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)
History & Position

- 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
Mission

- 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
• 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
Some Priorities in Basic Research

- 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
### Priorities in Life Sciences & Biotech

#### 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

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

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

#### 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

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

**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

**Material and Chemical Engineering**
- Green production
- Immobilization and utilization of CO₂
- 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
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 10th 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 11th five-year planning for information construction of CAS (2006-2010)
CAS Informatization Program
2001-2005

industry system web site

virtual museums

networking

Scientific Database

Supercomputing
# CAS Cyberinfrastructure Situation

<table>
<thead>
<tr>
<th>Infrastructure</th>
<th>Item</th>
<th>By 2000</th>
<th>By 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Networking</td>
<td>core</td>
<td>1Gbps</td>
<td>2.5Gbps-10Gbps</td>
</tr>
<tr>
<td></td>
<td>backbone</td>
<td>2Mbps</td>
<td>2.5G-155Mbps</td>
</tr>
<tr>
<td></td>
<td>Oversea link</td>
<td>55Mbps</td>
<td>620Mbps+15GbPs</td>
</tr>
<tr>
<td>HPC</td>
<td>Peak TFLOPS</td>
<td>0.13</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>Linpack TFLOPS</td>
<td>0.05</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>Storage</td>
<td>2.1TB</td>
<td>180TB(Disk+Ta)</td>
</tr>
<tr>
<td>Scientific Database</td>
<td>Member institutes</td>
<td>21</td>
<td>&gt;45</td>
</tr>
<tr>
<td></td>
<td>Databases</td>
<td>180</td>
<td>+503</td>
</tr>
<tr>
<td></td>
<td>Data volume</td>
<td>725GB</td>
<td>+16.6TB</td>
</tr>
</tbody>
</table>
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 11th Five-year Program 2006-2010
SDB status

- 45 institutes across 16 cities
- 503 databases
- 16.6TB total volume

45 institutes across 16 cities
503 databases
16.6TB total volume
Main Tasks in 2001-2005

• Six main tasks:
  – Database Resource
  – Data & Database Specification
  – IT Infrastructure Constructuring
  – Middleware 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)

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

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Number of sub-DB</th>
<th>Developing Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geosciences (151)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Database of China Geotectonics</td>
<td>4</td>
<td>Changsha Institute of Geotectonics</td>
</tr>
<tr>
<td>Scientific Database of Subtropical Agri-ecology</td>
<td>4</td>
<td>Institute of Subtropical Agriculture</td>
</tr>
<tr>
<td>Mountain Environment &amp; Hazards Database of China</td>
<td>3</td>
<td>Institute of Mountain hazards and Environment</td>
</tr>
<tr>
<td>Lake Database of China</td>
<td>2</td>
<td>Nanjing Institute of geography and Limnology</td>
</tr>
<tr>
<td>Soil Database of China</td>
<td>7</td>
<td>Nanjing Institute of Soil Science</td>
</tr>
<tr>
<td>Natural Resource Database of China</td>
<td>55</td>
<td>Institute of Geographical Sciences and Natural Resource Research</td>
</tr>
<tr>
<td>Data Library of Dynamic Geodesy and Resource Environment</td>
<td>2</td>
<td>Institute of Geodesy and Geophysics</td>
</tr>
<tr>
<td>The Mineralogy, Petrology and GeoChemistry Database of China</td>
<td>9</td>
<td>Guangzhou Institute of Geochemistry</td>
</tr>
<tr>
<td>Database of Soil and Water Conservation in the Loess Plateau</td>
<td>7</td>
<td>Institute of Soil and Water Conservation</td>
</tr>
</tbody>
</table>
# Database list (3)

