



## CAS' Scientific Database and its Application System

Dr. YAN, Baoping Principal of SDB Project Computer Network Information Center (CNIC) Chinese Academy of Sciences (CAS) 20th CODATA Conference, Oct.24, 2006, Beijing



## Agenda

- About CAS
- Background of SDB Project
- SDB in 2001- 2005
- SDB in 2006- 2010
- Conclusion



### Chinese Academy of Sciences (CAS)







- Founded on Nov. 1, 1949
- Highest academic institution in natural sciences in China
- Most comprehensive R&D center in natural sciences and high-tech development
- Highest national advisory body in S&T





1949



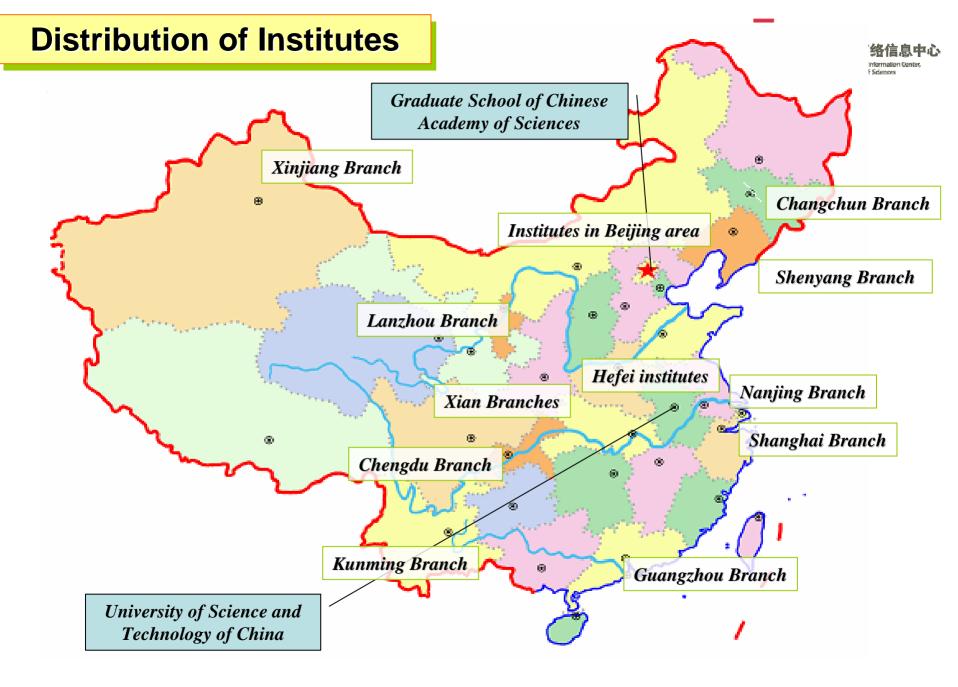


- Target at national strategic needs and world frontiers of science
- Mainly carry out basic and strategic research in an effort to solve major S&T issues of basic, strategic and forward-looking nature in national construction
- Play a key role in the national knowledge innovation system
- Train first-class S&T talents
- Provide scientific bases and tech-innovation sources
- Serve as a national think-tank

### **Research & Development**

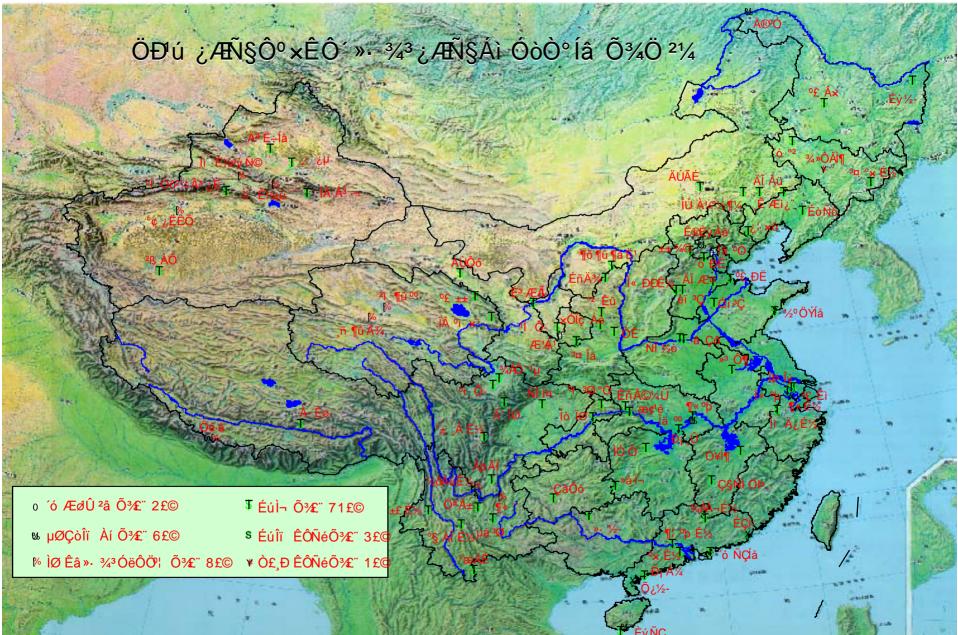


- Total staff: 44,000, of which 13,000 senior and 30,000 other S&T professionals
- Plus 30,000 visiting scholars, post-doctors, and graduates
- 12 branches
- 89 institutes
- Graduate School and USTC
- 9 supporting institutions (tech and docu)
- CAS Holdings Co., 10 major Co. & 490 others



### +200 Wild Field Observatories Distributed





中国科学院 **中国**科学院 **计算机网络信息中心** Conguter Metwork Information Center, Chinese Academy of Sciences

- Nano-materials and nano-devices
- Novel quantum phenomena
- Theoretical biophysics, structural and functional of biomacromolecules and bioinformatics
- Brain and cognitive science
- Complex systems
- Functional materials with new structures
- Physics under extreme conditions
- Molecular sciences and engineering
- Particle physics and evolution of universe

- Physics and chemistry in environmental S&T
- Scientific issues in national security
- Interdisciplinary theoretical studies
- Mathematics and interdisciplinary
- Future information sciences
- Space science and technology
- Future energy
- Interior earth and evolution of life in earth
- Large-scale scientific facilities and application of multi-subjects



### **Biomedical sciences**

- System Biology
- Neuroscience
- Brain Function and Cognition
- Reproduction and Development
- Mechanism of Main Diseases
- Immunity and Infection
- Metabolism and Nutrition
- Diagnosis Technique
- Drug Discovery
- Modernization of Traditional Chinese Medicine

### **Industrial Biotech**

- Bio-energy
- Biobased Chemicals
- Biomaterials
- Environmental Biotechnology
- Enzymes, Lipids and Glycose Biology

### **Agricultural Biology and Biotech**

- Crop Design
- Cloning
- Agricultural Functional Genome
- Agricultural Pest Management
- Marine Biotechnology
- Agricultural Resource Management
- Soil Monitoring
- Regional Agriculture

### **Integrated Biology**

- Taxonomy
- Biodiversity
- Ecology
- Global Change Biology
- Conservation Biology
- Gene and Germplasm Bank
- National Botany Garden System

### **Priorities in Resources and Environment**



- Basic theory and key tech for oil, gas and mines
- Lithosphere evolution
- Qinghai-Tibetan Plateau
- Geo-engineering technologies
- Water resources
- Costal marine ecosystems
- Deep sea environment and life
   process
- Ocean, continent and atmosphere interaction in Asian monsoon

- Earth system model
- Ecosystem functions
- Biodiversity
- Lake pollution and remediation
- Environment and health
- Eco-environmental effects of key engineering
- Remote sensing monitoring of resources and environment
- Global change

### **Priorities in High-tech R&D**



#### **Information Technology**

- High performance computing
- High performance processor
- Micro electro-mechanical systems
- Wireless sensor network
- Next generation internet
- Information security
- Cognition and computational intelligence
- Quantum information

### **Space Science and Technology**

- Scientific application on the National Spaceflight Program
- Lunar exploration
- Mini and micro satellites
- Space remote sensing
- Geospace environment research and space weather

### Energy

- Coal based co-production
- Clean coal technology
- Biomass energy
- Solar energy and wind energy
- Hydrogen energy and Fuel cell

