinese (<http://gridcafe.web.cern.ch/gridcafe/>)

Event	Date	Attendant	Venue
China Grid LCG Training	16-18 May 2004	40	Beijing, China
ISGC 2004 Tutorial	26 July 2004	50	AS, Taiwan
Grid Workshop	16-18 Aug. 2004	50	Shang-Dong, China
NTHU	22-23 Dec. 2004	110	Shin-Chu, Taiwan
NCKU	9-10 Mar. 2005	80	Tainan, Taiwan
ISGC 2005 Tutorial	25 Apr. 2005	80	AS, Taiwan
Tung-Hai Univ.	June 2005	100	Tai-chung, Taiwan
EGEE Workshop	Aug. 2005	80	20th APAN, Taiwan
EGEE Administrator Workshop	Mar. 2006	40	AS, Taiwan
EGEE Tutorial and ISGC'06	1 May, 2006	73	AS, Taiwan
EGEE Tutorial with APAN 23	26 Jan. 2007	30	Manila, Philippine
EGEE Tutorial with ISGC'07	26 Mar. 2007	90	AS, Taiwan



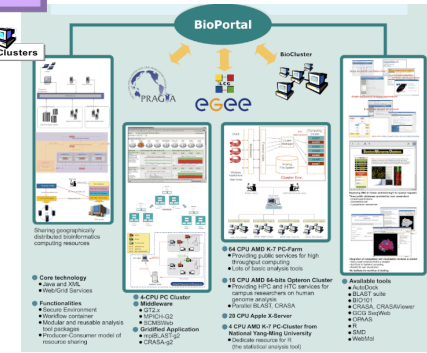
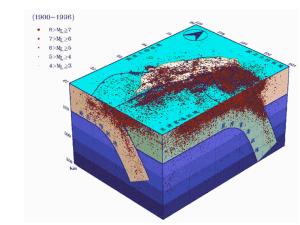
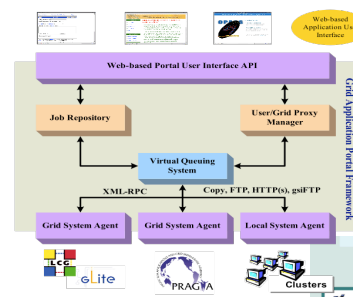
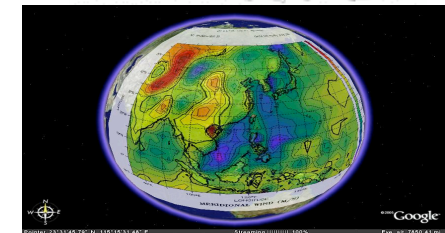
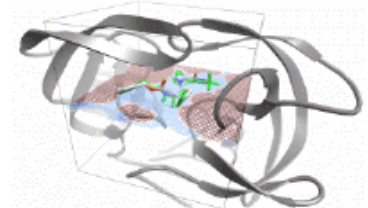
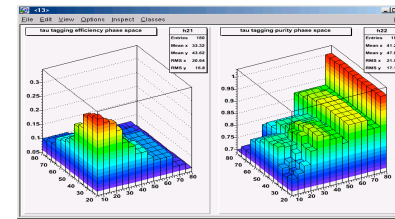


Applications



e-Science Applications in Taiwan

- High Energy Physics: **WLCG, CDF, Belle**
- Bioinformatics: mpiBLAST-g2
- Biomedicine: **Distributing AutoDock tasks on the Grid using DIANE**
- Digital Archive: **Data Grid for Digital Archive Long-term preservation**
- Atmospheric Science
- Earth Sciences: SeisGrid, GeoGrid for data management and hazards mitigation
- Ecology Research and Monitoring: EcoGrid
- BioPortal
- Biodiversity: TaiBIF/GBIF
- Humanity and Social Sciences
- **General HPC Services**
- e-Science Application Development Platform

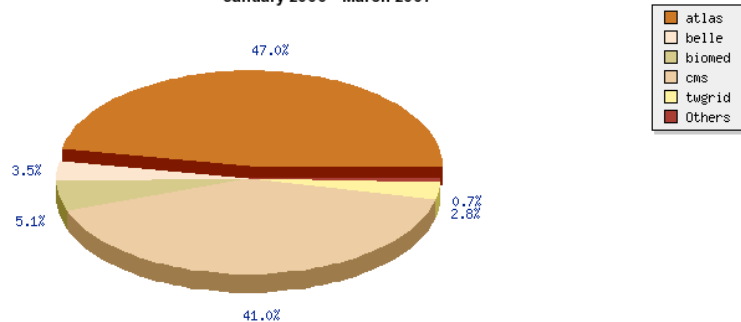


Sites and Applications

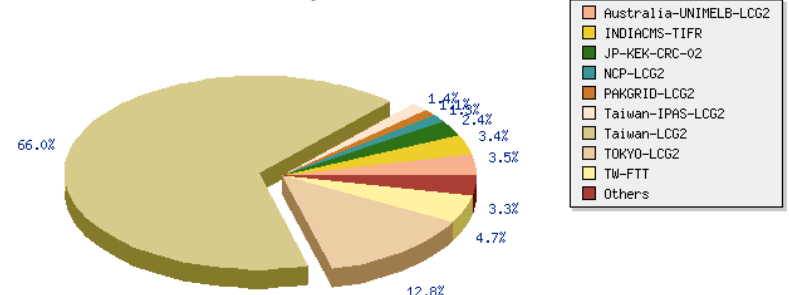
Normalised CPU time [units 1K.SI2K.Hours] by SITE and VO

SITE	alice	apdg	atlas	belle	blomed	cms	dteam	g4med	llc	lhcb	ops	ppj	twgrid	Total	%
Australia-UNIMELB-LCG2	0	0	45,845	0	17,587	0	14	0	0	0	20	0	0	63,466	3.45%
GOG-Singapore	0	0	6,588	0	4,798	1,562	28	0	0	178	6	0	0	13,160	0.72%
HK-HKU-CC-01	0	0	0	0	0	0	7	0	0	0	7	0	0	14	0.00%
IN-DAE-VECC-01	4,447	0	0	0	0	0	3	0	0	0	4	0	0	4,454	0.24%
INDIACMS-TIFR	0	0	0	0	0	63,229	79	0	0	0	5	0	0	63,313	3.45%
JP-KEK-CRC-01	0	2	0	11,660	0	0	58	5	17	0	19	0	0	11,761	0.64%
JP-KEK-CRC-02	0	0	737	43,477	0	0	12	0	0	0	14	0	0	44,240	2.41%
KR-KISTI-GCRT-01	430	0	0	0	0	0	0	0	0	0	0	0	0	430	0.02%
LCG_KNU	0	0	1,048	0	0	8,504	1	0	0	559	2	0	0	10,114	0.55%
NCP-LCG2	2	0	7,378	0	0	13,656	23	0	0	3,146	4	0	0	24,209	1.32%
PAKGRID-LCG2	2	0	1,093	0	13,230	4,435	12	0	0	1,306	13	0	0	20,091	1.09%
Taiwan-IPAS-LCG2	0	0	26,249	0	0	0	20	0	0	0	0	0	0	26,269	1.43%
Taiwan-LCG2	0	0	478,939	8,829	57,702	617,697	35	0	0	0	21	0	49,565	1,212,788	66.02%
Taiwan-NCUCC-LCG2	0	0	3,877	0	0	5,615	3	0	0	1,134	2	0	29	10,660	0.58%
TOKYO-LCG2	0	0	234,406	0	0	0	850	0	0	0	14	0	0	235,270	12.81%
TW-FTT	0	0	57,628	0	0	28,739	7	0	0	0	0	0	0	86,374	4.70%
TW-NCUHEP	0	0	0	0	0	9,111	7	0	0	0	0	0	0	9,118	0.50%
TW-NIU-EECS-01	0	0	0	0	0	0	9	0	0	0	18	0	1,238	1,265	0.07%
TW-NTCU-HPC-01	0	0	0	0	0	0	9	0	0	0	11	0	0	20	0.00%
TW-THU-HPC	0	0	0	0	0	0	8	0	0	0	7	0	0	15	0.00%
Total	4,881	2	863,788	63,966	93,317	752,548	1,185	5	17	6,323	167	0	50,832	1,837,031	
Percentage	0.27%	0.00%	47.02%	3.48%	5.08%	40.97%	0.06%	0.00%	0.00%	0.34%	0.01%	0.00%	2.77%		

AsiaPacific Normalised CPU time per VO
January 2006 - March 2007



AsiaPacific Normalised CPU time per SITE
January 2006 - March 2007





Summary of ASGC T1 Services (I)

- VOBOX/LFC: DDM
- CMS: Phedex/Frontier squid
- FTS
 - Data transfers services within AP
 - T1/T2/T3 data transfer services
- SRM: CASTOR at T1; DPM/dCache at T2
- HA services help improving single point failures of
 - DB → RAC (FTS, C2 catalogue/NS, LFC)
 - CE/RB/WMS → hard backup
 - R/R of BDII (site + top), and FTS.
 - Batch service
- QoS improvement
 - Catalogue service – Oracle backend
 - RR implementation for SRM (currently: C2: 3, and one for C1)
 - Network file system
- 24x7 Op: Service recovery std procedures



Summary of ASGC T1 Services (II)

- Provision of the pledged resources on schedule (by 1st July, 2007)
- Integrated testing with client tools and workflow (Users + T1/T2/T3)
 - Conduct user level testing of Grid services for their experimental researches
- Engage more Tier-2 sites join WLCG testing in all levels more proactively
 - Accounting data will be included in APEL repository and reported monthly no later than April
- ATLAS
 - T0-T1, T1-T1, and T1-T2 data distribution model verification
 - Our data distribution model requires a coupling between Tier-1s
 - BNL \Leftrightarrow IN2P3CC+FZK, NIKHEF/SARA \Leftrightarrow ASGC+TRIUMF+RAL, CNAF \Leftrightarrow RAL, PIC \Leftrightarrow NDGF
 - Build up a Data Management Supporting Framework among ASGC and all the T2s in Asia
 - Data distribution testing and improvement
 - Effective Supporting and debugging mechanism collaboratively



