Scientific Data Grid on NGI

Kai Nan

Computer Network Information Center
Chinese Academy of Sciences

CANS 2004, Miami

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

Agenda

- Background & History
- Scientific Data Grid
- Next Steps with NGI

Background

- 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, 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.

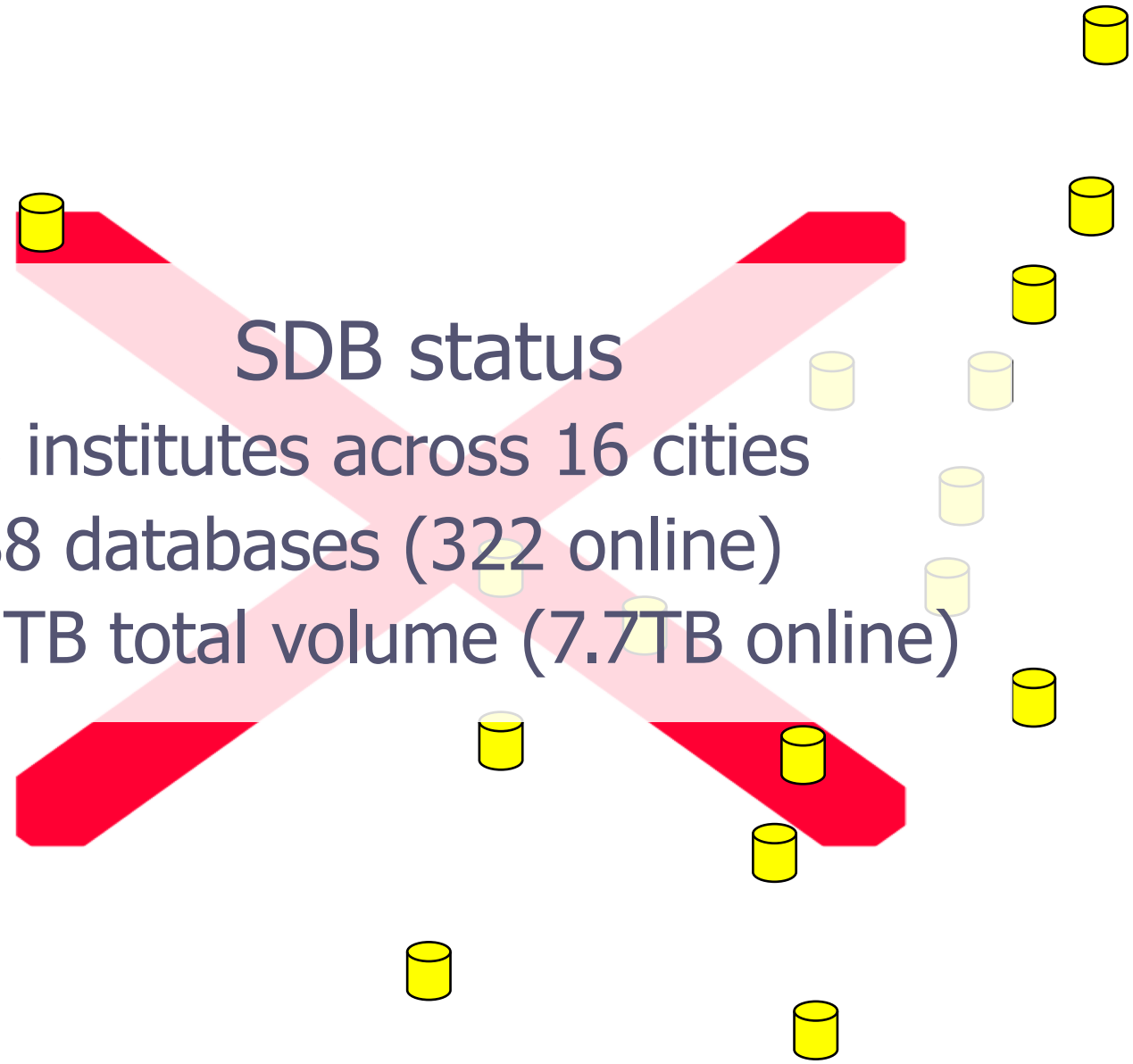
Scientific Database (SDB)

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

- SDB is a long-term project since 1983, 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
- data from research, for research

SDB status

- 45 institutes across 16 cities
- 388 databases (322 online)
- 13TB total volume (7.7TB online)



e-Science

- CAS Informatization Program
  - e-Science and ARP
- Scientific data is one of three poles of the e-infrastructure
  - Networks
  - Computing
  - Data
- SDG is a project of CAS e-Science

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

Milestones

- 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

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

Supported by

- Chinese Academy of Sciences (CAS)
  - Informatization Program 2001-2005
- Ministry of Science and Technology of China (MOST)
  - 863 Program/China National Grid 2002-2005
- National Science Foundation of China (NSFC)
  - Network-based Science and Research Environment (aka. NSFC e-Science) 2004-2006

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

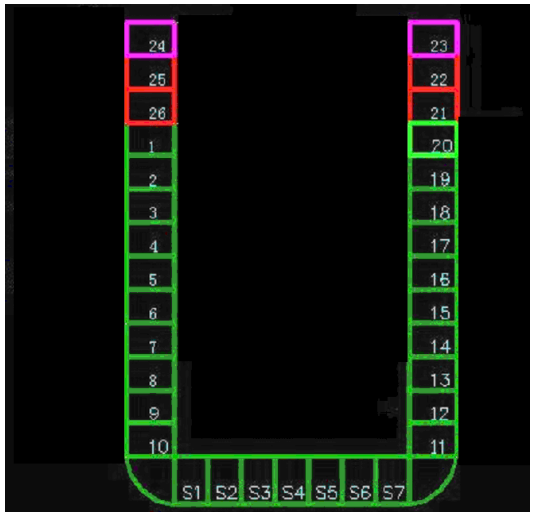
What we do for SDG

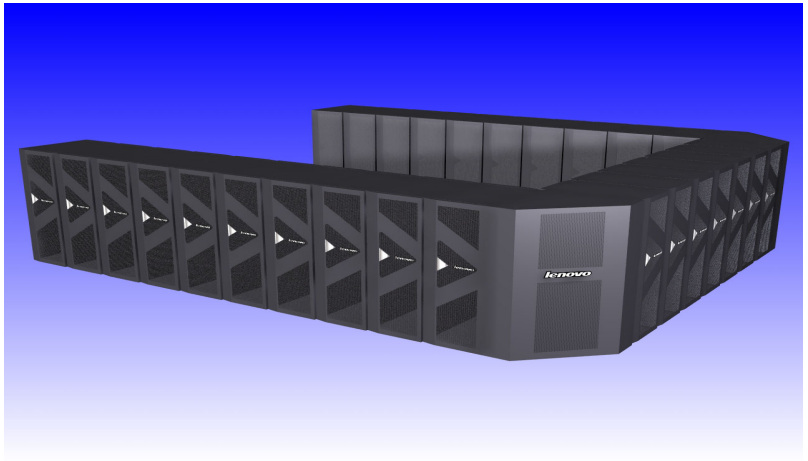
- System Platform
- SDG Middleware
- Demo Applications