### **Material and Chemical Engineering**

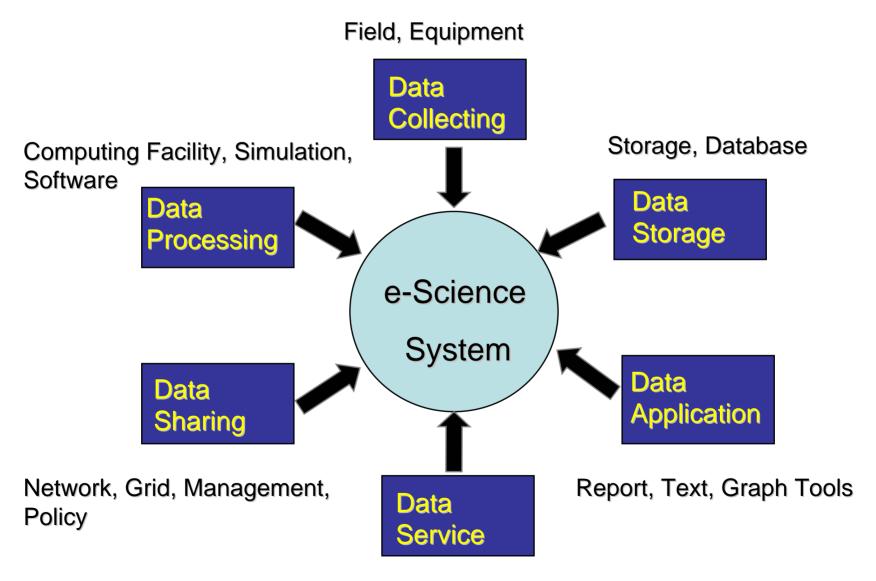
- Green production
- Immobilization and utilization of CO<sub>2</sub>
- Natural gas conversion
- High performance metallic material
- Advanced non-organic material
- Environment-friendly material
- Bio-material and medical material
- Material designing and computational simulation





## SDB in 2001-2005





Search & Retrieval, Content Management



## Scientific Database (SDB)

- Data is the one of the foundational elements in e-Science
- data from research, for research, drive e-Science
- SDB is a long-term project since 1982, in which there are multi-disciplinary scientific data accumulated through the course of science activities in CAS
- many institutes involved, long-term, large-scale collaboration



- In 1970s, some chemical institutes under CAS began to build specialized databases
- A large quantity of valuable scientific data have been produced during the long course of research activities at CAS
- In 1982, CAS initiated the idea for establishing "Scientific Database and its Application System"
- In 1986, CAS formally started the construction of SDB, 20th Anniversary this year

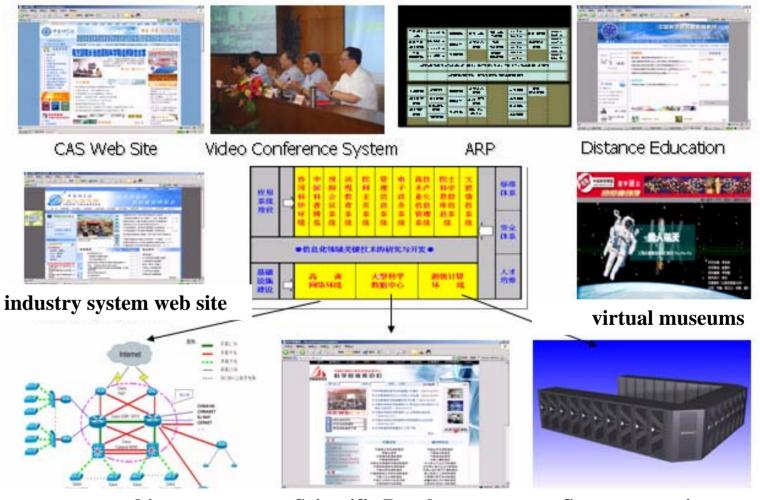
## Funding



As a collection of large-scale, multi-discipline, distributed, scientific databases, SDB is:

- Key engineering project of State Planning Commission (1986-1995)
- Key project of Chinese Academy of Sciences (1986-1990)
- Major project of network application of Natural Science Foundation of China (1995-1996)
- Basic research special support project of Chinese Academy of Sciences (1991-2000)
- Key-project of the 10<sup>th</sup> five-year planning for information construction of CAS (2001-2005)
- Key engineering Project of National Scientific Data Sharing of MOST (2004-2005)
- Key-project of the 11<sup>th</sup> five-year planning for information construction of CAS (2006-2010)

# CAS Informatization Programs Handle Construction Programs 2001-2005



networking

**Scientific Database** 

Supercomputing



Infrastructure	Item	By 2000	By 2005
	core	1Gbps	2.5Gbps- 10Gbps
Networking	backbone	2Mbps	2.5G-155Mbps
	Oversea link	55Mbps	620Mbps+15Gb ps
	Peak TFLOPS	0.13	5.5
HPC	Linpack TFLOPS	0.05	4.3
	Storage	2.1TB	180TB(Disk+Ta)
	Member institutes	21	>45
Scientific Database	Databases	180	+503
	Data volume	725GB	+16.6TB



## Milestones(2001-2005)

- In 2000, the Scientific Database (SDB) project renewed fund by CAS 10th Five-year Program
- In March 2001, proposed "Scientific Data Grid"
- In October 2002, SDG joined the China National Grid (fund from MOST)
- In Nov 2003, SDG Middleware v1.0 released
- In July 2004, SDG got fund from NSFC
- In Sep 2004, SDG renewed fund from MOST
- In Oct 2004, DeepComp 6800 for SDG installed
- In Nov 2004, SDG Middleware v2.0 released
- In Aug 2005, SDG Middleware v2.1 released
- Now, we're working for SDG in 11<sup>th</sup> Five-year Program 2006-2010

中国自然景观图

## SDB status

- 45 institutes across 16 cities503 databases
- 16.6TB total volume

Wulumuqi

Dalian ang

Harbin

Changehun

Shenyang

)ingdao

Fuzhou

Changsha

Guangzhou

Guiyang

Kunming

Xishuangbanna

njing Shanghai



- Six main tasks:
  - Database Resource
  - Data & Database Specification
  - IT Infrastructure Constructuring
  - Middle ware Platform Scientific Data Grid (SDG) Developing
  - SDB & SDG Service
  - Pilot Applications



## 1.Database Resource

- 45 Institutes and hundreds of researchers have participated in the construction of SDB.
- Data Volume: 16TB+
- The Number of Database:500 +
- Database Content covers Physics, Chemistry, Geosciences, biosciences, Ocean Science, Energy Science, Material Science, Astronomy, Space Science and etc.

## Database list (1)



Database Name	Number of sub- DB	Developing Institute
Bio-scien	ce(156)	
Chinese Botany Database	10	Institute of Botany
The Database of Tropical and Subtropical Botany	4	South China Botanical Garden
Plants Database of Southwest of China	10	Kunming Institute of Botany
Chinese Pictorial Flora Database	7	Wuhan Botanical Garden
Database of Zoological Science of China	5	Institute of Zoology
The Animal Resource Database of Southwest of China	10	Kunming Institute of Zoology
Microbial Resource Database of China	81	Institute of Microbiology
Database of Virus Resource	5	Wuhan Institute of Virus
Database of Hydrobiology in China	11	Institute of Hydrobiology
China Database of Nucleic Acid and Protein	2	Shanghai Institute of Biological Sciences
Bioinformatics Database	2	Beijing Genomics Institute
The Rice Genome Database	3	National Center for Gene Research
The Database of E-Clone-Genes and Genome Polymorphism	6	Institute of Biophysics

## Database list (2)



Database Name	Number of sub-DB	Developing Institute
Geosc	<mark>iences(151</mark>	)
Database of China Geotectonics	4	Changsha Intitute of Geotectonics
China Remote Sensing Satellite Image Catalogue System	1	China Remote Sensing Satellite Ground Station
Scientific Database of Subtropical Agri-ecology	4	Institute of Subtropical Agriculture
Mountain Environment & Hazards Database of China	3	Institute of Mountain hazards and Environment
Lake Database of China	2	Nanjing Institute of geography and Limnology
Soil Database of China	7	Nanjing Institute of Soil Science
Natural Resource Database of China	55	Institute of Geographical Sciences and Natural Resource Research
Data Library of Dynamic Geodesy and Resource Environment	2	Institute of Geodesy and Geophysics
The Mineralogy, Petrology and GeoChemistry Database of China	9	Guangzhou Institute of Geochemistry
Database of Soil and Water Conservation in the Loess Plateau	7	Institute of Soil and Water Conservation

## Database list (3)



Database Name	Number of sub- DB	Developing Institute
Geosc	<mark>iences(15</mark>	1)
Xinjiang Ecological and Environmental Resources Database	4	Xinjiang Institute of Ecology and Geography
Database of Chinese Resources and Environment Remote Sensing	11	Institute of Remote Sensing Application
China Wetland Database	11	Northeast Institute of Geography and Agricultural Ecolgoy
Cold and Arid Region Characteristic Database	18	Cold and Arid Regions Environmental & Engineering Research Institute
Agricultural Ecology Database for China's Northeast Black Land	5	Northeast Institute of Geography and Agricultural Ecolgoy
Paleontological and Stratigraphic Database of China	4	Nanjing Institute of Geology and Palaeonotology
Atmospheric Science Database	4	Institute of Atmospheric Physics
Ocean	Science(10	)
China Marine Science Database	3	Institute of Oceanology
Ocean Database of South China Sea	7	The South China Sea Institute of Oceanology