CMS Activities

- **CSA06**

- ttbar production
 - Done
 - CMS produced 5.8M events in total
 - 185K events (~3%) were produced at ASGC
 - All files have been transferred to CERN
- EWK soup production
 - CMS aims at 5M events
(2.6M $W \rightarrow l \nu$ + 2.2M DY + 0.1M $H \rightarrow WW$ + 0.1M WW)
 - 5.7M events have been produced
 - Production @ ASGC : ~255K events (~5%)
 - ~36 K events have been merged at ASGC
 - Merged files have yet to be registered in PHEDEX
- Next : $W \rightarrow e \nu$

- **Load Test Cycle 1**

target kind	target (or eventual sub-targets)	period	ASGC	CNAF	FNAL	FZK	IN2P3	PIC	RAL
single target	65% T0 -> T1 peak rate	1 week	17.1	23.9	68.3	17.1	20.5	6.8	17.1
simultaneous targets	50% T0 -> T1 aver rate	12 hrs	6.6	9.2	26.3	6.6	7.9	2.6	6.6
	50% T1->allT2 sum of aver. rate	12 hrs	38.0	47.0	124.0	31.5	48.0	26.5	42.0
	50% allT2->T1 sum of aver. rate	12 hrs	4.5	5.5	15.0	3.5	7.5	3.0	6.5
simultaneous target	T1 -> each T2 sustain	12 hrs	10	10	10	10	10	10	10
simultaneous target	each T2 -> T1 sustain	12 hrs	5	5	5	5	5	5	5



Atlas - DDM

TIER-1	Datasets Hosted	Datasets Subscribed	Transfer Completed	Transfer in Progress	Datasets Waiting	Last Subscription	LFC File Checked
ASGC	60	1102	401	478	223	Mar 24 06:11:07	Mar 26 10:30:36
BNL	390	794	358	388	48	Mar 24 06:44:35	Mar 26 12:04:42
CERN	78	1093	207	632	254	Mar 24 06:11:07	Mar 26 12:46:53
CNAF	75	1081	218	683	180	Mar 24 06:11:07	Mar 26 12:10:53
FZK	81	1082	484	540	58	Mar 24 06:11:07	Mar 26 12:09:48
LYON	130	1035	321	569	145	Mar 24 06:46:21	Mar 26 12:38:34
NDGFT1	190	968	514	421	33	Mar 24 06:46:25	Mar 26 14:47:12
NIKHEF	51	1122	212	610	300	Mar 24 06:44:40	Mar 26 11:29:10
PIC	70	196	71	125	0	Mar 19 17:10:09	Mar 26 14:49:18
RAL	81	1082	58	193	831	Mar 24 16:41:35	Mar 24 16:42:52
TRIUMF	32	1135	219	609	307	Mar 24 06:11:07	Mar 26 10:04:50



Atlas - DDM

TIER-1	Datasets Hosted	Datasets Subscribed	Transfer Completed	Transfer in Progress	Datasets Waiting	Last Subscription	LFC File Checked
ASGC	60	1102	401	478	223	Mar 24 06:11:07	Mar 26 10:30:36

Activity Summary (Last 24 Hours)

[Click on the cloud name to view list of sites](#)

	Transfers				Errors			
	Cloud	Efficiency	Avg Throughput	Files Done	Datasets Done	Transfer	Local	Remote
LYON	ASGC	80%	16 MB/s	734		180	0	0
NDGFT1	BNL	76%	63 MB/s	3133		977	0	0
NIKHEF	CERN	0%	0 MB/s	0		0	0	0
PIC	CNAF	33%	27 MB/s	1364		2726	0	0
RAL	FZK	75%	33 MB/s	1564		534	0	0
TRIUMF	LYON	74%	40 MB/s	1895		672	0	0
	NDGF	94%	19 MB/s	942		62	0	0
	PIC	75%	29 MB/s	1375		458	0	0
	RAL	0%	0 MB/s	0		0	0	0
	SARA	81%	41 MB/s	1912		457	0	0
	TRIUMF	65%	26 MB/s	1188		639	0	0



FTS – Perf/Stability Test

- **Functional test:** sites tested
 - FTT, IPAS, NCUHEP, NIUCC, THUHEP
 - AU, KEK, KNU, Beijing, Tokyo-LCG2
- **Performance test:** Average throughput MB/s

	AU	Beijing	IPAS	KEK	KNU	NIU	FTT	NCU	Tokyo
Rate	3.2	16.3	36.9	9.8	36.7	4.3	55	8.1	40.2

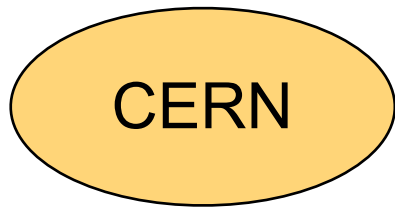
- **Stability tests:** Average throughput MB/s

	AU	IPAS	KNU	Tokyo
Rate	15.2	72.0	28.4	47.7

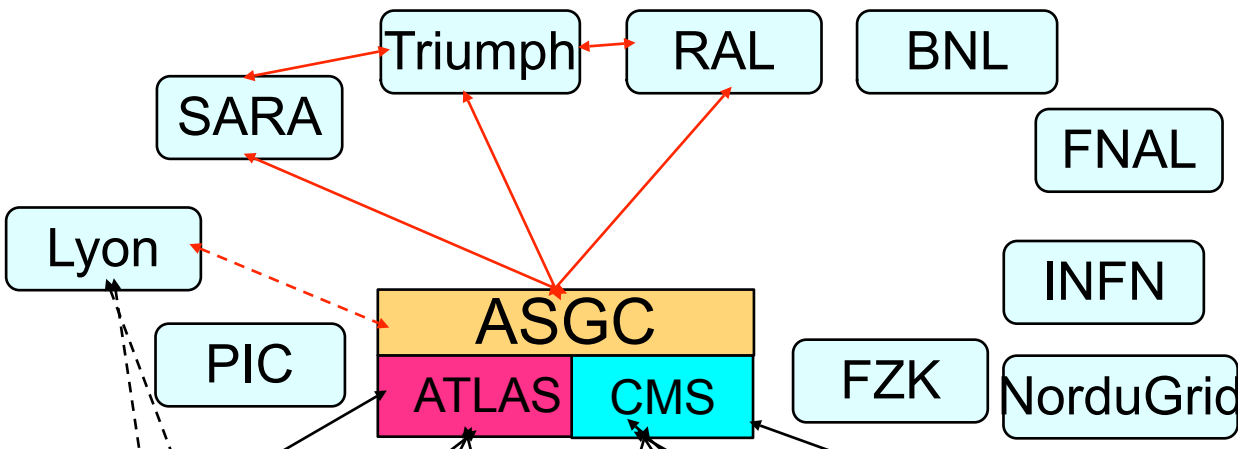
- **T2 → T1 Testing:**
 - KNU 31.8 MB/s
 - IPAS: 82.6 MB/s
- **RC → ASGC**
 - KEK: 14.6 MB/s