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

SDG System Platform

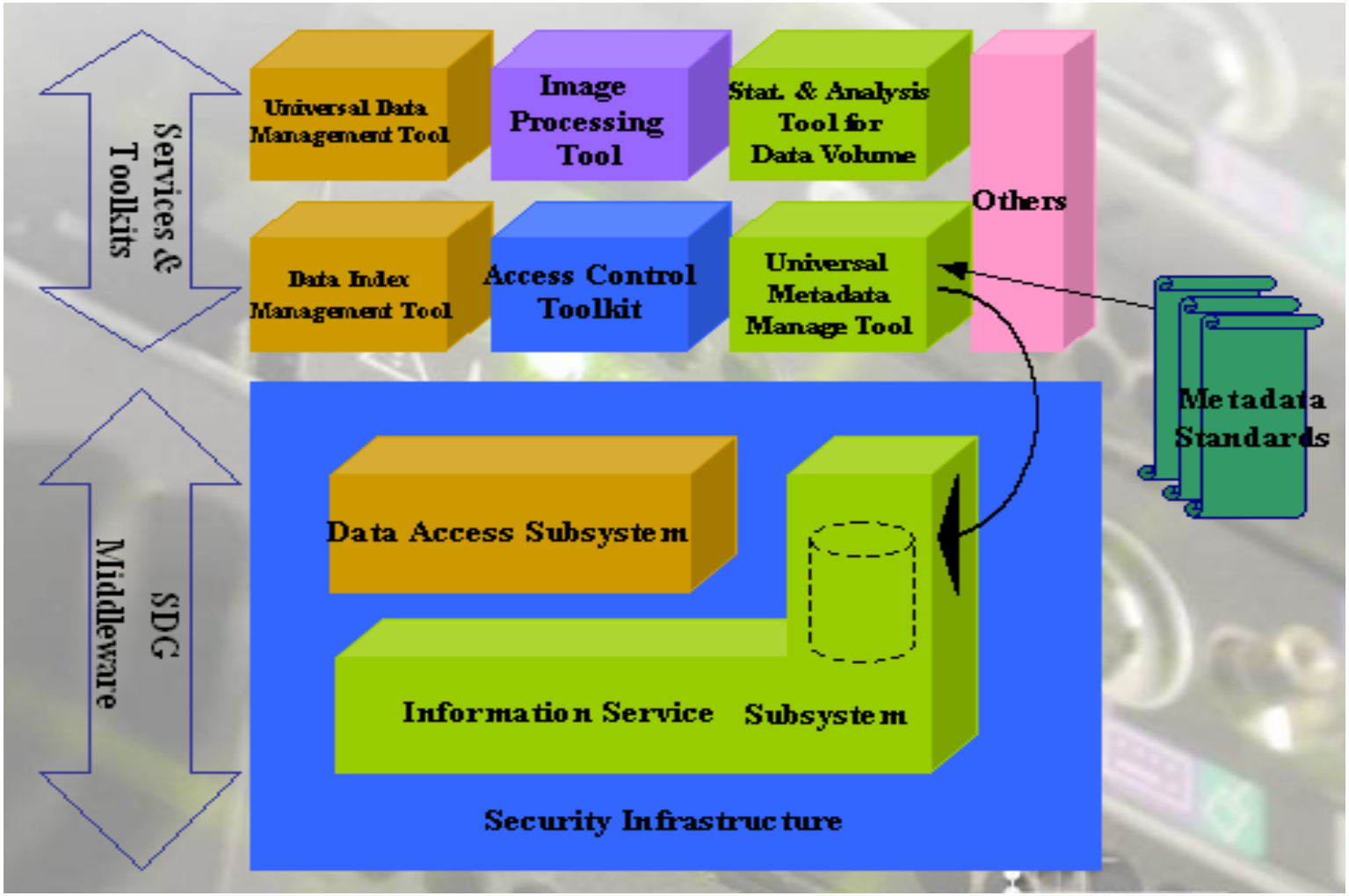
- Data Center
  - 59 nodes of DeepComp 6800
  - SAN Storage
    - 20TB Disk Array
    - 50TB Tape Library
  - TFLOPS-scale computing