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Number of sub-DB</th>
<th>Developing Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geosciences (151)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xinjiang Ecological and Environmental Resources Database</td>
<td>4</td>
<td>Xinjiang Institute of Ecology and Geography</td>
</tr>
<tr>
<td>Database of Chinese Resources and Environment Remote Sensing</td>
<td>11</td>
<td>Institute of Remote Sensing Application</td>
</tr>
<tr>
<td>China Wetland Database</td>
<td>11</td>
<td>Northeast Institute of Geography and Agricultural Ecology</td>
</tr>
<tr>
<td>Cold and Arid Region Characteristic Database</td>
<td>18</td>
<td>Cold and Arid Regions Environmental &amp; Engineering Research Institute</td>
</tr>
<tr>
<td>Agricultural Ecology Database for China’s Northeast Black Land</td>
<td>5</td>
<td>Northeast Institute of Geography and Agricultural Ecology</td>
</tr>
<tr>
<td>Paleontological and Stratigraphic Database of China</td>
<td>4</td>
<td>Nanjing Institute of Geology and Palaeontology</td>
</tr>
<tr>
<td>Atmospheric Science Database</td>
<td>4</td>
<td>Institute of Atmospheric Physics</td>
</tr>
<tr>
<td><strong>Ocean Science (10)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China Marine Science Database</td>
<td>3</td>
<td>Institute of Oceanology</td>
</tr>
<tr>
<td>Ocean Database of South China Sea</td>
<td>7</td>
<td>The South China Sea Institute of Oceanology</td>
</tr>
<tr>
<td>Database Name</td>
<td>Number of sub-DB</td>
<td>Developing Institute</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Physics and Chemistry (45)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Energy Physics and Related Subject Database</td>
<td>2</td>
<td>Institute of High Energy Physics</td>
</tr>
<tr>
<td>Optical System Database</td>
<td>6</td>
<td>Changchun Institute of Optics, Fine Mechanics and Physics</td>
</tr>
<tr>
<td>Professional Chemical Database</td>
<td>19</td>
<td>Shanghai Institute of Organic Chemistry</td>
</tr>
<tr>
<td>Engineering Chemistry Database</td>
<td>8</td>
<td>Institute of Process Engineering</td>
</tr>
<tr>
<td>Database of Applied Chemistry</td>
<td>4</td>
<td>Changchun Institute of Applied Chemistry</td>
</tr>
<tr>
<td>理化性能及分析数据库</td>
<td>3</td>
<td>大连化物所</td>
</tr>
<tr>
<td>Chemical Substance Toxicity Database</td>
<td>3</td>
<td>Computer Network Information Center</td>
</tr>
<tr>
<td><strong>Astronomy &amp; Space Science (12)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Astronomical Database</td>
<td>8</td>
<td>National Astronomical Observatories</td>
</tr>
<tr>
<td>Space Environmental Database</td>
<td>4</td>
<td>Center for Space Science and Applied Research</td>
</tr>
</tbody>
</table>
# Database list (5)

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Number of sub-</th>
<th>Developing Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Energy Science(17)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China Energy Database</td>
<td>5</td>
<td>Energy Research Institute</td>
</tr>
<tr>
<td>New Energy Source &amp; Environmental Protection Database</td>
<td>7</td>
<td>Guangzhou Institute of Energy Conversion</td>
</tr>
<tr>
<td>Gas Hydrate Database</td>
<td>5</td>
<td>CNIC and etc.</td>
</tr>
<tr>
<td><strong>Material Science(12)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials Database</td>
<td>7</td>
<td>Institute of Mental Research</td>
</tr>
<tr>
<td>Nano Science and Nano Technology Fundamental Database</td>
<td>5</td>
<td>Nano Science and Technology Center</td>
</tr>
<tr>
<td><strong>Others(1)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specialists of Chinese Academy of Sciences</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
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

Lenovo
DeepComp 6800
4. Data Service

• A Portal website of SDB has been established and put into service at http://www.csdb.cn

• Over 40 distributed data service websites have been built

• A portal website for technique communication and supporting in SDB community has been established, https://support.csdb.cn
5. Scientific Data Grid (SDG)

- 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)

- 45 institutes participated
- 503 databases
- 16.6 TB

- 236-CPU Superserver (1TF)
- 20TB Disk Array
- 50TB Tape Library
- VizWall & Access Grid
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
SDG Software Architecture
Data Access Service (DAS)