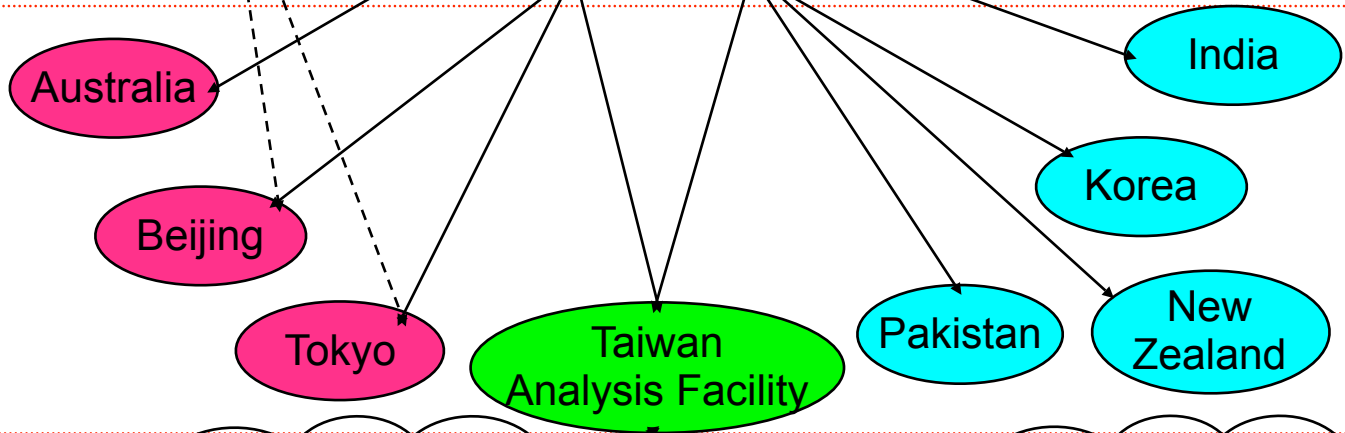
WLCG Architecture in Taiwan



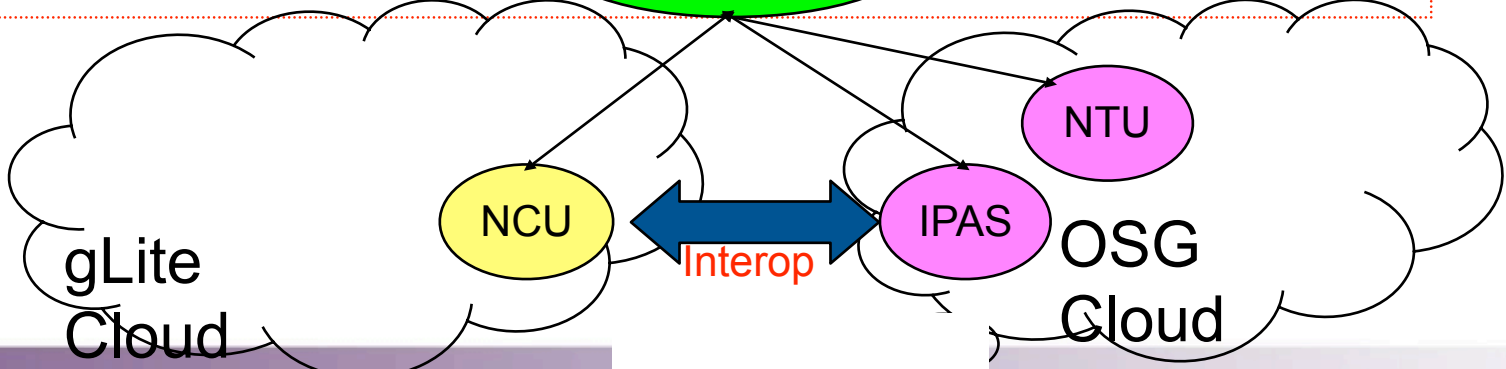
Tier-1



Tier-2s



Tier-3s





ASGC Tier-1 Reliability

- Based on SAM tests on CE, SE and SRM services
- Availability from Nov-2006 to Feb-2007 : **96%**
- One of two sites to reach 88% target
- Still much more effort needed to reach 99%



Site Reliability - WLCG Tier-1s + CERN

The table shows *availability* for May through August, and *reliability* from September on BNL included in average from November
Reliability = Availability/Scheduled_Availability (Scheduled_Availability=(1-Scheduled_Down_Time); tests are not run while scheduled down)

	CERN-PROD	FZK-LCG2	IN2P3-CC	INFN-T1	RAL-LCG2	SARA-MATRIX	TRIUMF-LCG2	Taiwan-LCG2	USCMS-FNAL-WC1	PIC	BNL-LCG2	average reliabilities	8 best sites average		
												availability	reliability	(% target)	
Sep-2006	89%	58%	78%	88%	70%	92%	63%	84%	25%	90%	0%	73%	80%	82%	93%
Oct-2006	96%	54%	85%	85%	77%	74%	80%	67%	55%	84%	27%	76%	80%	81%	92%
Nov-2006	90%	85%	62%	94%	87%	77%	87%	95%	77%	79%	56%	81%	86%	87%	99%
Dec-2006	93%	63%	22%	77%	86%	82%	91%	95%	79%	90%	70%	77%	86%	87%	99%
Jan-2007	99%	85%	96%	75%	80%	93%	79%	96%	84%	86%	90%	87%	91%	91%	103%
Feb-2007	91%	90%	74%	93%	82%	83%	88%	97%	67%	86%	57%	82%	88%	89%	101%
average last three months	94%	79%	65%	82%	82%	86%	86%	96%	77%	87%	73%	83%	--	87%	99%



General HPC Services

- Demonstrated the flexibility of gLite MPI environment, and it's good for embarrassed parallel applications
- Friendly UI in Grid environment
 - Build up a global file system between UI and CE (computing element) can reduce user effort of job submission.
 - Map UI account to real user account of CE to protect user data.
 - Provide a wrapper for job submission. User can submit serial or parallel (via GbE or IB) jobs by it easily without preparing JDL (job description language) file.
 - Chinese and English user guides : http://www.twgrid.org/Service/asgc_hpc/
- Single Sign-on
- Security enhancement by GSI
- Global file system (Keep input and output in home directory)
- Parallel jobs with GbE or parallel with IB jobs via the same script
- Current users are mostly Quantum Monte Carlo and Earth Science users



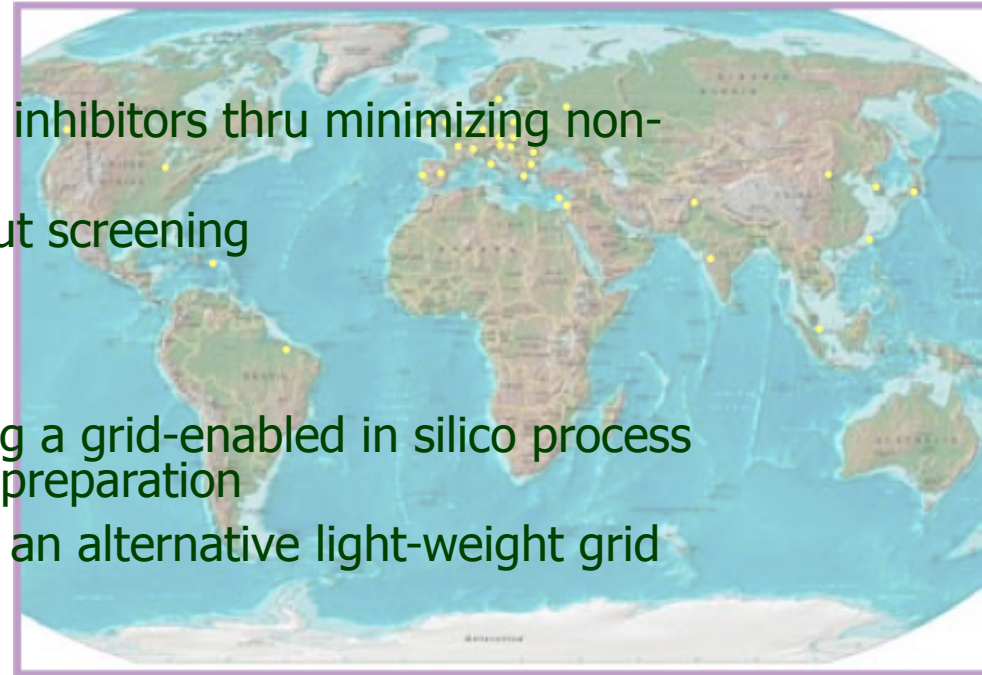
ASGC HPC User Environment

- Supported compiler and library
 - Intel compiler
 - PGI compiler
 - GNU branch for openMP
 - MKL library
 - Atlas
 - FFTW
 - MPICH for Intel, PGI and GNU compiler
- Mellanox version MVAPICH for Intel, PGI and GNU compiler
- Infiniband are deployed for high bandwidth and low latency HPC environment.

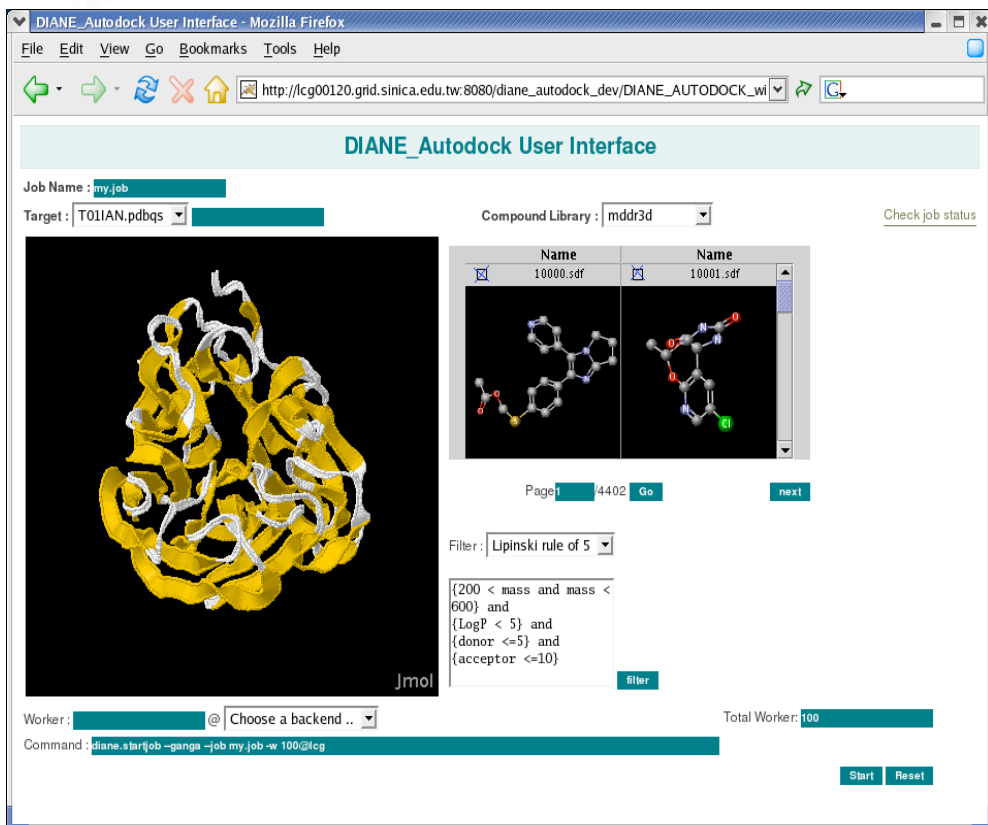


EGEE Biomed DC II – Large Scale Virtual Screening of Drug Design on the Grid

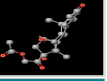



- **Biomedical goal**
 - accelerating the discovery of novel potent inhibitors thru minimizing non-productive trial-and-error approaches
 - improving the efficiency of high throughput screening
 - **Grid goal**
 - aspect of massive throughput: reproducing a grid-enabled in silico process (exercised in DC I) with a shorter time of preparation
 - aspect of interactive feedback: evaluating an alternative light-weight grid application framework (DIANE)
 - **Grid Resources:**
 - AuverGrid, BioinfoGrid, EGEE-II, Embrace, & TWGrid
 - **Problem Size:** around 300 K compounds from ZINC database and a chemical combinatorial library, need ~ 137 CPU-years in 4 weeks
- ⇒ a world-wide infrastructure providing over than 5,000 CPUs