中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

SDG Software Modules



SDG Middleware

applications

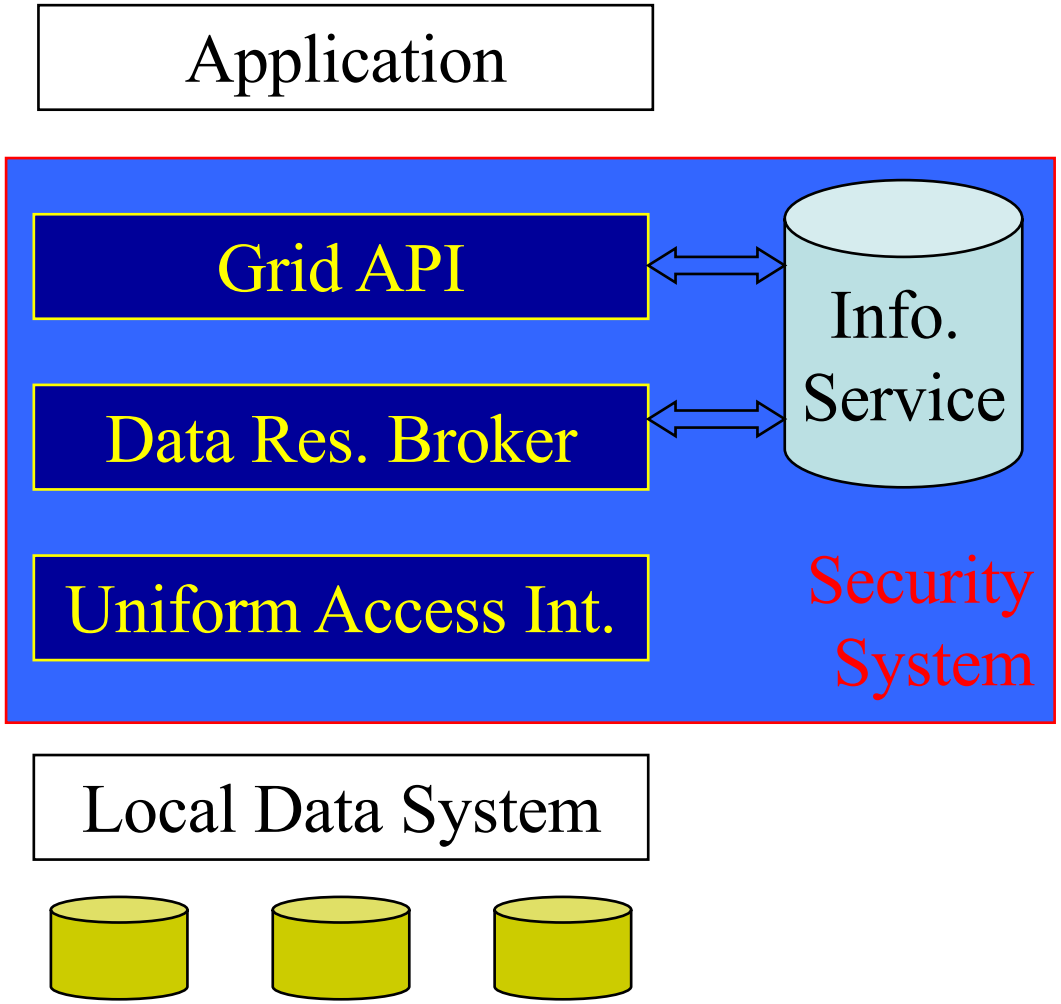
app-oriented, unified program interface

coordinated access to multiple data resources

uniform access interface to single data resource

local data management system, could be DBMS or file system

databases



中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

SDG Middleware and ToolKits

- SDG Middleware
  - **Grid Information System**
  - SDG Data Access System
  - SDG Security System
- SDG Toolkits