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

>Data View

Data Access Interface

Virtual Database

Physical Database

MappingBuilder
### 古代天象日食记录查询结果

<table>
<thead>
<tr>
<th>记录号</th>
<th>书名 (BookName)</th>
<th>卷 (Volume)</th>
<th>页 (Page)</th>
<th>内容 (Content)</th>
<th>朝代 (Dynasty)</th>
<th>皇帝 (Emperor)</th>
<th>年号及年代 (Year)</th>
<th>公元纪年 (AD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>清康熙朝《长兴县志》</td>
<td>卷9</td>
<td>页11</td>
<td>日食之，昼晦。</td>
<td>清</td>
<td>圣祖</td>
<td>康熙癸丑正月朔（1662.2.18）</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>清康熙朝《昌平县志》</td>
<td>卷9</td>
<td>页12</td>
<td>日食。</td>
<td>清</td>
<td>圣祖</td>
<td>康熙癸丑二月朔（1662.3.20）</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>清康熙朝《江宁府志》</td>
<td>卷6</td>
<td>页14</td>
<td>日食。</td>
<td>清</td>
<td>圣祖</td>
<td>康熙癸丑三月朔（1662.4.28）</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>清康熙朝《苏州府志》</td>
<td>卷7</td>
<td>页15</td>
<td>日食，日斗皆见。</td>
<td>清</td>
<td>圣祖</td>
<td>康熙癸丑四月朔（1662.5.16）</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>清康熙朝《扬州府志》</td>
<td>卷8</td>
<td>页17</td>
<td>日食。</td>
<td>清</td>
<td>圣祖</td>
<td>康熙癸丑五月朔（1662.6.16）</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>清康熙朝《扬州府志》</td>
<td>卷9</td>
<td>页170</td>
<td>日食。</td>
<td>清</td>
<td>圣祖</td>
<td>康熙癸丑六月朔（1662.7.16）</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>清康熙朝《长兴县志》</td>
<td>卷10</td>
<td>页83</td>
<td>日食。</td>
<td>清</td>
<td>圣祖</td>
<td>康熙癸丑七月朔（1662.8.16）</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>清康熙朝《长兴县志》</td>
<td>卷11</td>
<td>页1416</td>
<td>日食。</td>
<td>清</td>
<td>圣祖</td>
<td>康熙癸丑八月朔（1662.9.16）</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>清康熙朝《长兴县志》</td>
<td>卷12</td>
<td>页1217</td>
<td>日食，于斗宿二十一度二十分，食分五十四秒，申初一</td>
<td>清</td>
<td>圣祖</td>
<td>康熙癸丑九月朔（1662.10.16）</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>《隋书》</td>
<td>卷1</td>
<td>页3</td>
<td>中食，日食八分。</td>
<td>清</td>
<td>圣祖</td>
<td>康熙癸丑十月朔（1662.11.16）</td>
<td></td>
</tr>
</tbody>
</table>

**SDG Services**
<table>
<thead>
<tr>
<th>功能分类</th>
<th>数据库名称</th>
<th>说明</th>
</tr>
</thead>
<tbody>
<tr>
<td>自然保护</td>
<td>中国自然保护区数据库</td>
<td>存储中国自然保护区相关信息</td>
</tr>
<tr>
<td></td>
<td>县级地名数据库</td>
<td>存储县级地名信息</td>
</tr>
<tr>
<td></td>
<td>脊椎动物标本数据库</td>
<td>存储脊椎动物标本信息</td>
</tr>
<tr>
<td></td>
<td>云南森林昆虫分布名录数据库</td>
<td>存储云南森林昆虫分布信息</td>
</tr>
</tbody>
</table>
MappingBuilder & Dataview
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
“科学数据网格”项目是在863支持的“高性能计算机与核心软件”重大专项的应用项目（2002-2005），科学数据网格的建设以中国科学院科学数据网为主要基础。中国科学院科学数据网是自1993年就开始建设的一个大规模综合数据集群，是目前国内信息量最大、学科专业最广、服务层次最高的科学信息资源系统。“十五”期间，在中国科学院信息化建设委员会的安排下，科学数据网的发展进入一个新的阶段。目前已有45个库藏单位（中科院的研究所），2004年11月专业数据库388个，总数据量13TB。

科学数据网格项目的研究目标是科学数据密集型数据资源的整合，通过网络计算的相关技术和科学数据密集型数据资源的整合，科学数据密集型数据资源的整合，科学数据密集型数据资源的整合，科学数据密集型数据资源的整合，科学数据密集型数据资源的整合，科学数据密集型数据资源的整合，科学数据密集型数据资源的整合，科学数据密集型数据资源的整合，科学数据密集型数据资源的整合，科学数据密集型数据资源的整合。
Collaborations

• PRAGMA
  – www.pragma-grid.net
• EUChinaGrid
  – www.euchinagrid.org Interconnection and Interoperability of Grids between Europe & China
• IGTF / ApGrid PMA
• …
5. e-Science applications

- High Energy Physics
- Astronomy
- Biology
- Natural Resources
- Disaster Reduction
- ...
YBJ-ARGO/AS$\gamma$

- 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.
Transfer, Process and Distribution of Mass Cosmic Ray Data from Tibet