## Database list (4)



Database Name	Number of sub-DB	Developing Institute
Physics an	d Chemisti	ry(45)
High Energy Physics and Related Subject Database	2	Institute of High Energy Physics
Optical System Database	6	Changchun Institute of Optics, Fine Mechanics and Physics
Professional Chemical Database	19	Shanghai Institute of Organic Chemisty
Engineering Chemistry Database	8	Institute of Process Engineering
Database of Applied Chemistry	4	Changchun Institute of Applied Chemistry
理化性能及分析数据库	3	大连化物所
Chemical Substance Toxicity Database	3	Computer Network Information Center
Astronomy &	z Space Scier	nce(12)
Astronomical Database	8	National Astronomical Observatories
Space Environmental Database	4	Center for Space Science and Applied Research

## Database list (5)



Database Name	Number of sub-	Developing Institute
Energy	Science(17	)
China Energy Database	5	Energy Research Institute
New Energy Source & Environmental Protection Database	7	Guangzhou Institute of Energy Conversion
Gas Hydrate Database	5	CNIC and etc.
Materia	al Science(12	2)
Materials Database	7	Institute of Mental Research
Nano Science and Nano Technology Fundamental Database	5	Nano Science and Technology Center
0	thers(1)	
Specialists of Chinese Academy of Sciences	1	



## 2.Data & Database Specification and Standard

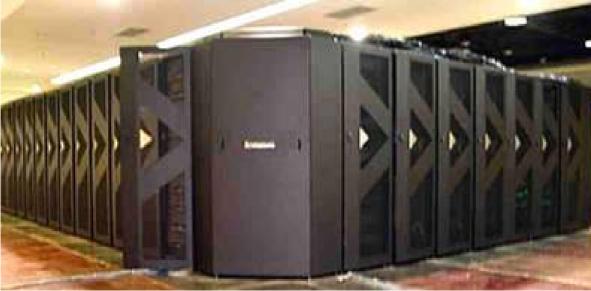
In order to Standardize the process of database construction and database Schema for data integration, Series of specifications for SDB have been published.

- The standard process of scientific database construction and document specification
- Data Sharing Policy and specification for data sharing statement
- Core Metadata Specification for SDB (Ver2.0)
  - A metadata repository and clearing house has been established in the Scientific Data Center
- Some metadata specification for special domains
  - Flora Images, Ecological Data, biological species and so on.
- The Framework for Data quality control and evaluation



## **3.IT Infrastructure Construction**

- Data Center
  - 20TB SAN Storage
  - 50TB Tape Storage
  - TFLOPS-scale computing capacity



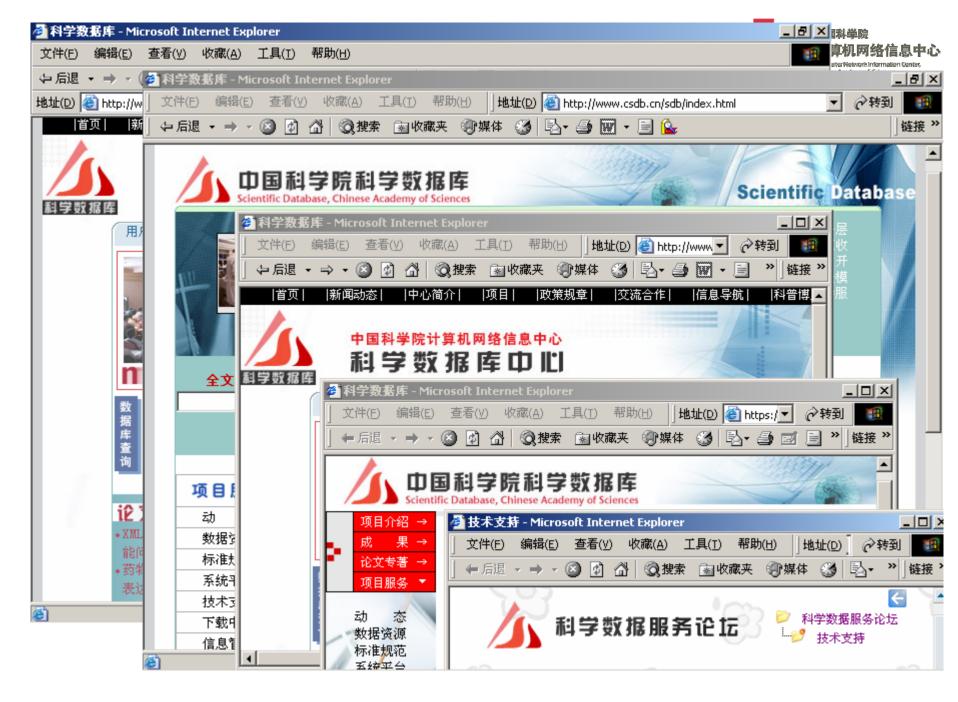
	4	23
2	25	22
2	6	21
		20
2		19
3		18
		17
5		16
6		15
7		14
8		13
		12
1	0	11
	S1 52 S3 S4 S5 S6 S7	1

Lenovo DeepComp 6800

## 4.Data Service



- A Portal website of SDB has been established and put into service at <u>http://www.csdb.cn</u>
- Over 40 distributed data service websites have been built
- A portal website for technique communication and supporting in SDB community has been established, <u>https://support.csdb.cn</u>





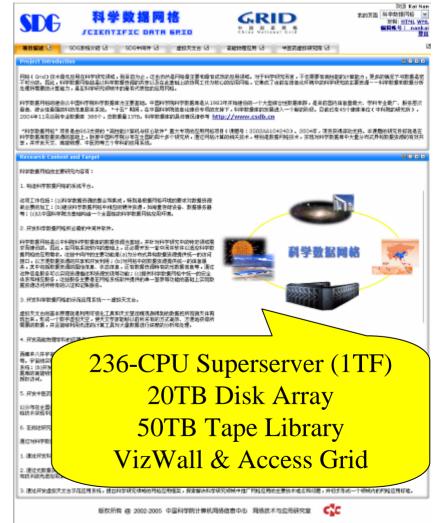
- Scientific data is one of three poles of the cyber infrastructure of CAS
  - Networks
  - Computing
  - Database
- SDG is a sub-project of SDB

## Scientific Data Grid

- SDG is built upon the mass scientific data resources of the Scientific Database (SDB).
- Scientific Data Grid (SDG) is a typical project of CAS e-Science based on SDB, also a pilot.
- The vision of SDG is to take valuable data resources into full play by benefiting from advanced information technologies, in particular, the Grid technology.

## Scientific Database (SDB) & Scientific Data Grid (SDG)

· BRAN			A +6R0 12E
usegradue.		化学专业数据库	
(a) NUMBER		等的重要组成部分,是中制限和 等。上海有机业学研究所的教授 方的综合性优惠系统,可以遵保	所申指讓我的地台科於信息於道 (這個)工程信息化讓我的動力中 常期是是呆子化学化工时究竟开 化合物育笑的自名、结构、最不 化学玩心、更否求的边唱、灭然 通 
• 伊利分类			Google
はまたやまたりの     はまたやまたりの     はまたいたけの     はたいたけの     はたいたけの     はたいたけの     はたいたけの     はたいたけの     はたいたけの     はたいたけの     はたいたけの     にたいたいたい     の     に     れ     に     れ     に     れ     に     れ     に     れ     に     れ     に     れ     に     れ     に     れ     に     れ     に     れ      に     れ      に     れ      に      れ      に      れ      に      れ      に      れ      に      れ      に      れ      に      れ      に      れ      に      れ      に      れ      に      れ      に      れ      に      れ      た      れ      に      た      れ      に      た      れ      に      た      れ      に      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た      れ      た     た     た        た        た        た        た        た        た        た	또해의보습科부 귀집정보하入관한체험, 제정권권 환大권학 가집절권 환大권학 지정권적 관차권학 지정권학 관차권학 지정권학 관차권학 가집권학 관차권학 가집권학 관차권학 관련 2014년 2014	地理科学 午頃なた料学はなり、 市場にた料学なり、 市場になり、 市 日本 市 日本 市 日本 市 日本 市 日本 市 日本 市 日本 日本 日本 日本 日本 日本 日本 日本 日本 日本 日本 日本 日本	管理得免差 之学生生於根準 之学生生於根準 還出定其以子的影響 理论定其以子的影響事件 有些的單子和使用 有些的單子和使用 不可能的單子和使用 的單子和使用 的單子和使用 的單子的一個人。 一個人的一個人。 一個人的一個人。 一個人的一個人。 一個人的一個人。 一個人的一個人。 一個人的一個人。 一個人, 一個人。 一個人, 一個人。 一個人, 一個人。 一個人, 一個人。 一個人, 一個人。 一個人, 一個人。 一個人, 一個人。 一個人, 一個人。 一個人, 一個人。 一個人, 一個人。 一個人, 一個人。 一個人, 一個人。 一個人。 一個人。 一個人。 一個人。 一個人。 一個人。 一個人。
市政主義不會影響			
45	institute	s parti	icipate
	503 d	atabas	es
	16	6 TB	
	10.		