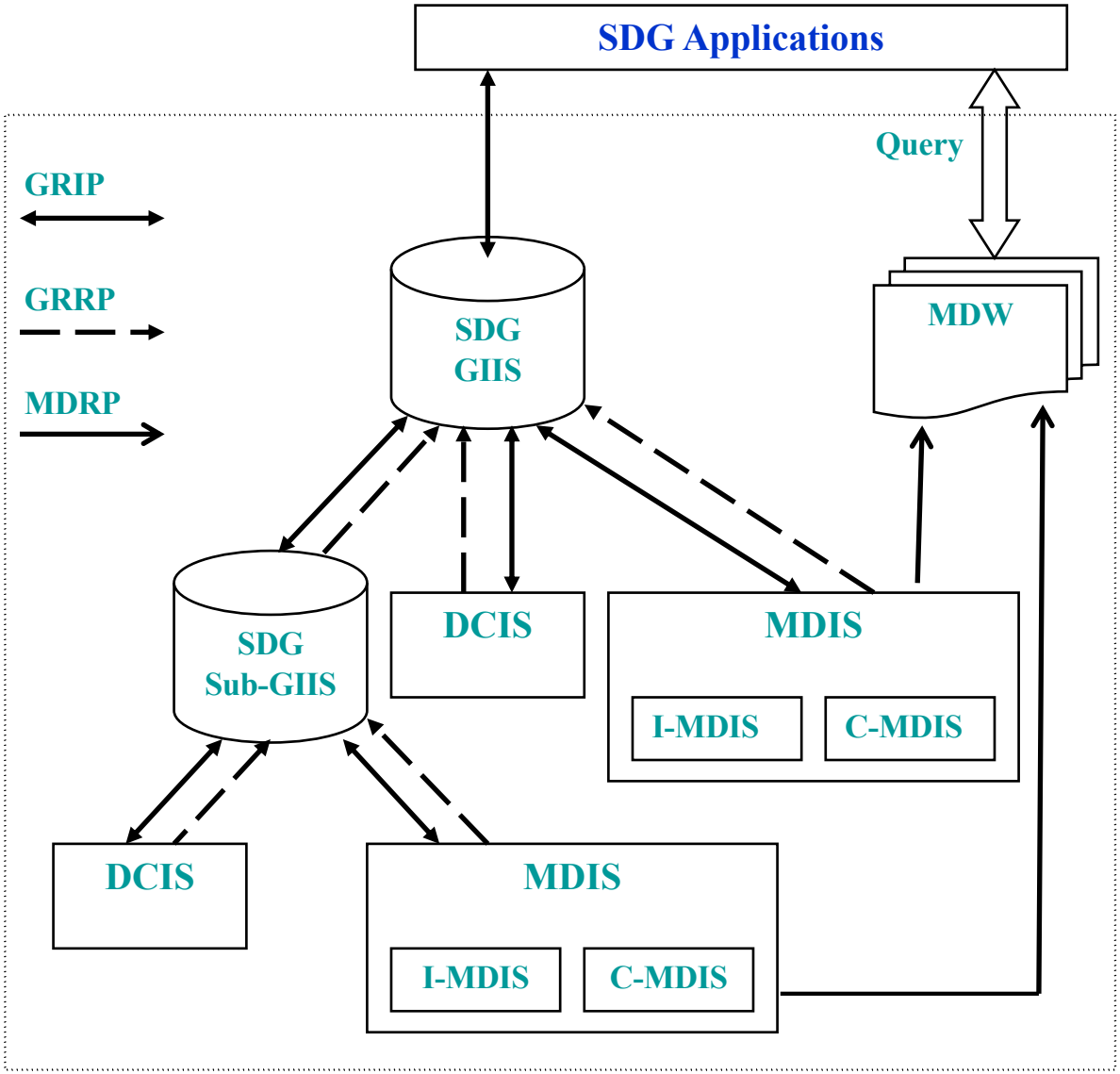
SDG GIS V1.0

Universal Metadata Tool V2.0

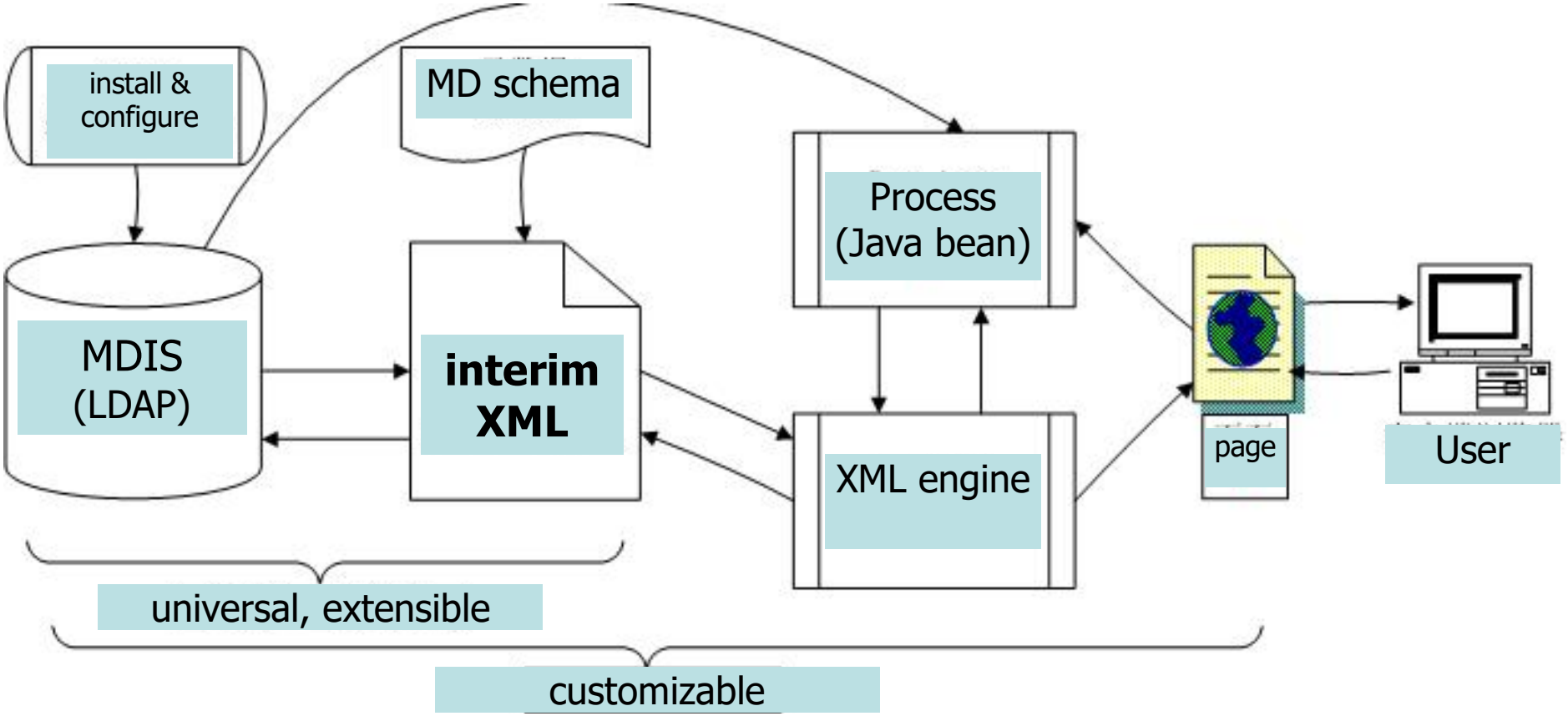
Statistics Tool V1.1

SDG GIS V1.0

- Backend MDS/LDAP
- Two types of Information
  - System info
  - Metadata
- Management and Service
  - Centralized
  - Distributed



SDG Universal Metadata Tool



-metadata is tree-like and more flexible than fix-column tables, difficult to deal with on web UI

-use xml files to store interim results

您的位置： 元数据工具个性化定制

元 数 据 工 具 个 性 化 定 制

查询结果显示界面设定

您的位置： 元数据添加

元 数 据 添 加

首页 | 选择所有元数据 | 用户密码修改 | 单位内部查询元数据 | 添加元数据 | 设定界面风格

您的位置：查询结果

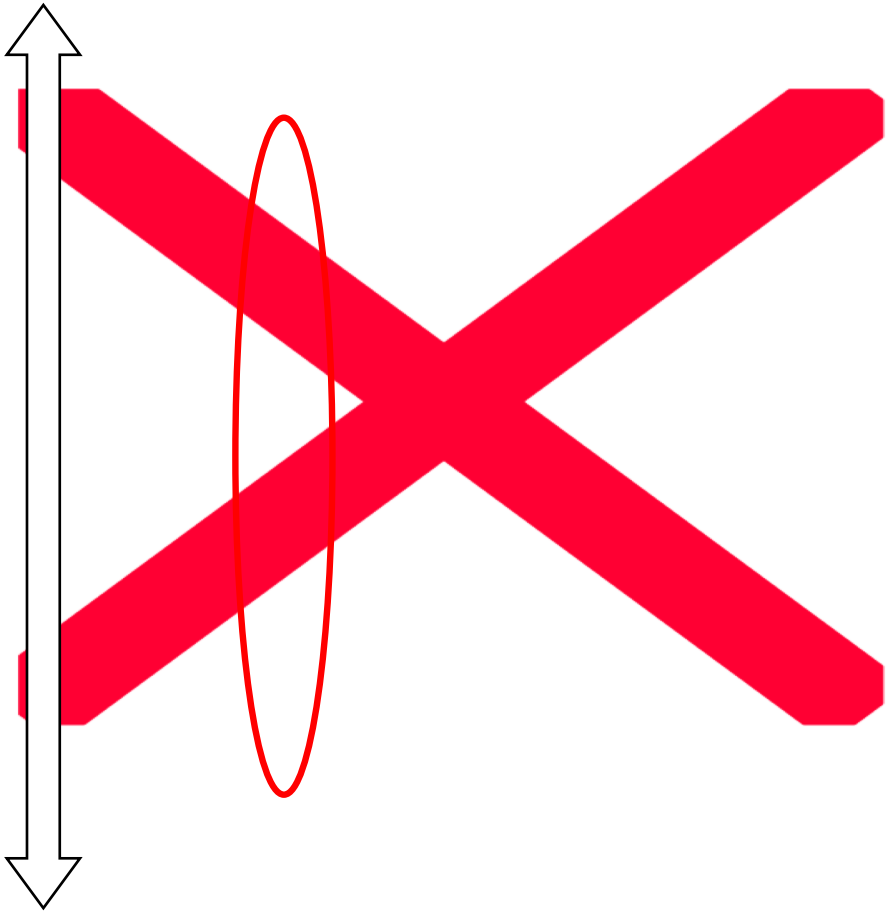
共找到 相关记录 2 条	
ID：1 开发部	**数据集中文名称：国外纳米专利数据库** 详细信息
数据集标识	INF105-SDB-1-33-1
描述	国外纳米专利库的数据收集了从1985年至2001年12月，主要来自美...
数据集提供者	计算机网络信息中心
ID：2 信息部	**数据集中文名称：古代名人库** 详细信息
数据集标识	a-person
描述	古代名人：包括白居易，邓世昌等
数据集提供者	计算机网络信息中心

共1页 第1页 转到 页

返回跨单位元数据查询页面

技术支持：manager@sdb.ac.cn 推荐分辨率：1024*768

Windows 2k/xp
Java 1.4
GT3 Core



Statistics Services

Statistics & Analysis Tool (SAT) for Data Volume

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

- Features
  - Win2000/XP, Linux
  - Java 1.4
  - Globus Toolkit 3 Core
  - Oracle, SQL Server, File System
- Deploy
  - Data nodes: 45 institutes at CAS, across 16 cities in China
  - Mediator: CNIC
  - Service Monitor


Scientific Data Grid Portal - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
后退 搜索 收藏夹 媒体 链接 Norton AntiVirus
Scientific Data Grid Portal
Home | Introduction | Information Service | Security System | Data Access | Application | Links
数据量统计服务
按数据子库名称
请选择数据子库名称
请选择统计信息
提 交 查 询
按数据库名称
请选择数据库名称
请选择统计信息
提 交 查 询
按建库单位名称
请选择建库单位名称
请选择统计信息
提 交 查 询
按学科名称
请选择学科名称
请选择统计信息
提 交 查 询
统计SDB所有数据集
请选择统计信息
提 交 查 询
Computer Network Information Center, Chinese Academy of Sciences. All Rights Reserved
Internet