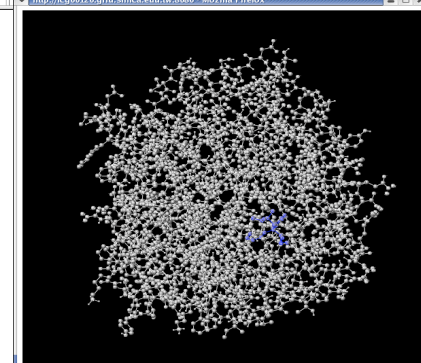
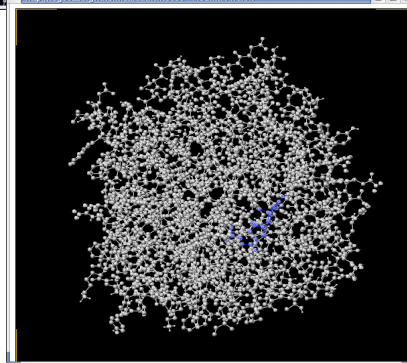


The Interfaces



- job-submission page
- define the docking target and library
- choose filter for database
 - Lipinski rule of 5
 - Lead-likeness
- choose worker and backend

Rank	Molecule	Complex	Run	Group	Docking Energy(kcal/mol)	Binding Energy(kcal/mol)	RMSD(A)	Docking Energy(Group Average)(kcal/mol)	Binding Energy(Group Average)(kcal/mol)	RMSD(Group Average)(A)
1		112.T01IAN	7	1	-12.91	-12.13	67.71	-12.91	-12.13	67.71
2		112.T01IAN	8	2	-12.65	-11.82	68.58	-12.07	-11.40	68.43
3		123.T01IAN	1	1	-12.56	-11.39	68.17	-12.55	-10.91	68.57
4		111.T01IAN	5	2	-11.11	-9.69	74.22	-11.11	-9.69	74.22



- job result page
- show the pose of docking compound and complex structure
- sort by binding energy
- show docking/binding energy and RMSD information
- download structure file of the complex and compound



Lessons Learnt from the 1st DC

- Flexibility and performance of Grid Resources/Services was demonstrated, but
- Lack of a well annotated ligand database:
 - Ligands were selected from variant sources with different indexing schemes.
 - Time consuming to find associated information of each ligands
- Workflow and I/O Issues to the underlying Grid Services
 - Abstraction of Grid filesystem is available but the efficiency and ease-of-use still need to be improved.
 - Search and retrieval the results for analysis should be as easy and efficient as possible
- Friendly Web-based User Interface coping with Application Workflow is required:
 - Biologists prefer an a “virtual” form of traditional in-vitro screening
 - Should be as easy as possible without the knowledge of Grid
- Analysis Pipeline could be further automated:
 - “screening – filtering – screening” cycle approach is used to narrow down the targeted ligands.
 - Screening by distributed docking jobs was implemented very well on Grid, but the pipeline automation and optimization should be taken care as well.

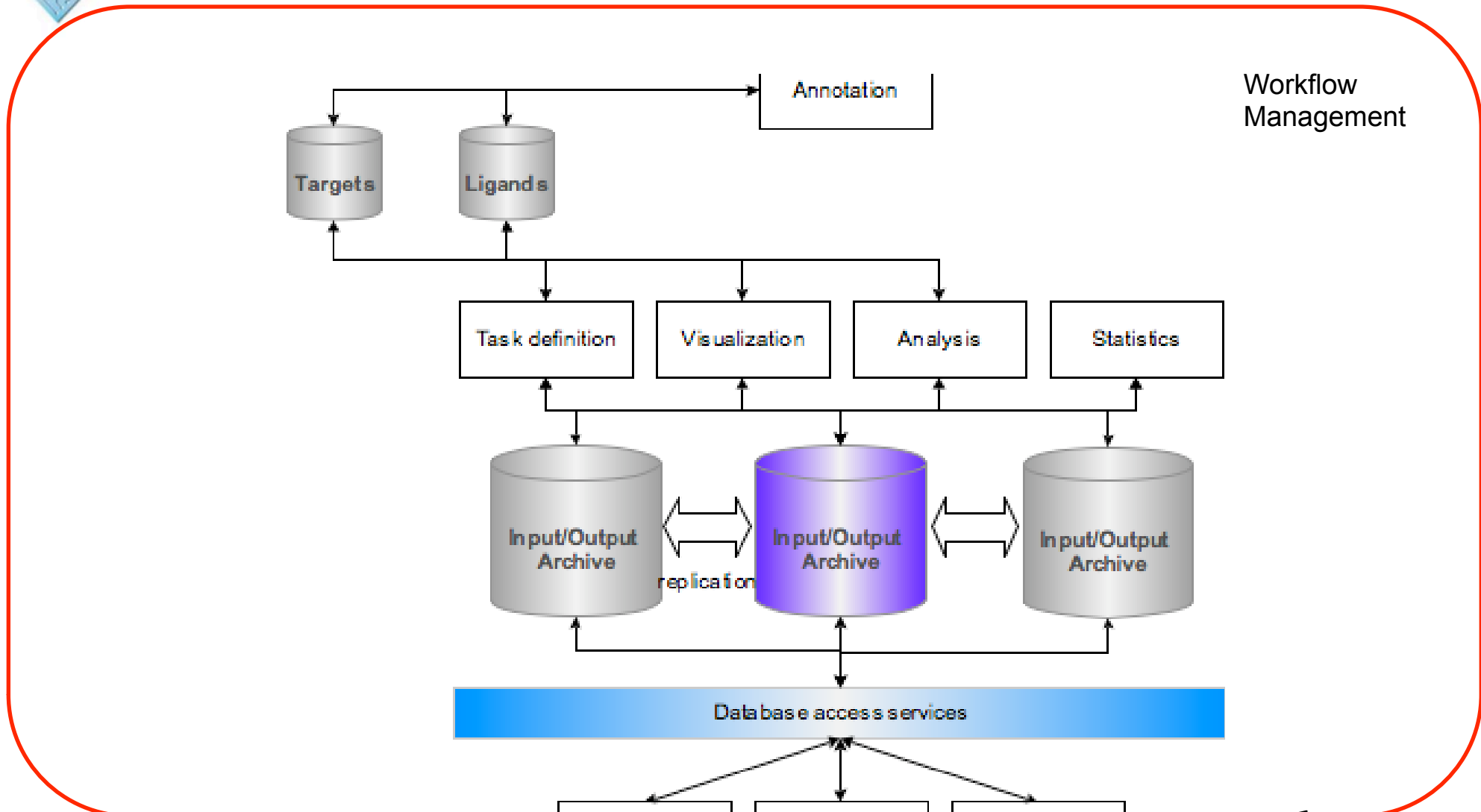


Objectives of DC II

- **Biology:**
 - To further analysis the effect coming from the open form observed by Russell et al and from the variations on the amino acid Try344.
 - To extend the collaboration to wet lab as well
- **Data analysis:** to better represent the virtual screening results and identify the workflow management possibility for overhead reduction.
- **Grid:**
 - To enable the pipeline refinement of virtual screening and GUI enhancement on the Grid
 - To integrate the docking agents (DIANE and WISDOM, etc.) to the Grid Application Platform (GAP) for the full advantage of Grid Services and Heterogeneity



System Architecture





Estimated Resources

- Number of targets: 4 Neuraminidases structures
- Number of ligands: 500,000 chemical compounds
- Estimated elapsed time of each docking in the 1st phase screening: 15 mins
- Estimated size of each dlg file produced by the 1st phase screening: 60 KB
- Estimated elapsed time of each docking in the 2nd phase screening: 30 mins
- Estimated size of each dlg file produced by the 2nd phase screening: 130 KB
- According to the pipeline, the required computing time on an average PC (Xeon 2.8 GHz) will be about: 114 CPU-years
- The total size of the produced docking results will be about: 260 GB



Digital Archives Long-Term Preservation

Objectives



- To conduct Grid-related R&D and integration tasks to help digitize and network the collections and resources of different institutes in NDAP.
- To provide long-term preservation and unified data access services by taking advantage of Grid technology.
- To support the complete information life cycle and persistent values of archives
 - relationship between information sources, history, and provenance
 - Integration with NDAP collection/content Metadata Framework
- These services will be built upon the e-Science infrastructure of Taiwan, by integrating the data management components of the underlying middleware.
- Link the digital archive management tools and applications to take advantage of the Grid infrastructure.