ComputerNetworkInformation Cent Chinese Academy of Sciences

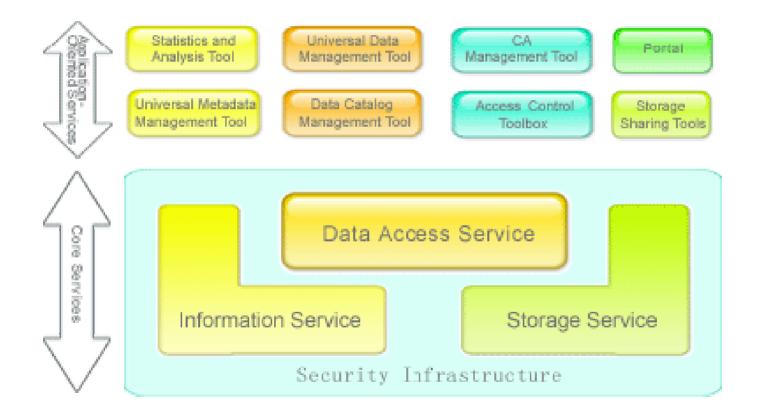
★研算数1010-59812350 010-58812300



## Requirements and SDG

- How to FIND the data I want from hundreds or thousands of databases
- How to ACCESS large-scale, distributed and heterogeneous scientific data uniformly and conveniently
- How to make sure all this goes always in a SECURE and proper way





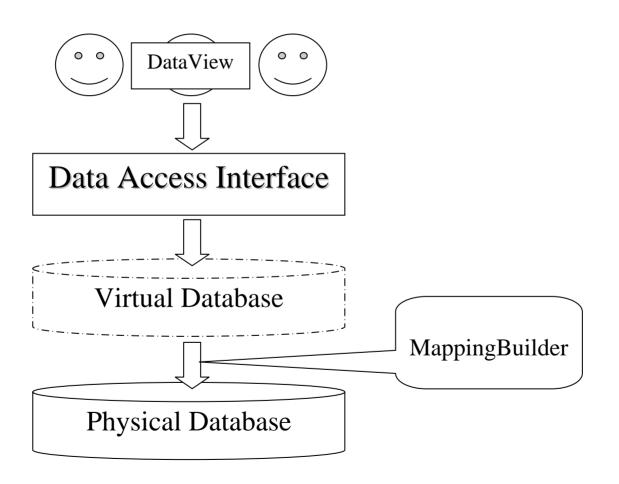


# Data Access Service (DAS)

- Uniform Access Interface (read-only)
- Rich metadata
- Easy publish on web
- flexible configuration and extensibility



### DAS modules



### **SDG** Services



] 古代: 文件 @		<b>日食记录查询结果</b> - 繊促) 查看(V) 收							
(3)后退					🔊 - 📚 🔳 - 🗔 🕄 🛍 🙆				
地址 (2)	🕘 h	.ttp://www.astronomy.	csdb. cn:80	80/dataview/Re	sultDisplayer.jsp			🖌 🏓 转翻	刮 🌀 SnagIt 🛃 链
	Datav	iew首页>>古代天象日報		Data	访问服务 Access Service	3		$   \neq_{\rightarrow} $	<b>DAS</b> 使用帮助
Ę	以下是	<u>16W自风</u> 22 <u>日1(八家日)</u> 第 1 页; 总 17 页			10135-302-37	±0/0	* *		下页尾页
1	记录 号	书名 (bookname)	卷 (bookno)	页码 (bookpage)	内容(cont)	朝代 (dynasty)	帝王 (emperor)	年号年代(years1)	公元纪年 (years2)
1	1	清康熙浙江《长兴县 志》	卷9	页11	日有食之,昼晦。	清	圣祖	康熙元年正月朔	(1662.2.18)
2	2	清康熙广东《吴川县 志》	卷9	页12	日食。	清	圣祖	康熙元年二月朔	(1662.3.20)
3	3	清光绪上海《江东 志》	卷1		日食。	清	圣祖	康熙三年正月朔	(1664.1.28)
4	4	 清康熙江苏《扬州府 志》	卷3	页21	日食,星斗皆见。	清	圣祖	康熙三年十月	(1664.11.18→ 12.16)
5	5	《大清圣祖仁皇帝实 录》	卷13	页17	日食	清	圣祖	康熙三年十二月戊 午朔	(1665.1.16)
6	<u>6</u>	《清史稿?9?9圣祖本 纪一》	卷6	页170	日有食之。	清	圣祖	康熙三年十二月戊 午朔	(1665.1.16)
3		~ 清康熙安徽《望江县 志》	卷3	页83	日有食之。	清	圣祖	康熙三年十二月戊 午朔	(1665.1.16)
8	8	《清史稿?9?9天文十 二》	卷37	页1416	申时,日食九分弱,次于南斗。	洁	圣祖	康熙三年十二月戊 午朔	(1665.1.16)
9	9	《清朝文献通考》	卷263	页7217	日食,在斗宿二十一度二十分,食八分五 十四秒,申初一	洁	圣祖	康熙三年十二月戊 午朔	
1	<u>10</u>	《阅世编》	卷1	页3	申初,日食八分。	洁	圣祖	康熙三年十二月戊 午朔	(1665.1.16)
_	下页月	<b>≩</b> 页						跳转到	副第1页 查看

✓



### **Discovery and Access**



〕中国自然保护区数据库五询界面 -	Microsoft Internet Es	splarer		_ 1 🛛 🔀		中国科学院
170 MAN TOD ATO						▲ ▲ ↓算机网络信息中
3 月後 - 〇 - 💽 🗟 🎸 🔑 秋東	👷 收留天 🛛 🙆 - 🍇	* Google · 百人一刻	★ 後親来 646 内容 + 100×100 合 422	个已赶税 🛛 医浸泡 🎽		Computer Network Information Center,
La D) 👩 Su (k.//158, 555, 148, K7:0000//	utoviev/gazyTates Carriènera	te:js/%oxiedR1%fox///s	🕽 R. 🌒 Sung" 🔡	ernet Explore	Y.	Chinese Academy of Sciences
中國自然保护区換据库 一 <u>中期沿空间</u> 一定等等方法		中国自然保护	●区数据座	""""一团访	* Coogle - EAD-M ter before collecting and a l D III Software con Software	▲ ● 後親来 ### 内容 + 「空話書 日本: 个日志和 同意考 <sup>**</sup> ● 5*** ● 5**** (1) 日* ● 5**** (1) 日* ● 5**** (1) 日*
大會會菜風虫公布名景說 報和 氯隆 → <u>松達直道</u>	《 <b>仙儿。</b> 图錄中文名    李翰英文名	名 条件运输将	neas laterite His new Horse Construction	LULERERN	<sup>油面的</sup> 县级地名蓼	<sup>#用電動</sup> 数据库
年期 式南朝戦後遊遊分布名景設 編隊 年均	当町 unaTang 深前景 unaraig	22		Jatabase descip	ti xa	
▲ 并推动物标本数据库五询界面		包含 v			条件运算符	
※「100-000000-4-2000-10-000000 文字(2) 建築(2) 亚ケ(2) 収置)		Explorer			35 8	-
0 TE - 0 - R 2 6 A		& <sup>20</sup> Google · 百人开始	★ 読録案 646 内容 + Popeliet ₽	422 个已扩张 医液泡 **	¥† 🖌	
8.12 D) 🛃 50.57//158.225.148.4740					ernet Explorer	_] # 🛛
数数字は描葉共存 中国台紀県が区境担体 ・特 <u>地球で</u> が ・考試構成 大協会林尾島公布名書数	Data- ewing -> @ BabibRas	脊椎动物	林本数据库	<b>计引动实验探索</b>	前间服务 and Service	
氯体 → <u>物油面</u> 直					云南森林昆虫分	たな子野田中
十百次市 河	宇藤中立名 宇殿英立 物州中立名 CALASSE	23 音作运算表 122 M	m			
云南蝴蝶资遵分布名景談 据库	中国紅豆形 (1603	28 ¥		(中位含糖谱)	275千月春,内容包括制作生物。为6	24日、我们我们没有你这些资料的特担关证具。
- <u>Baas</u>	critis critis			产级英文名		âi de la companya de
一十 <b>百</b> 次百代 强新山区陆生脊椎动物偏	Nega comu	22 ×		VESE	包含 v	
日数据库	日中京名 C_033258	22 ×		WOLV	包含 V	
	亚科中文名 C_SOMIZ	28 ×		PENT:	<b>包含</b> 💌	
<b>潮汕江流域动物冬祥竹编</b>	分析地代码 34	2合 💌		DEB	변상 · · ·	
□私新門	物性天光名 4.4885	HS M		LES	包含 V あか G	
	科名 2,4012	28 ×		K IS	12 C	
爭構动物分布名录破發展	傷神代吗 口	22 m			包含 ¥	
→ <u>物油素</u> → 再現畫道	18082000 19082000	122 = =		LE_COOZ	<u> 현송</u> 💌	
作用动物标本 <b>和</b> 新用	8 % DOGDER	28 ×		×		0.4
<u>總</u> 元中				🔮 Internet 🛛 🕸	ارك. ⊛ دو هما 10 م	
					與東豐宗 10 m 余t	
			→ <u>●秋田</u> 竹根动物花本数 新	-		
				•		💣 Internet 🗰