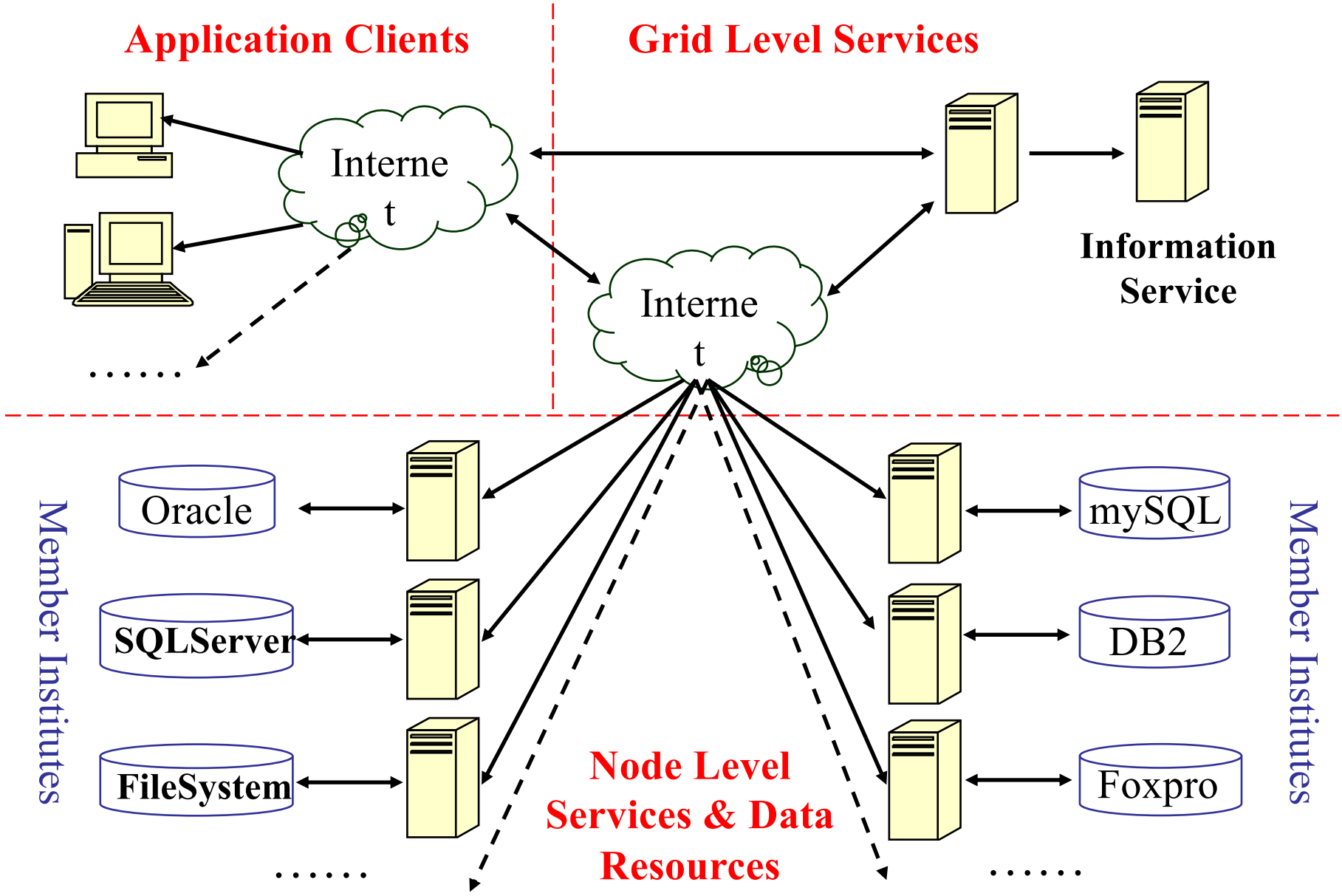
中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