Layered Service Framework



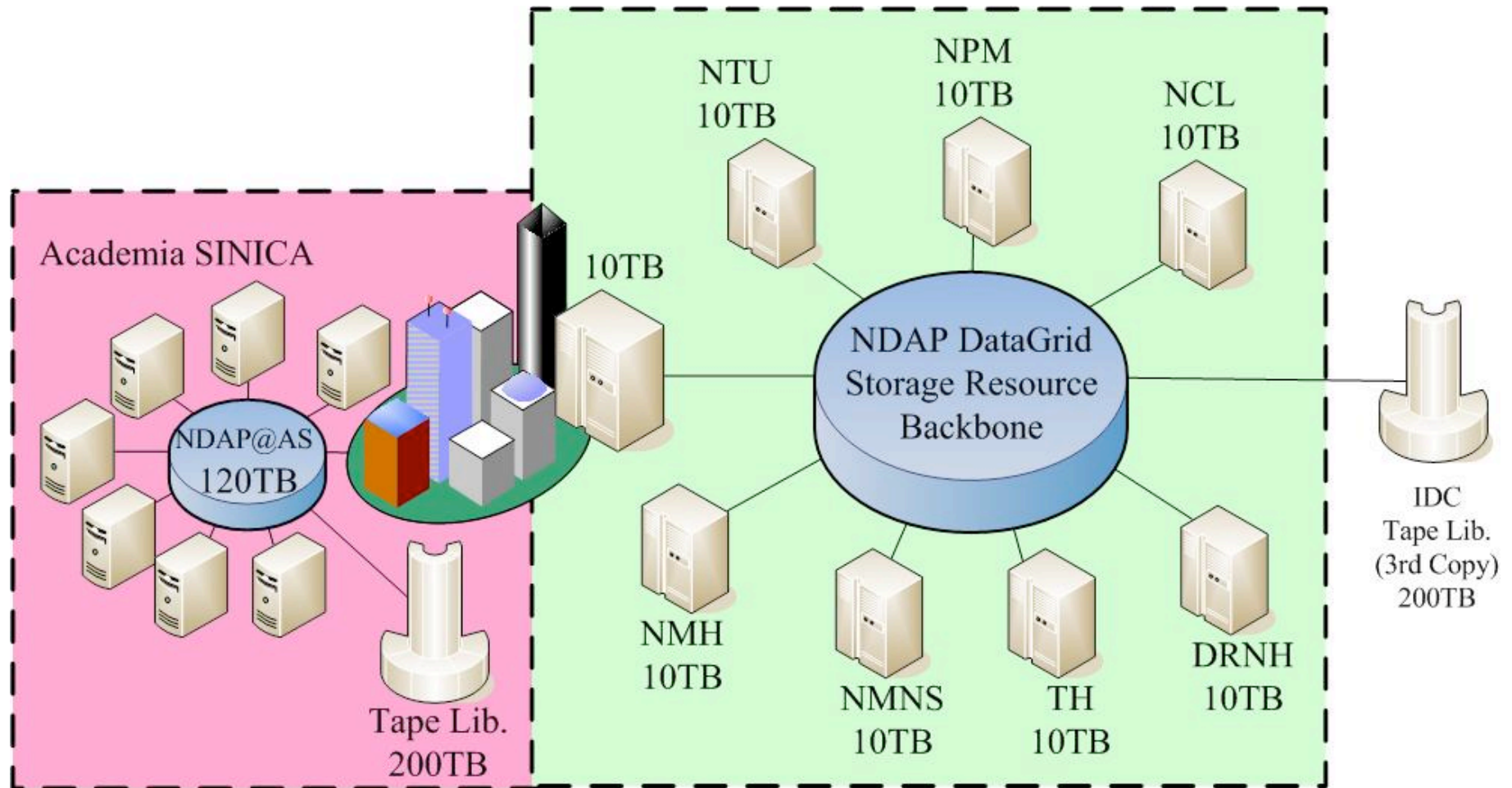
- Customized Application
 - Mediation of heterogeneous Repositories
 - Semantic level information exploration and Knowledge Discovery
- Visualization & Presentation
- Workflow Management
- Distributed Content Management
 - Standardized Digital Object with Metadata
 - Information Retrieval of integrated heterogeneous content sources
 - Federation of distributed resources
- Archive: Long-Term Preservation and efficient access
 - replicated by three remote copies at different sites automatically
 - Secure Access
 - Integration with distributed storage management
 - Uniform name space

Workflow Management



- Optimization of the required services
 - Find Data
 - Registries & Human communication
 - Understand data
 - Metadata description, Standard / familiar formats & representations, Standard value systems & ontologies
 - Data Access
 - Find how to interact with data resource
 - Obtain permission (authority)
 - Make connection
 - Make selection
 - Move Data
 - In bulk or streamed (in increments)
 - Transform Data
 - To format, organisation & representation required for computation or integration
 - Combine data
 - Standard DB operations + operations relevant to the application model
 - Present results

Current Digital Archive DataGrid Architecture in Taiwan



Long-Term Archives for AS NDAP Contents



Table I. Size of Digital Contents of NDAP

	2002	2003	2004	2005	Total
Total Data Size (GB)	22,810.00	38,550.00	63,480.00	70,216.02	195,056.02
AS Production (GB)	22,800.68	31,622.17	47,430.79	55,757.47	157,611.11

Table II. Details of NDAP Production in 2005

	Metadata Size(MB)	Metadata Records	Data Size(GB)
All Inst.	56,204.40	1,035,538.00	70,216.02
AS	53,434.13	763,431.00	55,757.47

User	Project	Total Files	Total Size (Byte)
museum.asmss	珍藏歷史文物	110,235	7,415,558,134,658
srbadm.asmss	管理員	11,096	981,332,124
malacolg.asmss	台灣貝類相	28,077	111,619,505,203
gis.asmss	近代中國歷史地圖與遙測影像資訊典藏計畫	77,734	1,152,358,082,818
la.asmss	語言典藏計畫	1	7,049,563
daal.asmss	技術研發分項計畫	474,262	845,258,179,628
fishdb.asmss	魚類資料庫	32,070	4,199,317,364
ithda.asmss	台史所	121,346	168,408,646,949
muchwood.asmss	台灣本土植物	32,542	1,640,317,250,276
archives.asmss	近代外交經濟重要檔案計畫	642,485	21,926,873,280,602
twnative.asmss	台灣原住民	601,715	1,516,242,052,811
Total		2,131,563	34,781,822,831,996



Grid for Earth Sciences

- SeisGrid (TEC and ASGC)
- GeoGrid (NCKU, NSPO, AIST, ASGC)
- AtmosphereGrid (NCU, NNU, NTU, ASGC)
- GISGrid



SeisGrid

Data Centre

Seismological Resources Integration

- Archiving/ QC/ Links

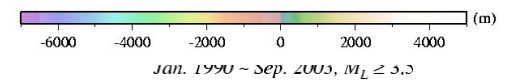
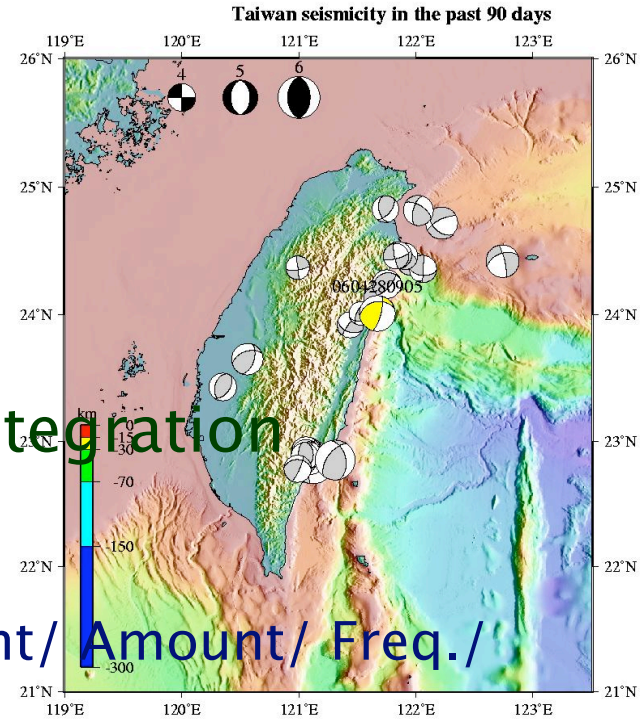
Platform for data access, sharing and integration

- On-line databases
- Utility provider: Software/ Systems/ Scripts
- Requesting Log: Who/ Where/ Time/ Content/ Amount/ Freq./ ...

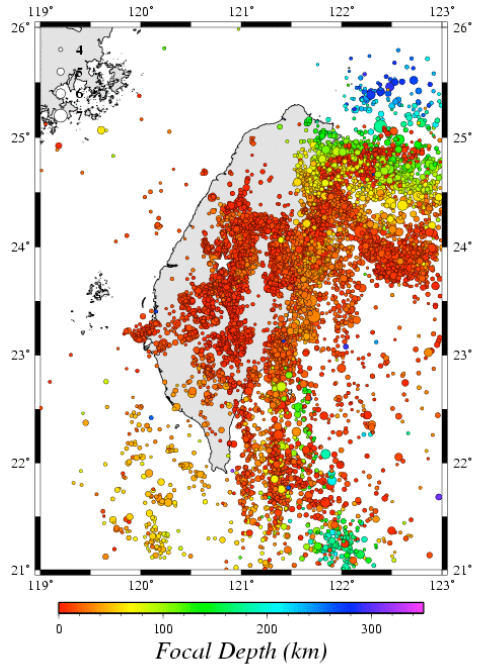
Data Contents

- Seismic Data (with event catalog and station info)
 - Waveform data
 - Parameter data
- Geodetic/ GPS Data
 - Raw/ processed
- Geological Data
- Summary of Seismogenic Structures
- Taiwan Reference Model - Version 0.1