High Speed Network
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

http://goc.grid.sinica.edu.tw/gstat/BEIJING-CNIC-LCG2-IA64/

Monitoring Info on BEIJING-CNIC-LCG2-IA64
VO = World Wide Telescope
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

上千支鱼鸥、棕鸥、斑头雁死亡
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

- Model Database
  - Model Evaluation System
  - Distribute Model

- Model Storage
- Model Verification

- Host data
- Survey data
- Virus data
- Avian trade routes

- Winter Survey Data
- Survey on source
- Predicting

- SDB
<table>
<thead>
<tr>
<th>新闻动态 (565)</th>
<th>学术文献 (228)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IAP Program</strong></td>
<td><strong>Global Natural Hazards and Disaster Reduction</strong></td>
</tr>
<tr>
<td>• 盲目扑杀家禽并不能阻止禽流感的传播</td>
<td>• Avian H5N1 Influenza in Cats.</td>
</tr>
<tr>
<td>• 越南今年12人死于禽流感</td>
<td>• H9N2 chicken influenza viruses display...</td>
</tr>
<tr>
<td>• WHO：关注禽流感新动态</td>
<td>• Molecular determinants within the su...</td>
</tr>
<tr>
<td>• 越南26省发现禽流感 两名女孩已被证实...</td>
<td>• H5N1 influenza: a protein pandemic t...</td>
</tr>
<tr>
<td>• 亚太经合组织将在美举行禽流感研讨会</td>
<td>• Assessment of risk to human health a...</td>
</tr>
<tr>
<td>• 《新英格兰医学杂志》：禽流感病毒可能...</td>
<td>• Public health. Public health risk fr...</td>
</tr>
<tr>
<td>• 泰国研究证实 禽流感可通过人际传播</td>
<td>• Use of the DNA flow-thru chip, a thr...</td>
</tr>
<tr>
<td>• 世卫称可能发生通过人际传播的禽流感</td>
<td>• Reemerging H5N1 influenza viruses in...</td>
</tr>
<tr>
<td>• 泰国又发现禽流感</td>
<td>• Antiviral chemotherapeutic agents ag...</td>
</tr>
<tr>
<td><a href="http://www.avian-flu.info">更新新闻...</a></td>
<td>[更新新闻...]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>最新疫情 (112)</th>
<th>知识介绍 (76)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 越南今年12人死于禽流感</td>
<td>• 人禽共患病之不完全统计</td>
</tr>
<tr>
<td>• 越南26省发现禽流感 两名女孩已被证实感染</td>
<td>• 台湾今年将建成人禽共通传染病信息库</td>
</tr>
<tr>
<td>• 泰国又发现禽流感</td>
<td>• 日本专家：吃经过加工的鸡肉不会感染禽流感</td>
</tr>
<tr>
<td>• 越南禽流感患者增至7人 其中5人已经死亡</td>
<td>• 解读禽流感：禽流感溯源</td>
</tr>
<tr>
<td><a href="http://www.avian-flu.info">更新新闻...</a></td>
<td>[更新新闻...]</td>
</tr>
</tbody>
</table>
6. Cooperation & Communication

• CODATA
  – Secretariat of China CODATA
  – Scientific data & database development and sharing
7. SDB Organization chart

CAS

CNIC

SDB Specialist Committee

SDB Center

SDB Office

Inst. of Botany

Inst. of Geography

Inst. of Zoology

......

Inst. of Microbiology
SDB in 2006 - 2010

“SDB Driving e-Science of CAS”
Framework of CAS e-Science
Technical View of CAS e-Science
-- China Science Grid

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

China Science Grid

- Computing Grid, Data Grid, Storage Grid
- global lambda networks

Sensor Grids, Instrument Grids, ...

Applications

Operation/Support/Policy, ...

e-Science Facility

Infrastructure

web/grid services, data integration, distributed processing, security, ...

portal/PSE, metadata, semantic web, visualization, collaborative tech, ...

Application Grids

China Science Grid
SDB Architecture

- **Sharing Mechanism**
  - Standard
  - Public SD service
  - E-Science oriented SD service
- **Sharing Service**
  - Main Body SDB
  - Motif SDB on domain
  - Special SDB based Key project
- **Operation and management**
  - Technic supporting
SDB Resource Architecture

Main Body SDB

Motif SDB

Subject SDB

special

专业库

专业库

专业库

专业库
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!