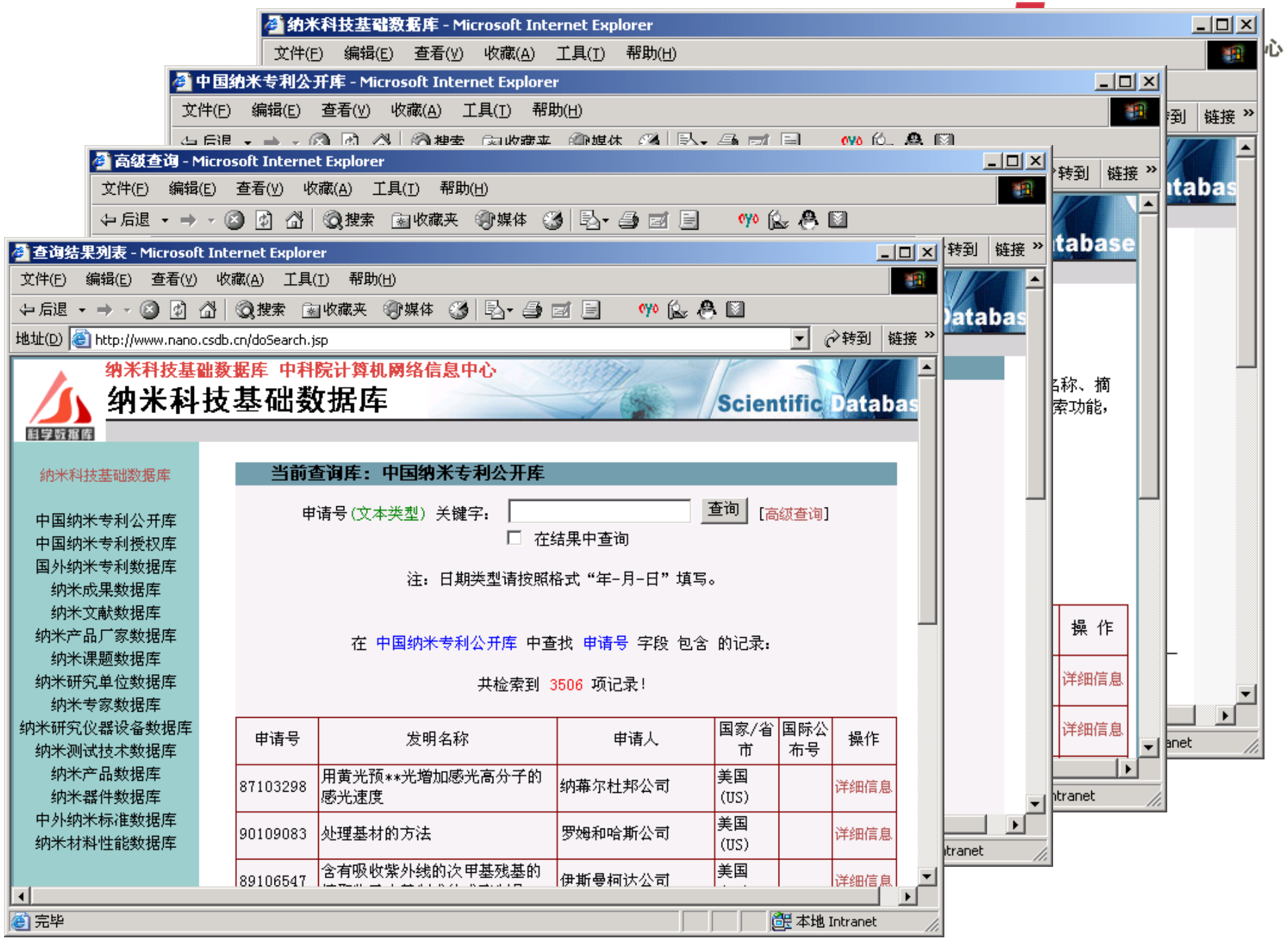
SDG Middleware and ToolKits

- SDG Middleware
  - Grid Information System
  - **SDG Data Access System**
  - SDG Security System
- SDG Toolkits

Data Access Subsystem 1.0

SDG Data Access Service Framework




纳米科技基础数据库 - Microsoft Internet Explorer
中国纳米专利公开库 - Microsoft Internet Explorer
高级查询 - Microsoft Internet Explorer
查询结果列表 - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
后退 搜索 收藏夹 媒体
地址(D) http://www.nano.csdb.cn/doSearch.jsp
转到 链接
纳米科技基础数据库 中科院计算机网络信息中心
纳米科技基础数据库
Scientific Databas
纳米科技基础数据库
中国纳米专利公开库
中国纳米专利授权库
国外纳米专利数据库
纳米成果数据库
纳米文献数据库
纳米产品厂家数据库
纳米课题数据库
纳米研究单位数据库
纳米专家数据库
纳米研究仪器设备数据库
纳米测试技术数据库
纳米产品数据库
纳米器件数据库
中外纳米标准数据库
纳米材料性能数据库
当前查询库：中国纳米专利公开库
申请号 (文本类型) 关键字：
查询
[高级查询]
在结果中查询
注：日期类型请按照格式“年-月-日”填写。
在 中国纳米专利公开库 中查找 申请号 字段 包含 的记录：
共检索到 3506 项记录！
申请号 发明名称 申请人 国家/省市 国际公布号 操作
87103298 用黄光预**光增加感光高分子的感光速度 纳幕尔杜邦公司 美国(US) 详细信息
90109083 处理基材的方法 罗姆和哈斯公司 美国(US) 详细信息
89106547 含有吸收紫外线的次甲基残基的 伊斯曼柯达公司 美国 详细信息
操 作
详细信息
详细信息
完毕
本地 Intranet

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

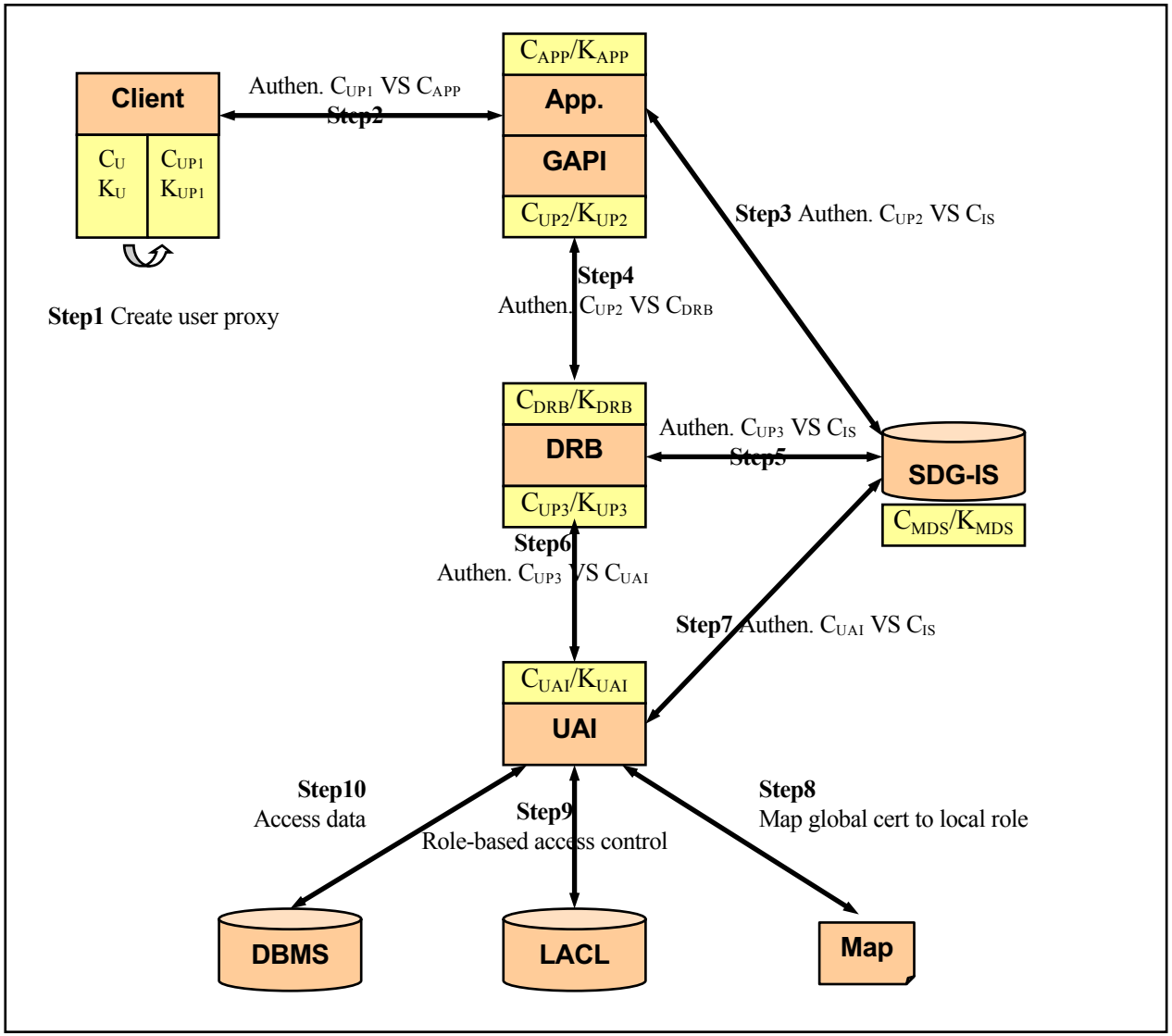
SDG Middleware and ToolKits

- SDG Middleware
  - Grid Information System
  - SDG Data Access System
  - **SDG Security System**
- SDG Toolkits