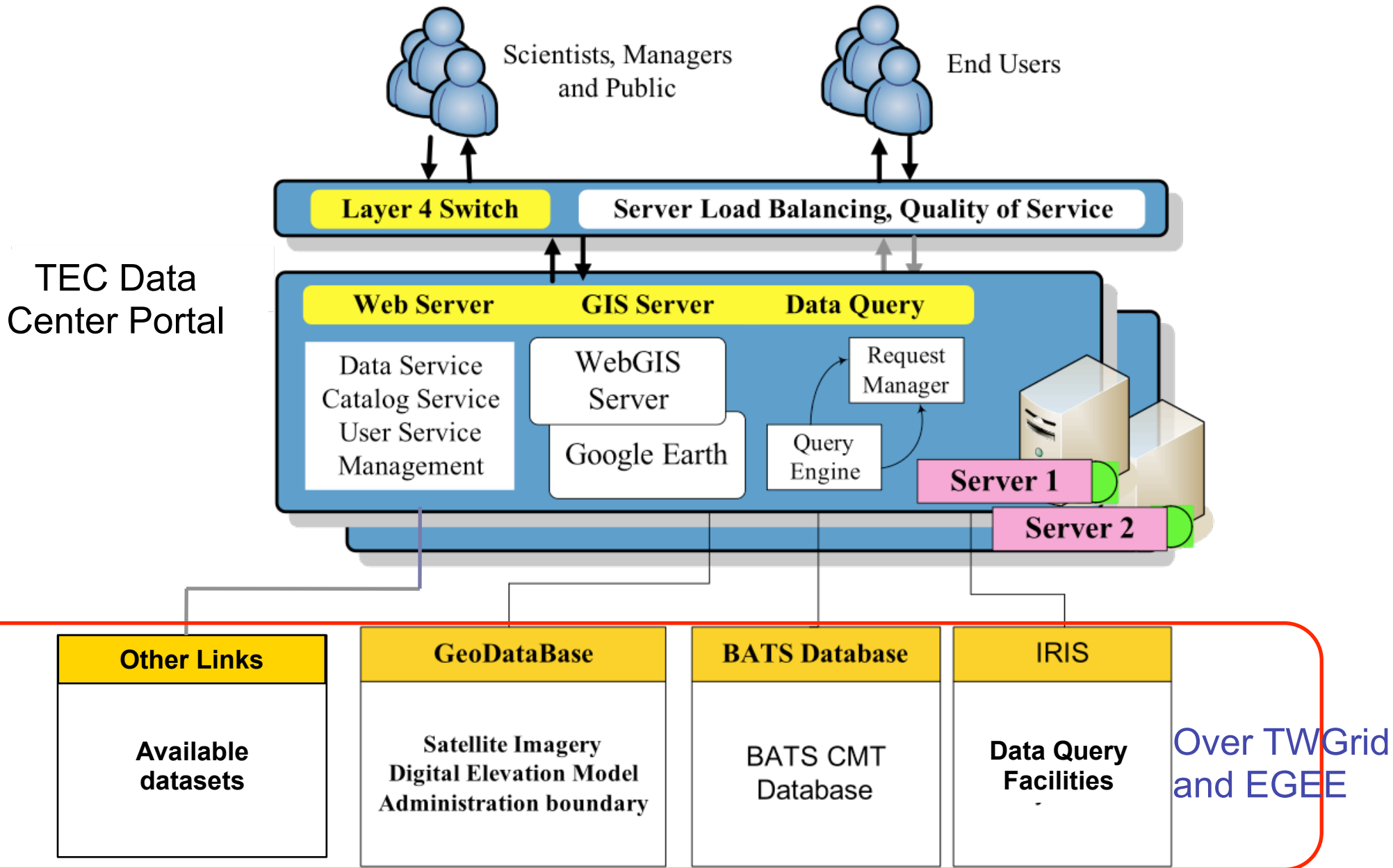
Research and Analysis



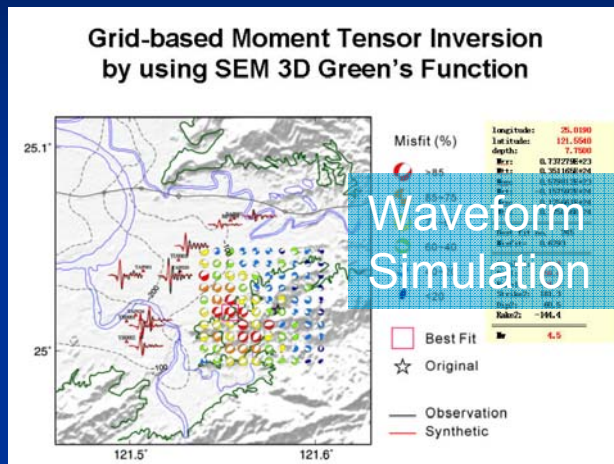
Jan. 1990 ~ Sep. 2005, $M_L \geq 3.5$



TEC Data Center Portal Architecture

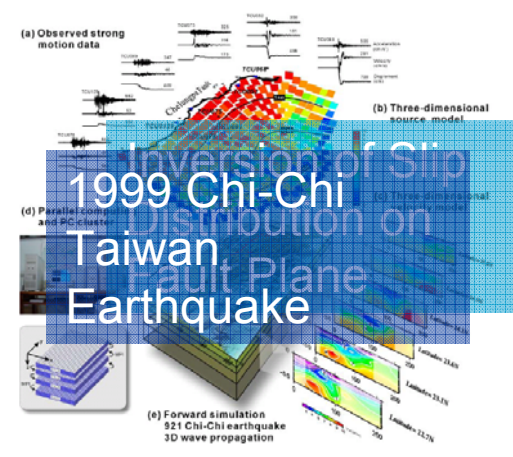
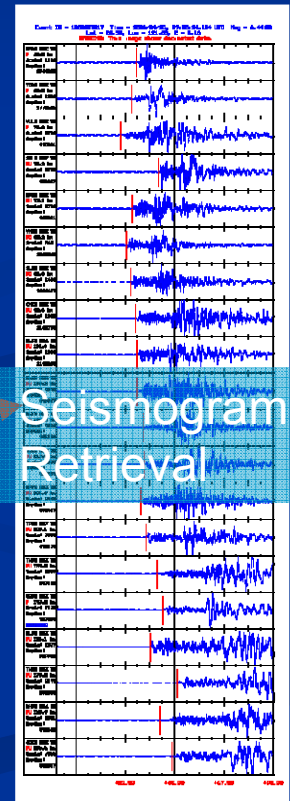
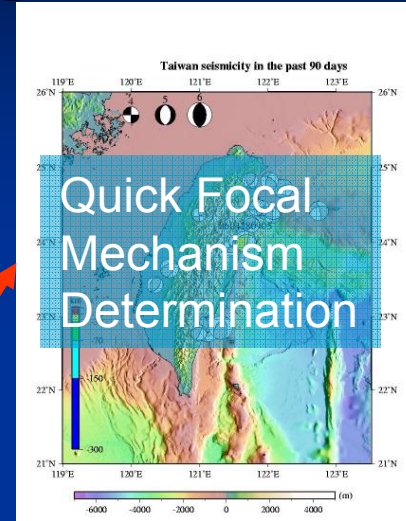