SEREN CAPEN		3 3 🛱						
Cattor Ca	AZET Novio N	PZST SS PZST SS Prestar Prest Prest Prest Prest	21 0000 15 15 10737416 10737416 31474636 21474836	2428 Bet	6872 8 2 2	14270 	\$2814	Rah R R R R R R R
Control Brid 2555 Frage State Frage State Support Support Controls Con	9 <b>a</b>							
	SBZS Distances	s cargores catalogi	0.					





### SDG Today





### sdb6800 Superserver

- 59 nodes
   /236 CPUs
- official service started in Apr. 2005
- node usage 79.7% storage usage 87% (by Sep 2005)





### SDG Storage System





### **Visualization System**





### portal.sdg.ac.cn

🚰 科学数据网络门户 - Ticrosoft Internet Explorer	
文件 (2) 编辑 (2) 查看 (Y) 收藏 (A) 工具 (1) 帮助 (A)	1
🔾 后退 🔹 📀 🔹 🛃 🏠 🔎 搜索 🧙 收藏夹 🤣 🍃 🔂 💭 🎒 🦓 🦓	
地址 🕕 🥘 http://portal.sdg.ac.cn/jetspeed/portal/media=type/html/user/anon/page/default.psml/js_pane/P=10498a90af6=10000	✓ → 转到 链接
Google - 💽 Search 🔹 🐝 🖓 8 blocked 🥙 Check 🔹 👯 AutoLink 🔹 🗐 AutoFill 🔩 Options 🖉	
科学数据网格 CIENTIFIC DATA GRID 中国家网格 Cina National Grid	用户名: 态码: 登录 <u>登录符助</u> 效迎济HSDG Portal
<b>项目介绍</b> 网格门户 硬件平台	
科学数据网络项目简介	
"科学数据网格"项目是由863支持的"高性能计算机与核心软件"重大专项的应用网格项目(2002-2005)。科学数据网格的建设以中国科学院科学 科学院科学数据库是从1983年开始建设的一个大型综合性数据库群,是目前国内信息量最大、学科专业最广、服务层次最高、综合性最强的科技信息服务 中国科学院信息化建设专项的支持下,科学数据库的发展进入一个新的阶段。目前已有45个建库单位(中科院的研究所),2004年11月达到专业数据库 13TB。	紧系统。"十五"期间,在
科学数据网格项目的研究目标是科学数据库数据资源的基础上,通过网格计算的相关技术,特别是数据网格技术,促进科学数据库中大量分布式异构数据 上开发对科学研究有直接实用意义的应用系统"虚拟天文台"。	资源的共享,并在此基础

本项目的主要研究内容包括:构造科学数据网格的系统平台、开发科学数据网格所必需的中间件软件、开发科学数据网格的示范应用系统一一虚拟天文台,并在这些研究工作成果的基础上,提出一种科学研究领域的网格应用框架,争取成为领域内的网格应用标准。构造科学数据网格系统平台的工作主要是整合和集成现有的数据资源,即科学数据库。科学数据网格中间件软件由以下三个部分组成:网格信息服务系统、科学数据网格数据访问系统、科学数据网格安全体系。在科学数据网格中间件的基础上,我们还开发了一些实用的工具,如:通用数据管理工具、数据目录管理工具、数据量统计工具、通用元数据管理工具、访问控制工具包及图像数据处理工具等。虚拟天文台的建设将利用网格技术和科学数据网格中间件所提供的服务,实现国内天文数据的无缝透明融合,并与国际资源进行互联共享。同时,配合国家重大科学工程LAMOST项目和天文创新工程的需求,逐步建成连接国内外天文研究资源的"网关",为我国的天文学科发展提供重要的基础科研环境。



## Collaborations

- PRAGMA
  - -<u>www.pragma-grid.net</u>
- EUChinaGrid
  - <u>www.euchinagrid.org</u> Interconnection and Interoperability of Grids between Europe & China
- IGTF / ApGrid PMA
- •



# 5.e-Science applications5

- High Energy Physics
- Astronomy
- Biology
- Natural Resources
- Disaster Reduction



# YBJ-ARGO/ASγ

- Italy, Japan-China cosmic ray observatories in Tibet.
- 200TB raw data per year.
- Data transferred to IHEP and processed with 400 CPUs.
- Rec. data accessible by collaborators.





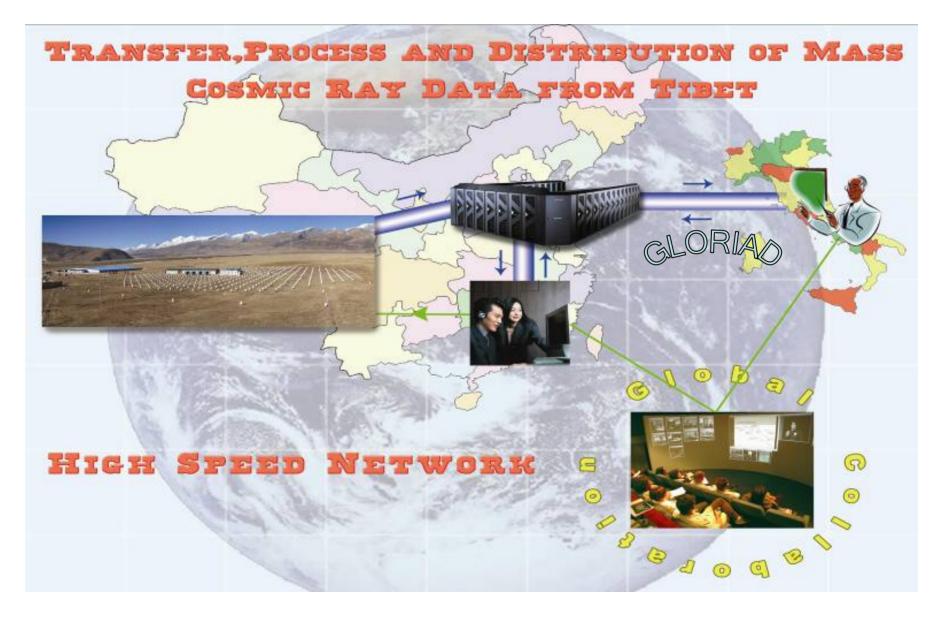
## **YBJ-ARGO**

 Established a 8Mb/s link from Tibet to Beijing in March 2005, by CNIC of CAS. Upgraded to 155Mb/s in March 2006.

- Stopped bringing tapes half year ago.

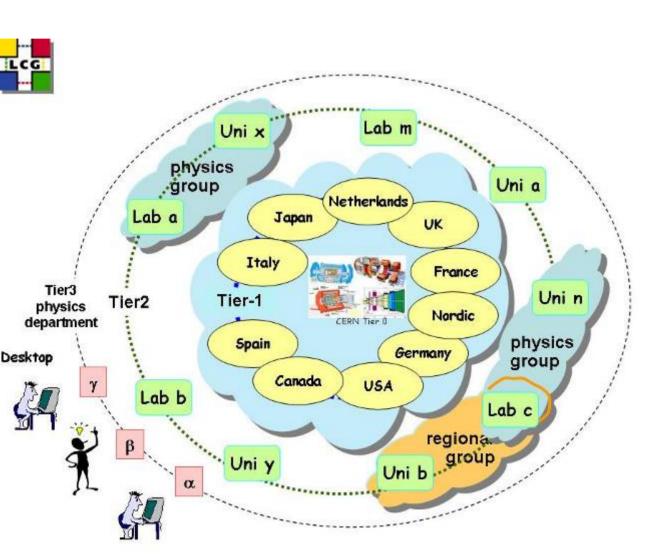
 Building a computing system based on LCG, collaboration of IHEP of CAS, CNIC of CAS, INFN of Italy, EU-China Grid application under EU FP6.