SDG CA V1.0

Access Control Toolkit V1.1

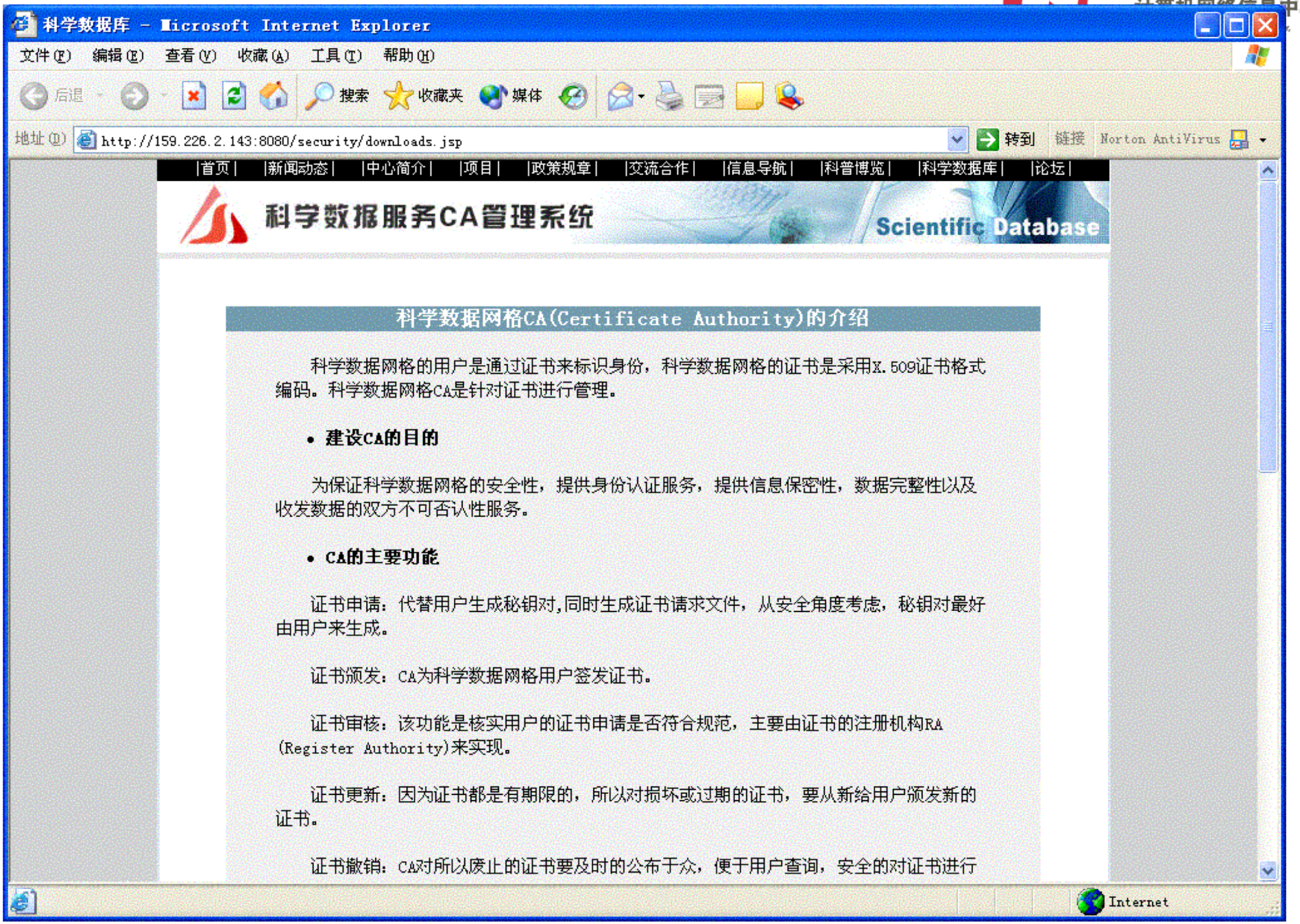
SDG Security System



- GSI based
- Use certificates to identify users
- Role-based local access control

C_X, K_X	X's Cert & Key
UP1, UP2, …	User Proxy, 2nd-level User Proxy, …

Full Process of security-related operations under SDG Security System




科学数据库CA - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
链接
Google
搜索 Web 内容
PageRank
453 个已拦截
选项
科学数据库CA管理系统
用 户 接 口
Scientific Database
导航信息
CA
获取 CA 证书
证书撤销列表
用户
申请证书
获取成功申请证书
测试证书
撤销证书
证书
有效
已到期
已撤销
挂起
申请
证书申请列表
证书撤销申请列表
公共PKI服务
本页面功能：帮助用户提交证书、申请以及查看证书相关信息。
初始化
获取 CA 证书
[CA证书导入浏览器]
证书管理
申请一个证书
[适用于证书申请]
获取成功申请的证书
[获取证书导入浏览器]
测试证书
[测试您的证书]
证书撤销申请
[撤销一个证书]
列表证书信息
证书撤销列表
[CRL 下载页面]
关于我们
联系方式
版权信息
网站地图
版权所有 中国科学院计算机网络中心 Copyright 1995-2004
Internet

SDG Middleware and ToolKits

- SDG Middleware
  - Grid Information System
  - SDG Uniform Access Interface
  - SDG Security System
- **SDG Toolkits**