TEC SRB-based Digital Library



Waveform Simulation

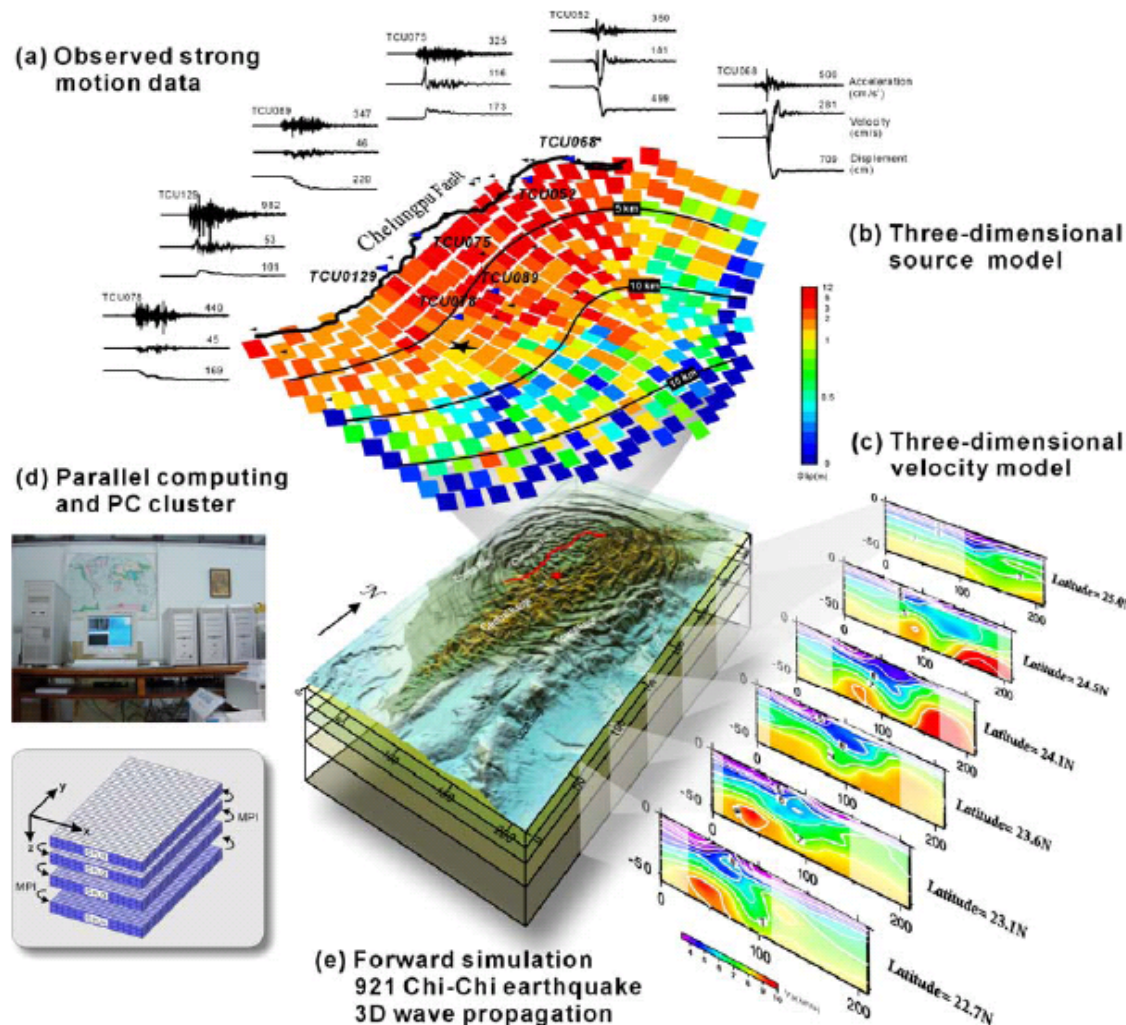
Outputs



SRB-based Digital Library

- 9 Terabytes of on-line disk
- More than 100 Terabytes of tape archive

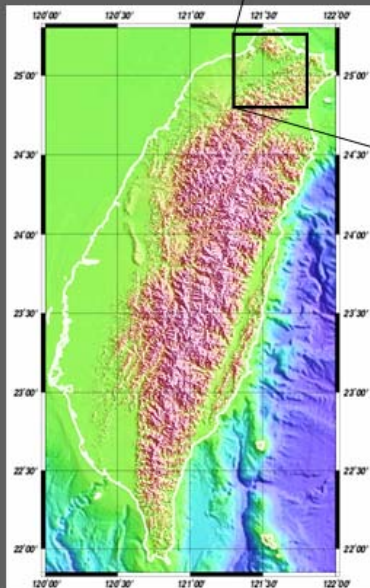
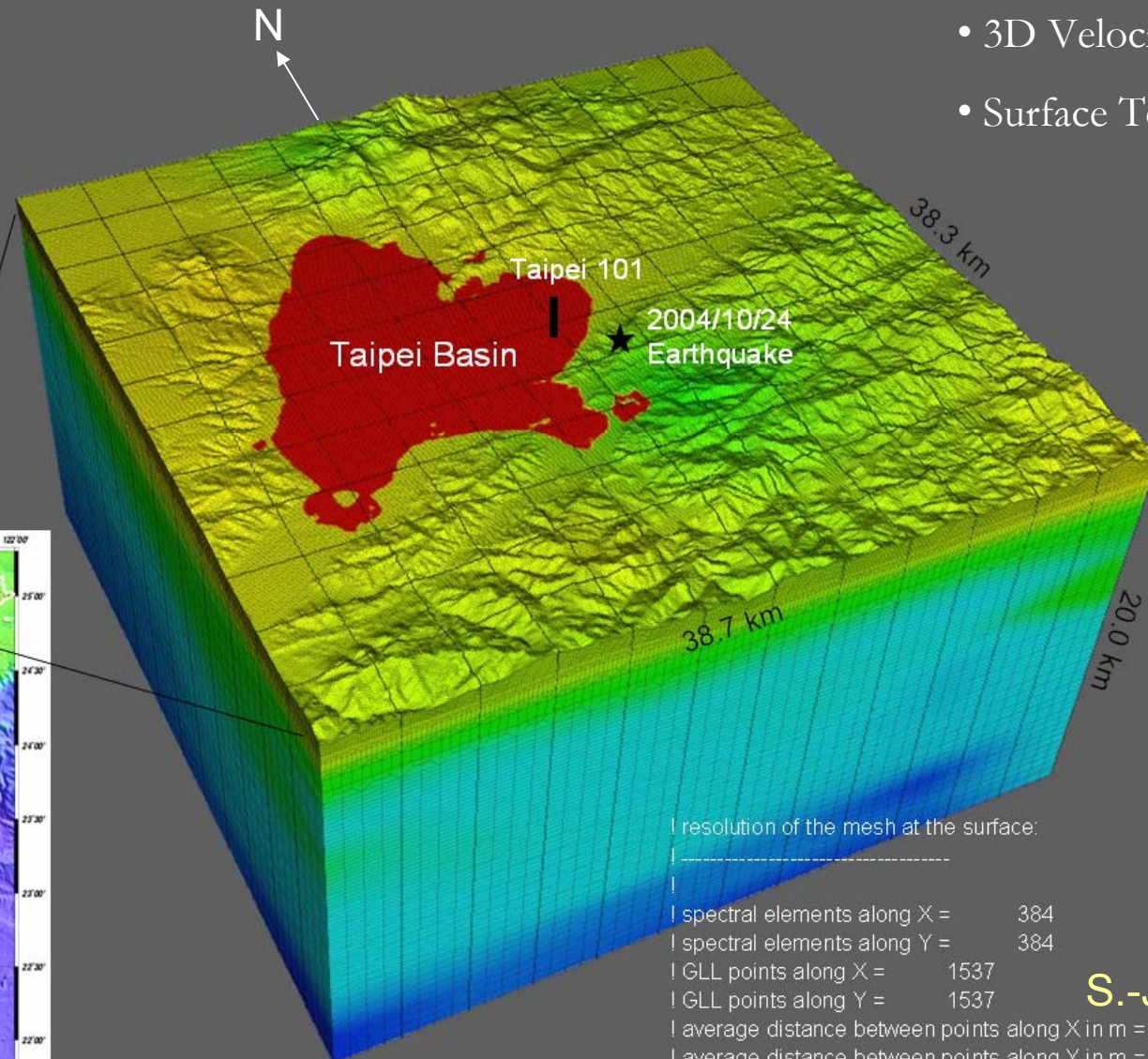
Finite Source Inversion and 3D Wave Propagation



S.-J. Lee, 2005

Taipei Basin Spectral-Element Mesh

- 3D Velocity Structure
- Surface Topography



resolution of the mesh at the surface:

spectral elements along X =	384
spectral elements along Y =	384
GLL points along X =	1537
GLL points along Y =	1537
average distance between points along X in m =	25.23148
average distance between points along Y in m =	24.96805

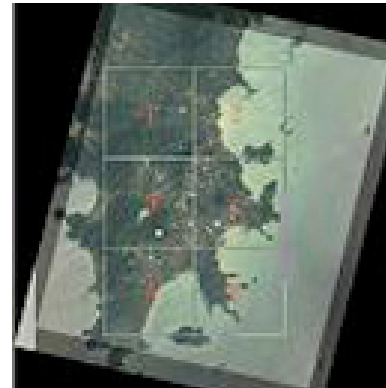
S.-J. Lee, 2006



Taiwan GeoGrid

- Applications

- Grid for Geoscience, Earth Science and Environmental Research and Applications
- Land Use and Natural Resources Plan/Management
- Hazards Mitigation
 - Typhoon
 - Earthquake
 - Flood
 - Coast line changes
 - Landslide/Debris flow



- On-the-fly overlay of base maps and thematic maps,

- from distributed data sources (of variant resolution, types, and time) based on Grid Data Management
- WebGIS/Google Earth based UI
- Integration of Applications with Grid



Grid Application Platform (GAP)