### LCG Tier-1/2



- to build a LCG Tier-1/2 node in China
- Institute of High Energy Physics of CAS
  - CNIC providing support and working together with IHEP

# LCG2 production site @CNIC

GStat: 03:31:38 12/02/ 文件で 編辑で 音看 0 ゆ     G高退 · ○ · ▼ 2 ☆     http://goc.g	文藏 ( <u>A</u> j	〕 工具(I) 帮助(H) 搜索 👷 收藏夹 🧑 🔗	• 🚴 🗷 • 🖵 🕄 🛍 (		FIUNG.	CNIC.		A <u>r</u>	
GStat: 03:31:38 12/02/05									
<u>home</u> alert table service regiona	<u>l ser</u>	vice metrics <u>links</u> ? pro	d <u>test seegrid</u>						10000 00000 P* 10000 0000 P*
ServDuplicate									
AEGIS01-PHY-SCL	<u>ok</u>	ALBERTA-LCC2	BEIJING-CNIC-LCG2-	IA64 18	BEIJING-LCG2	BelGrid-UCL	OK BG-INRNE CT		in Appendix Section 1. 2 Sector Sector Sector 1. 2. Sector Sector Sector 1. 2. Sector Sector Sector 1. 2. Sector Sector 1. 2. Sector Sector 1. 3. Sec
BG01-IPP	<u>ok</u>	BG02-IM CI	BG04-ACAD	<u>0K</u>	BHAM-LCG2 OK OK	BIFI	OK BITLab- LCG CTCT		INCOMENTS & Income Television Control of Co
BNL-LCG2	л	BRISTOL-PP-LCG <mark>ok</mark>	BUDAPEST	<u>ok</u>	CARLETONU- LCG2	CAVENDISH- LCG2	OK CEA- DAPNIA- SACLAY		$ \begin{array}{c} \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\$
CERN-CIC	<u>ok</u>	CERN-PROD OK	CESGA-EGEE	Л	CGG-LCG2 CT	CIEMAT-LCG2	OK CNB-LCG2 OK		Terrent and a discretion of index of a discretion of the set of the set of the set of the set of the set of the set of the set of the set of th
<u>cpDIASie</u>	<u>SD</u>	CSCS-LCG2	<u>csQUBuk</u>	<u>0K</u>	<u>csTCDie</u> OK	<u>csUCCie</u>	SD CY01- LCG2 OK		A second
CYFRONET-IA64	<u>ok</u>	CYFRONET- LCG2	DESY-HH	<u>OK</u> <u>OK</u>	Durham OK	egee.man.poznan.p	1 OK ekplog2 OK		
ESA-ESRIN	<u>SD</u>	FMPhI-UNIBA OK	FZK-LCG2	<u>OK</u> OK	giAITie OK	giDCUie	ок <u>giDITie</u> ок		1000000
<u>giTTCie</u>	<u>CT</u>	giTTTAie <mark>ok</mark>	giTTTRie	<u>ok</u>	giNUIMie <b>ox</b>	<u>giRCSIie</u>	ok <u>giULie</u>		t : Marid areas Marid areas
giWITie	<u>SD</u>	GOG-Singapore 💴	GR-01-AUTH	<u>0K</u>	GR-02-UoM or		CD		
GR-05-DEMOKRITOS	<u>ok</u>	<u>GRIF</u> ok ok	GSI-LCG2			Monito	oring Info	on	
HPC2N	<u>ok</u>	HPTC-LCG2ia64	HR-01-RBI		BE	JING-C	NIC-LCO	G2-IA6	4

# VO = World Wide Telescop





TRACE









1.2m



### **China Virtual Observatory at SDG Portal**

SDG

82.1508 0.16873

10000 BUILDING DAMAGE

42(10)2010/02/2010/2010/2010

REPAIRS NO. 1 THE SALES

> E人の3数配定お子白

> > 8000

Marine Services

10 10 10 10 10 10



CONCERNENT PROPERTY AND	10-10-10-10-1	COMPANYA-C	Construction in the last	C. BRITTER CONTRACTOR	-

at the second second

de l'anne de la constante de la
minter.

ALC: NO. OF A REAL PROPERTY.
al- <u>voeiner</u>
8668
AERLINWIGHT, TAR

(4) 長井田文計業長年

INVESTIGATION OF A DESCRIPTION OF A DESC - 42 の構成的も実現実現象、実現会で500人が0.00(開発系の構成、影響能力的ない.40、目的を学び発展系が発展するまたないの実現を認識の構成的にや中国も実質的定時的も Balaka, Balaka Manana, Sanaka Manana, Kanana Manana, Kanana, Ka Kanana, Kan

GRID

進入USAO、A2型書給金鋼板服务

**Application Tools** 

CHEERENHAMT, BUT CONSCIONDERNITOR, DESCRIPTION DESCRIPTION DE LA CONSCIENCE DE LA CONSCIENC

科学数据网格

CONTRACTOR OF THE SECOND

EALS & RANKER & PART & CREW &

CHARLES SCORES SCORES & SCORES & SCORES & SCORES

Addition and interaction and even is to addie addies the user to invalid or definition images of employed of the other to premimente addition in the associated catalogues. In extra basis, which interactively addies and the state data of the other additional and the interactive has been added additional and the basis areas and other additions for all source classes in the left. The others work-addred has been added additional is used at an extra state of the additional basis and the basis of the left. The others work-addred has been added addred is used at an extra very the nutlework-additional basis. The others addred addred addred addred addred is used at an extra very the nutlework-addred basis. It is the other addred addred addred addred addred addred others addred in the other addred a

As image dete base called "'Aladin image server": this protect is managed by Energoia Dormanel.
 A integrator scar tool called "Made tays" or also "Made": this protect is managed by Financial Called "Made tays".

BO CORE Louise the cases of SHI SHISHOPPING SHIELD AND A THE SHIP

5		15	
3		10	
3 - 200	diam'r.	B	
1.1			

• 医氯化物化物品。GH中的氯化和合用不利用的通知的实现的有品质是不可能的通过,并可以为自己的原则和含化和含化和含化和合用。 HPADADA、FRADADA FRAMENTATIONAL AND ALL CONTRACTION AND ALL 中国語には自然的には、日本は主要な目的ではなか。その、「中国が自然にないからななかの」。 名からが多く、たませたまた時にかどなかの事業が設計するだかである新聞語れたができた。満年でない話を、可能に話をらび置きただまを思想をも位置きたけが加かかき取扱では を取得させた。このでありたちからが行われたから使きた何かとしてます。

通过这二个问题是不要了这话中,并且你们会会的中型量的以及其了,我们问题是在此大学们不知道我们要不知道你是我们的事故还是不知道,我都不能不能是不知道你们是我们的时候。 第二章是我们已经会学师大学们不知道我们们是对我们不会问道:"你我们是你那么什么你没有你们可能是你的人们可能。"

油入服用工业学习家主要都是放出分析

可以高lona構成容易を始めまたの最多、存在気候VOG素やら

NOTINg是是我又去自己们的公司这一个XML回答着,以实现在Court的HoutDate(上外ACTableASE用的的样子中,但是我都是我们又 的教师的是中午的方式系统的中国的这个时候,我们就是自己不会是你的情况的的话。

verifies is an VM. (Rev for openetifies CAR developed by the chean-to peoplet, triag the first, yes can open vertable files, with and analysis them with CoerOffee Carls. WOTfor a fully integrated application, nour need not inow writting about 24 dex, pargomming calls.