SDG Portal

Image Process Tool 1.0

Storage Sharing Service

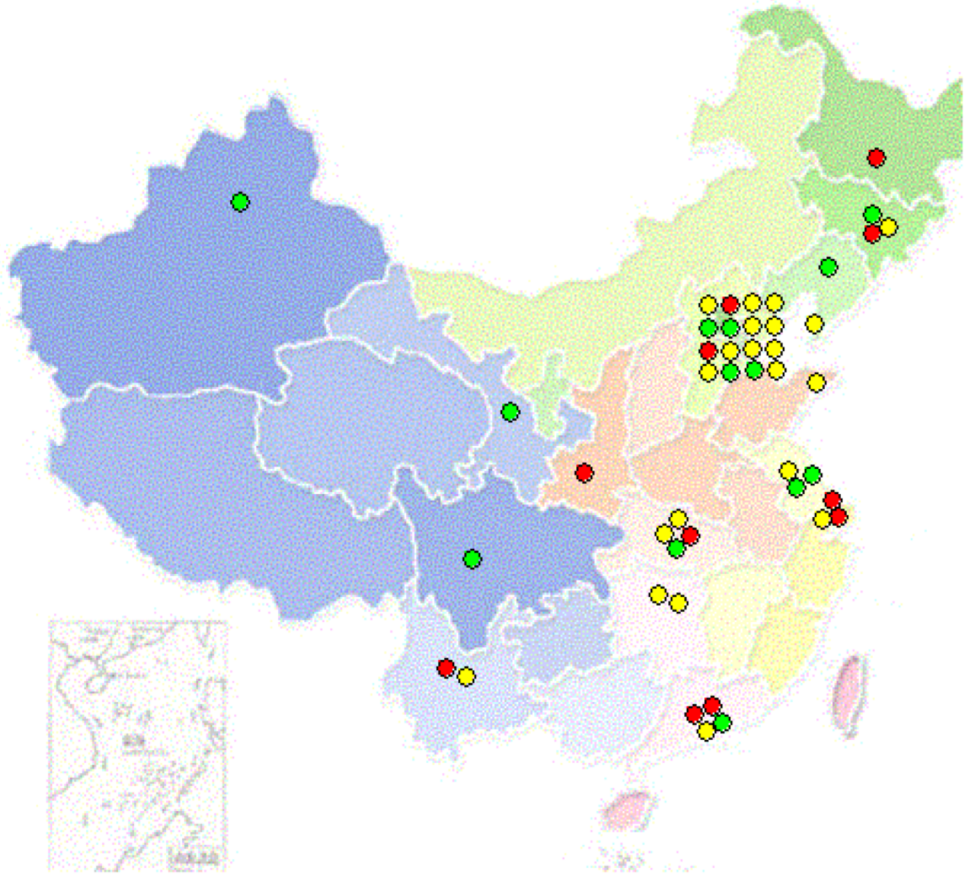
Scientific Data Grid Portal - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

地址(D) http://159.226.2.184:8080/sat/portal/index_en.htm 转到 链接 Norton AntiVirus

Scientific Data Grid Portal

Home | Introduction | Information Service | Security System | Data Access | **Application** | Links



Scientific Data Grid (SDG) is an application grid which focuses on scientific data resources sharing and collaboration. Its goals include connecting and integrating data resources in more than 40 institutes under Chinese Academy of Sciences, enhancing effective sharing of distributed and heterogeneous data in Scientific Database

Normal Service
Abnormal Service
No Service
Refresh on the hour

小应用程序 TestApplet started

Internet

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

Demo Applications

- China Virtual Observatory
  - National Astronomical Observatory, CAS
  - Grid Services wrapping up astronomical data and code
  - quite a few services ready now
- HEP
- …

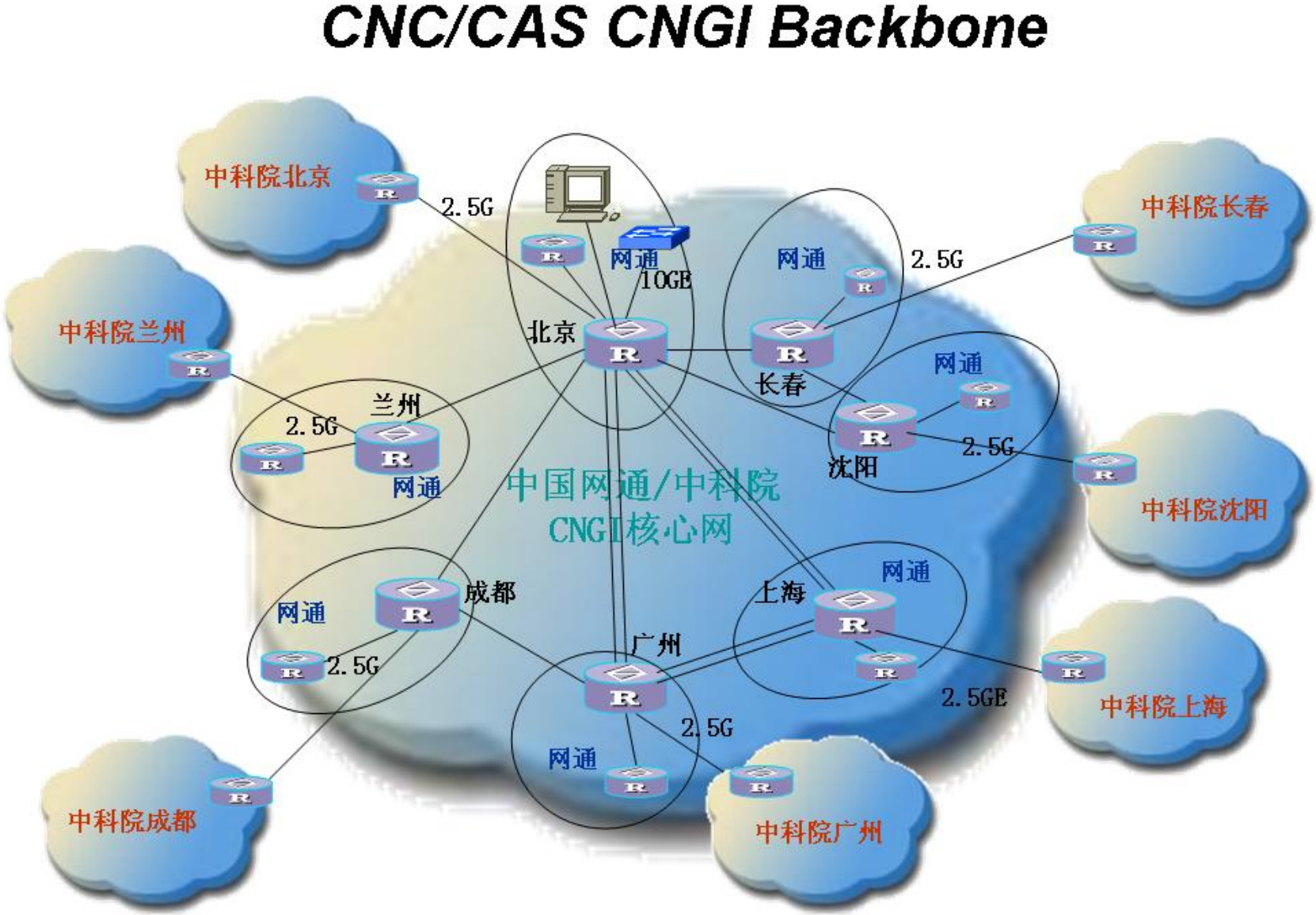
Training and Deployment

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

- SDB Technical Training
  - more than 100 participants
  - once a year
- SDB Work Evaluation Online
  - important to impel deployment of SDG middleware
- Distance training with partners
  - CNIC-UCSD/SDSC, February 2004

PRAGMA

SDG - good application on CNGI



Next Steps with NGI

- With higher bandwidth (e.g.. CNGI)
  - Mass data transmission, better data services
  - Data intensive applications using distributed superservers (not easy now)
  - Share data securely (often underestimated)

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

Next Steps with NGI

- 11th five-year Program (2006-2010)
  - to build some subject data centers, which are well connected by NGI
  - to run SDG system platform routinely
  - to get SDG Middleware aware of NGI, and steady
  - to expand SDG beyond CAS
  - to develop more *real* science applications

Thank you!