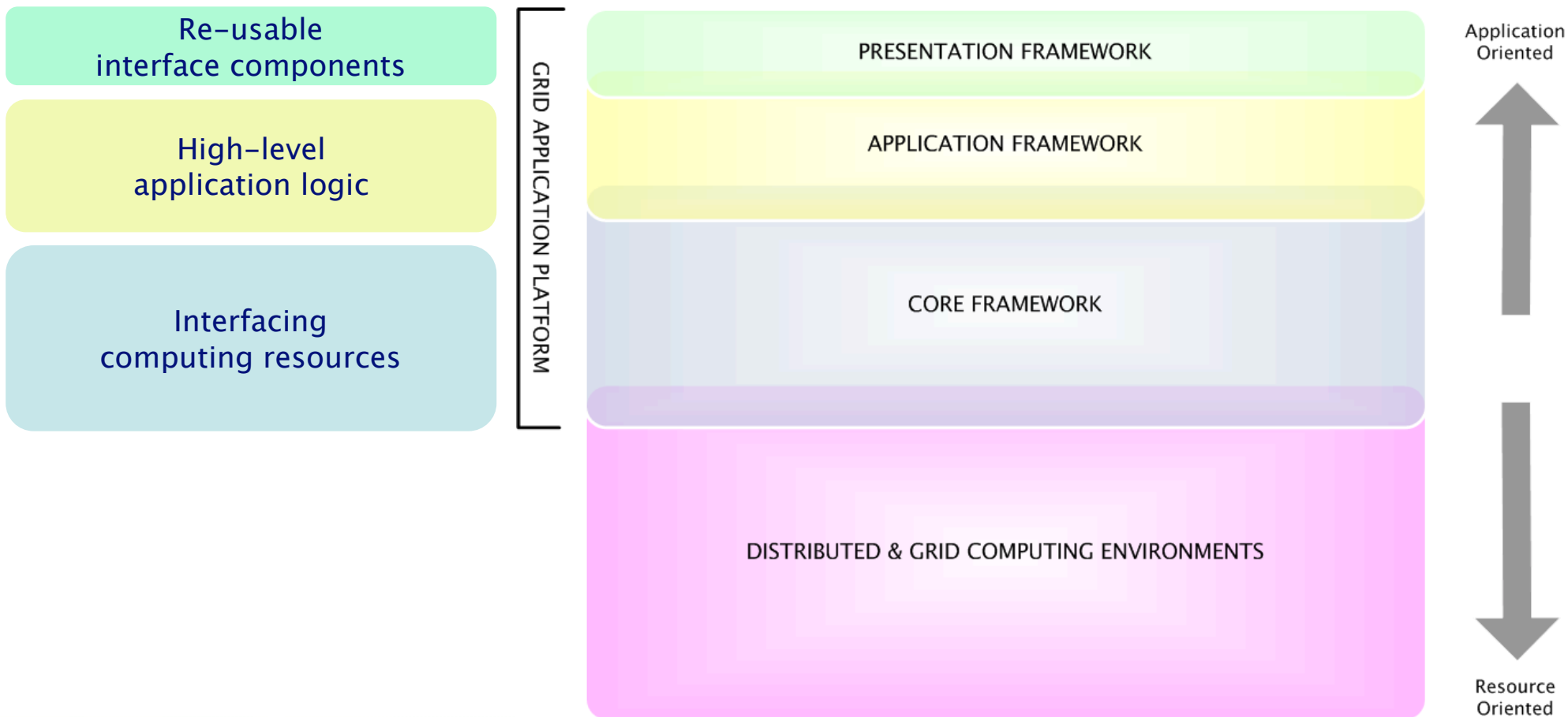


The layered GAP architecture

Reduce the effort of developing application services

Reduce the effort of adapting new technologies

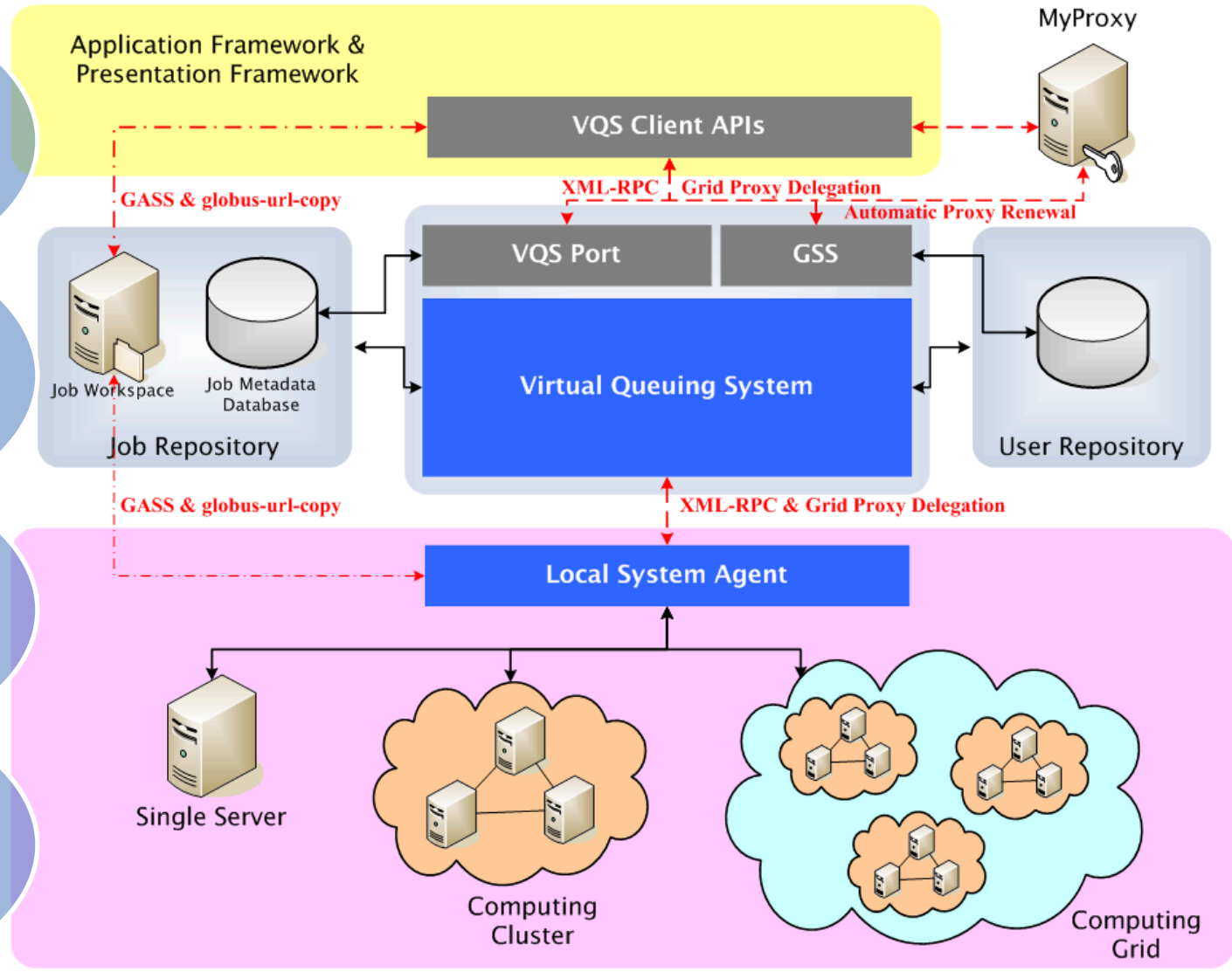
Concentrate efforts on applications





The architecture overview

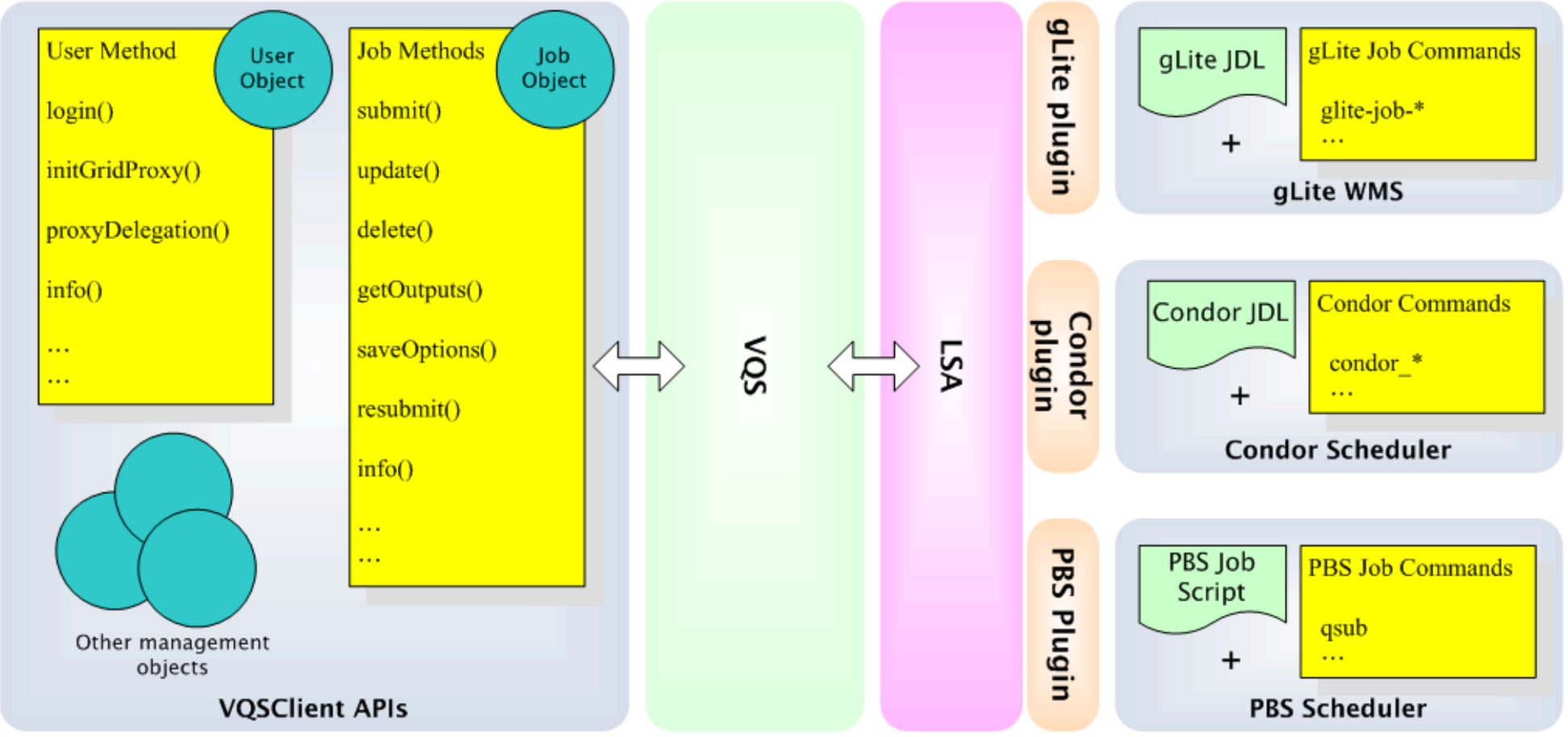
- Portable & light-weight Client
- Common Interface to Heterogeneous Environment
- Multi-user Environment
- Service Oriented Architecture





Common interface to different resources

← common interface → different resources →





Grid Interoperation



Data Management

- Data Interoperation among SRB, gLite and OSG (thru SRM)
- Requirements & Spec : the use cases analysis
 - storage (system/service/space) virtualization
 - automatic replication and version management
 - robust, secure and high performance catalog service
 - reliable, flexible, and quality data transmission
 - Workflow optimization
 - Long-Term Preservation Policy
- Implementation
 - SRM-SRB development
 - based on SRM V 2.2

Challenges

- **Port of SRM interface as client API to a SRB collection**
 - Established as a collaboration
 - “Wayne Schroeder” schroede@sdsc.edu
 - “Wei-Long” wlueng@twgrid.org
 - “Eric Yen” eric@sinica.edu.tw
 - “Ethan Lin” ethanlin@gate.sinica.edu.tw
 - “Abhishek Singh Rana” rana@fnal.gov
- **Wiki created at**
 - <http://www.sdsc.edu/srb/index.php/SRM-SRB>
 - Initial draft document published on high-level approach





Roadmap

- Stage I: ~ end of June 2007
 - API development which are compliant to SRM v2.2
 - SRB-SRM clients will be developed as well
- Stage II: July ~ Sep. 2007
 - Interact and test between data management systems: DPM -- SRB, Castor -- SRB, and dCache -- SRB
- Stage III: Oct. 2007 ~
 - Interoperation with gLite to provide the uniform access interface
 - Develop higher level services for data look-up, data transmission services, etc., based on the user requirements (as FTS, LFC etc.)



Summary

- Application-Driven and Innovative Collaboration are the major drivers to the success of Grid
- Global e-Infrastructure should be composed of all the production Grid systems, whatever it's national, regional or international level -- Grid of Grids
- Asia Pacific Region is of virtuous potential to adopt the e-Infrastructure :
 - More and more Asia countries will deploy Grid system and take part in the e-Science/e-* world
- Easy-of-use still the most essential : friendly interface and workflow support