10.0 連絡入場入最後に利用力を消費用を取 Cotalog Name 教育会会的COS業務長年(Cotalegue Name)等業者目標中線回路開展2045年(/m Name)。他認知業者特別的原因用中級医力的分配要素長年,但如為人業者長年 U100、最近常業務業券中有高級一個新生活性の支付後、資源局別的新聞、1007的新業業の支付後、業業長期、新生活年期時期に最近常常的COS所有高額。 WEARSHOLD AND A CAMPANESS WARRENTSPACED Former 重用

06			

10026200

建输入输送器数数的数数数数数 Catalog Name

### CONVERSE

4.集合会会会委任任何、利用的公司、可以成为公司、建立社会公司委員会会会会、直知目前公司公司合理委員会、自己的利用目的公司委員会、可以可以提供

http://wkweb.acstresby.h/ogitbin/goet20/262 /i
They TF2 with two to: It gos/Antibarts and strates ling Ary and s / 11/2022/ 4+4 Mitgary / Analos Jennas Lang / and kines / DATS / and along /
Homepage of BAYC: http://batc.bao.ac.cn

### World Date Center for Astronomy

科学数据网格

ACCENTIFIC DATA SELO

BANKS STREET & FROM & CROWN &

CONSIDERT CONSIDER CONSIDER CONSIDER CONSIDER CONSIDER

大教师: [1] 从大教师: [1] 只有这些话: (1) 现于常常: (1) 主要无效: (1) 和他们的目录: (1) 高校教师: (目前的常常有法

期料经验量无效数据中心电子设置Stranbourg无效力,通常用上量事品的无效数据中心,这中心考虑的发生无效中标的通用数据,还是通常数据的内容包括↓1.元件数据数据↓1.3 例

了她就是我不能是我的这个生活是我不能会说,这是为了你的老师不同的工具来说得这些天空的后,consel了我么ACO我们来来这种老师是这些天文是常爱爱,而我知道了一 

出身上面的外后,就们我们又这些确定在马来完全要要做我们了教师,她又中的大家的问题是这人用方面都要要帮助中,用为它的帮助我们了这些马来要做我用干分。我外,他们以 希望是你是当我的学校们的没有了会教教你?"你说是这些你,我们就是这些你。我们没有这些你说,我们又是你会帮助我们的问题?

MATC House Antone Convertineの第一下15世紀大学員、大阪子 2009年、モナキロ代表了他で2008年2月、GAN 日本の日本の小学生の子、大学学校学校会、日本の日本 美国会社会社会社。

BLTC instant Briging dataser Clamater-Connections, which is a multi-outor sky survey project conclusted in Rulational Automatical Discussions of Disrus (AreCC). The taking a used by the project is a 64/06m 03 Series of telescope is carried at implants protoce of MAC. The filter ration induces a provide the series of the series of the following in the series of the CCD in Series of the Series of the CCD carries with 15 months and provide the series of the Series of the followings. The field of new of the CCD in S<sup>CCD</sup> with a solar project IS and CCD carries with 15 months and provide at the Series field carries and carries of the Arc CLO in S<sup>CCD</sup> with a solar project IS and the CCD carries with 15 months and the series of the Arc CLO in S<sup>CCD</sup> with a solar project IS and the CCD in Series and the Series (Series of the Arc CLO in Series and the Series (Series of Series CCD) and the solar project at the Series of the Series (Series of the Series (Series of the Series of the Series of the Series of the Series (Series of the Series of t

**Data Services** 

SDG

COMPRESSED

The later.

THE BAT

himone The mak

The World State Center solver was marked as artistic and distributed where related in the absormational programs of the USS-1028 International complexical mark colocation emails and a the analysis and assess the most contract relation to come contractions and too mark solutions and the WOC solution mark includes 55 Centers in 12 countings. To holdings indiced a winde related to solve contraction, devincements, and homan developed stats.

The solid data Center (NOC) for Attorney, also inner, as all in Attorney Data Center (MOC), in the only data center an attorney in the NOC and the solid data center attorney and and the solid data center attorney attorney attorney and and the solid data center attorney attorney attorney and and the solid data center attorney attorney

Veto-extronomy are co-funded by chinese tational astronomical observatory (NASC), data draring funding projects of Ministry of sulence and technology (MOST) and Chinese Academs of Sciences (OAS), and maintained by (AMOST, NAOC).

Hernebege of W00 Astronomy http://hado.leneost.arg/website.

<u>\$1</u>

/anshiwns/ 

GRID

にの意思考査府
構成化ないのと思想がなくていたなない。Name ) 生活的成素を加えた影響性学校学校生活を発生するの。作用的人成素のないのの、効果の生活が生活の影響性なが…10%、成素のな ないた高いで見たくていたがななた。

Artives of MCC-Artistrane, http://back.iwww.en	к
Germ Detakane Name	

Course Detabase Name
CDA最后者 and



## Avian Bird Flu Alarming & Predicating System

By: Institute of Microbiology, CAS Institute of Zoology, CAS Institute of Virology, CAS CNIC, CAS

### Avian Bird Flu in Gangcha, Qinghai Province, May 2005网络信息中心





### Tasks

- Integrate bird-flu basic databases from multiple institutes
- Field survey on bird-flu
- Establish bioinformatics comprehensive analysis system for bird-flu
- Establish bird-flu alarming and predicting system
- Establish international cooperative work environment
- Establish information publishing system (web)

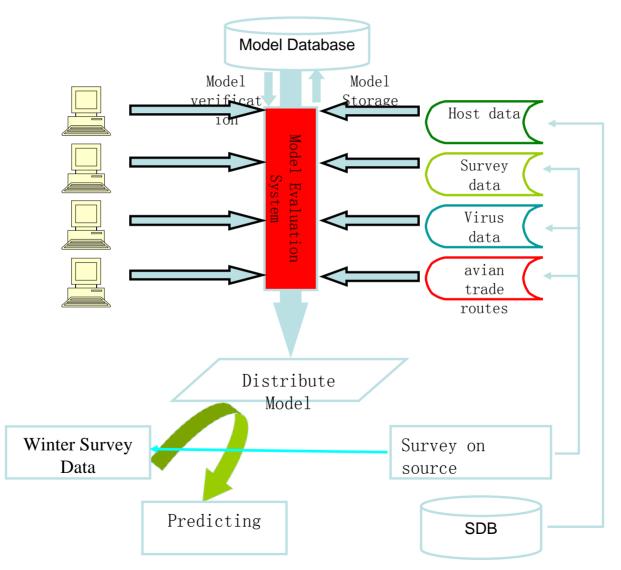


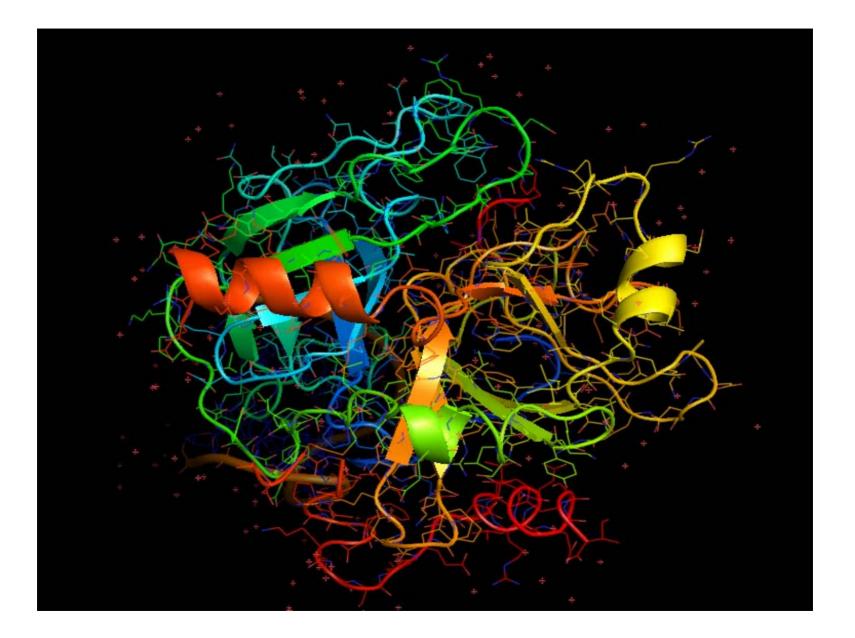
### Bird-flu basic databases

- Standards
  - Bird-flu basic database's model and data standard
  - Metadata specification and description language of bird-flu information
- Data resources
  - Bird-flu virus resource database
  - Bird-flu virus inherent resource database
  - Bird-flu history database
  - Bird-flu dynamic monitoring database
  - Bird-flu host database
  - Bird-flu information database
  - Bird-flu international DNA database
  - Bird-flu international research progress database



# Technical architecture





🚰 禽 瀛 感 信 息 阿 - Microsoft Internet Explor	er			- 7 🛛
文件(E)编辑(E)查看(Y) 收藏(A) 工具(E) 帮助(H)				A.
🕜 后退 🔹 🕥 🔹 💽 🐔 🔑 搜索 🌧 收藏夹 🤗 😥	3• 头 👿 • 🖵 📂	A 🛍 🏠 🚳		
地址 @) 🥘 http://www.avian-flu.cn/		₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩	🝷 🔁 暂停拦截 🍕 清瑪	里 💽 修复
				<u>^</u>
禽流感信息网		http://ww	vw.avian-flu.info	
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Charles 1	<b>1</b> 金洁咸	信息网	
	A MARS		on Network	
	H		rian-flu	
新闻动态 (56	i5) <	学术文献 (228)	<	
• 盲目扑杀家禽并不能阻止禽流感的传播	(02-01)	• Avian H5N1 Influenza in Cats.	(09-02)	
	D D (02-01)	77501 chicken influenza viruses disp.	L (09-01)	
	P Program			
► ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■	ards and Di	<ul> <li>H5N1 influenza: a protean pandemic 1</li> </ul>		
• 亚太经合组织将在美举行禽流感研讨会		• Assessment of risk to human health :	a (05-21)	
• 《新英格兰医学杂志》: 禽流感病毒可)	<u></u> (01-28)	• Public health. Public health risk for	(05-14)	
• 泰国研究证实: 禽流感可通过人际传播	(01-24)	• Use of the DNA flow-thru chip, a th	c (05-01)	
• 世卫称可能发生通过人类接触传染的禽	<b>氟</b> (01-23)	• Reemerging H5N1 influenza viruses in	n (05-01)	
• 泰国又发现禽流感	(01-21)	• Antiviral chemotherapeutic agents ag		
	更新新闻		更新新闻	
新闻检索:	提交查询	文章检索:	提交查询	
	2)	知识介绍 (76)	<	
• 越南今年12人死于禽流感	(02-01)	<ul> <li>人畜共患病之不完全统计</li> </ul>	(12-21)	
<ul> <li>- 越南26省发现禽流感 两名女孩已被证实</li> </ul>			(08-22)	
<ul> <li>泰国又发现禽流感</li> </ul>	(01-21)	• 日本专家: 吃经过加工的鸡肉不会感染禽流	<b>Š</b> (03-01)	
• 越南禽流感患者增至7人 其中5人已经列	亡 (01-20)	<ul> <li>解读::::::::::::::::::::::::::::::::::::</li></ul>	(02-27)	~
			🥑 Inte	

# 6. Cooperation & Communication

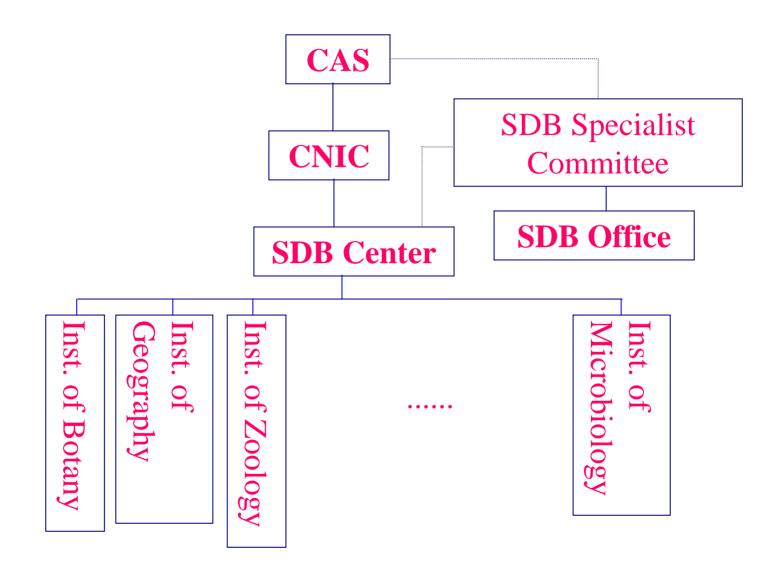


### CODATA

- Secretariat of China CODATA
- Scientific data & database
   development
   and sharing









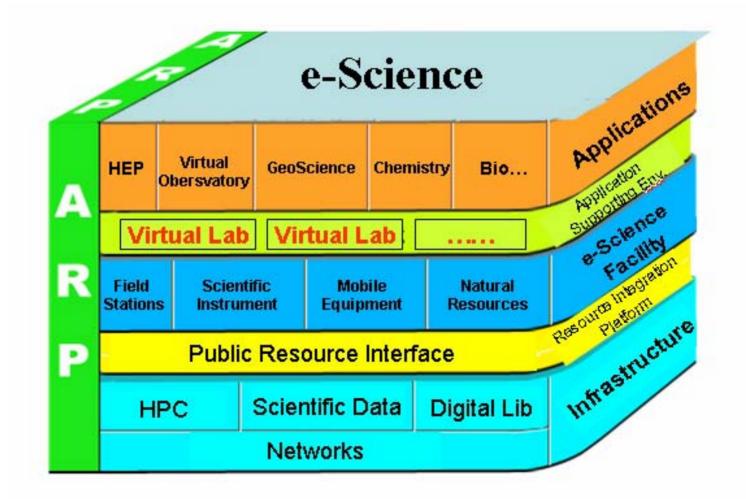


### SDB in 2006 - 2010

### "SDB Driving e-Science of CAS"

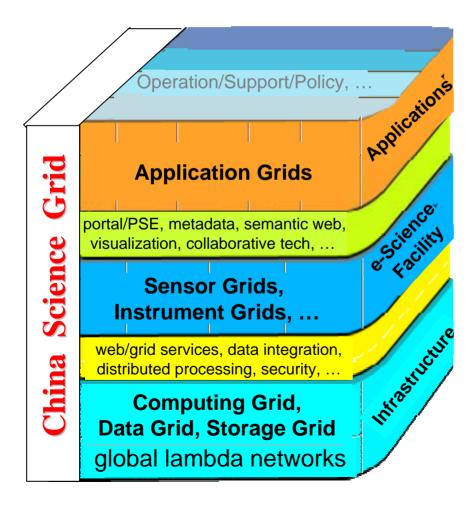


### Framework of CAS e-Science





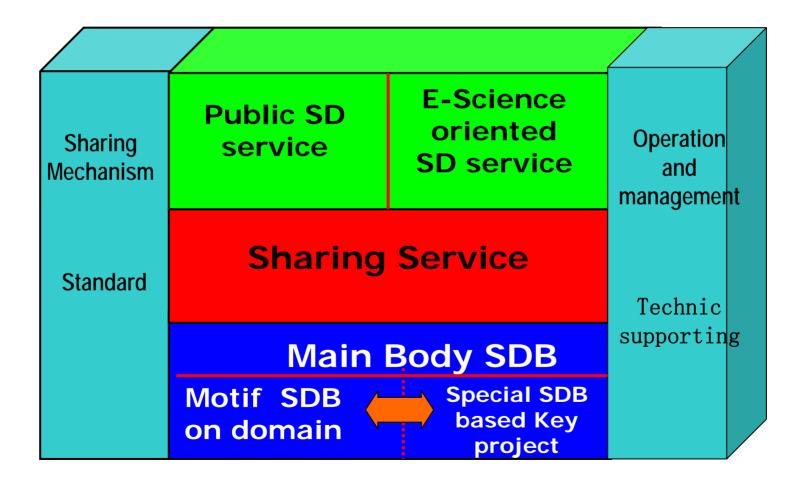
# Technical View of CAS e-Science



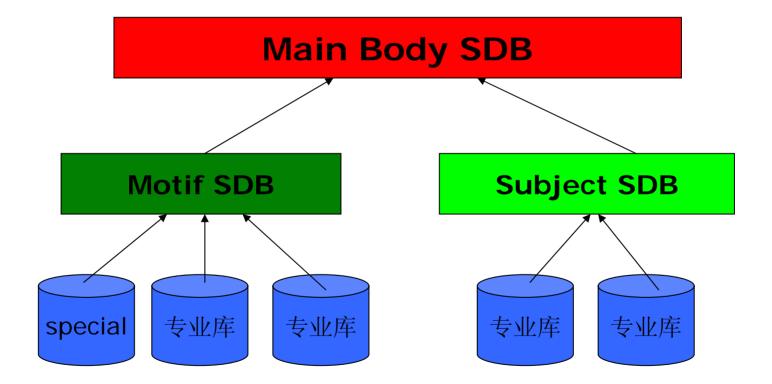
- Grid-oriented
- Open
- Sharing
- Collaboration and Virtual Organization
- Security



### **SDB** Architecture









### Main Tasks on SDB

- 60 motif SDBs, 600 special SDBs, 60TB sharing
- Continuing standard
- Platform for sharing service
- Platform for running: 300TB disk, 2-3PB tape, parallel wall visualization based LCD, software, ....
- Pilot applications



### summary

- SDB is a key foundation for e-Science of CAS
- New challenges:
  - Data technic, data engineering, data science
- Data producing, data management, data service, data using
  - Data quality and maturity
  - Data security
  - Data Policy: sharing and property right, ....
- Drive pilot applications
- Sharing and international cooperation



# Thank You !