

CADSL: China Archives of Digital Scientific Literature Initiative



Xiaolin Zhang

National Science Library, Chinese Academy of Sciences

<http://www.nstl.gov.cn/> zhangxl@mail.las.ac.cn

2006.10.10

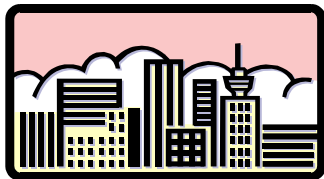


China Archives of Digital Scientific Literature

- 1. Background
- 2. CADSL Initiative Framework
- 3. CADSL Requirements
- 4. CADSL Implementation and Future Efforts

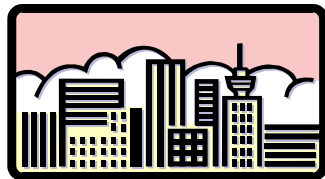
China Archives of Digital Scientific Literature

- 1. Background:
- 1.1-1 Basic Info: Chinese Libraries



Public Libraries
3,000

M of Culture



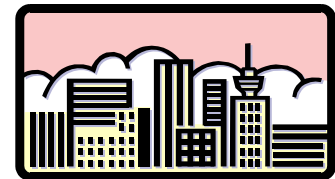
Academic Libraries
1,700

M of Education



Special Libraries
1,500

M of S&T, CAS,
Other Ministries
Enterprises, etc.



School Libraries
5,000

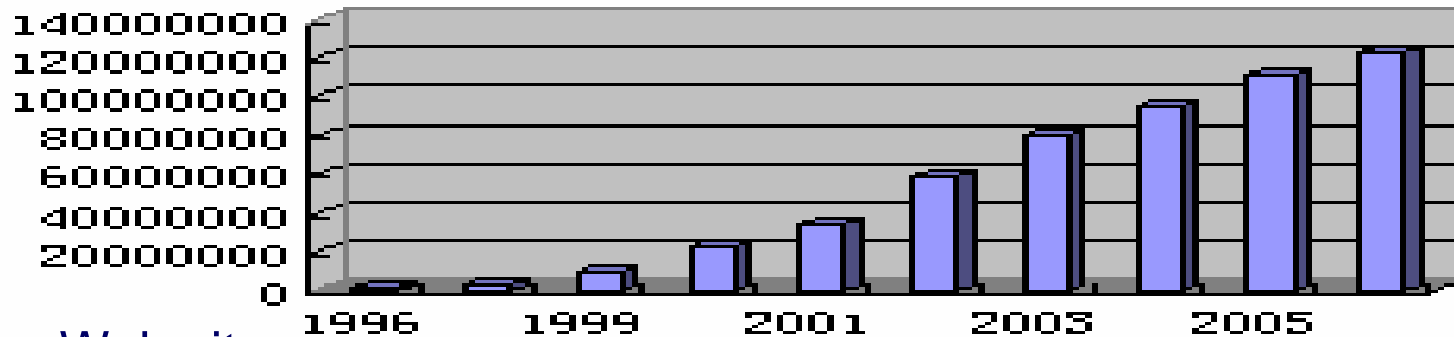
Various Cities

China Archives of Digital Scientific Literature

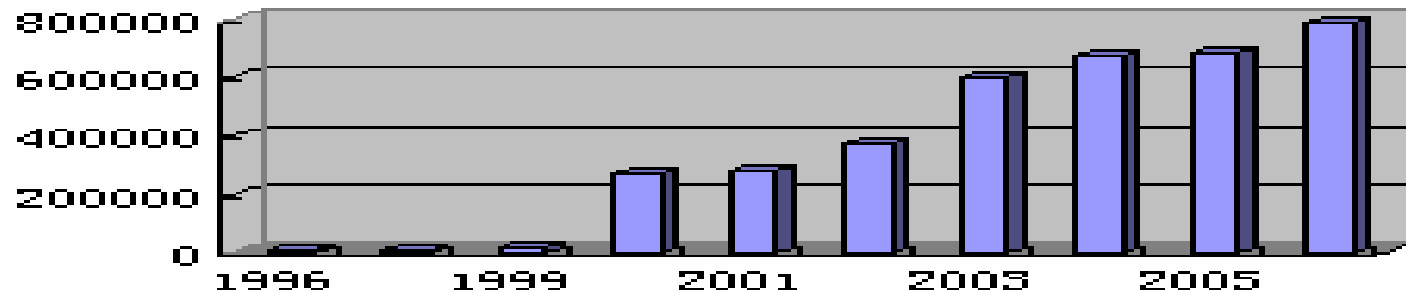
- 1. Background

- 1.1-2 Basic Info: Internet development in China

- Internet Users



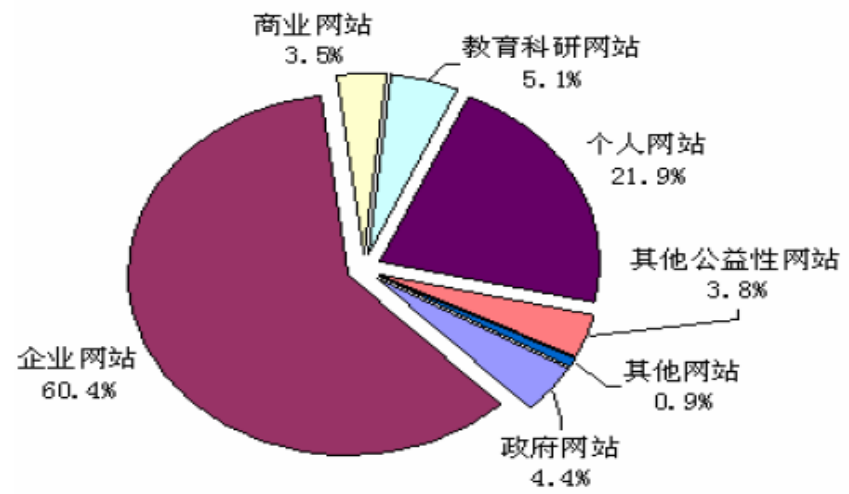
- Web sites



- CNNIC China Internet Report (06/07/01) <http://www.cnnic.cn>

China Archives of Digital Scientific Literature

- 1. Background
- 1.1-3 Basic Info: Internet resources in China
- CNNIC 2005 China Internet Resource Report
 - Web sites 694000
 - Enterprises 60.4%
 - Personal 21.9%
 - Research & Ed 5.1%
 - Government 4.4%
 - Public Interest 3.8%
 - Sales 3.5%
 - Others 0.9%
 - Web pages 24000 Millions, 63932 GB
 - Online Databases 295000





China Archives of Digital Scientific Literature

- **1. Background**
- **1.2-1 Digital Info Resources: Foreign**
 - CALIS (Chinese Academic Info and Lib System, MoE)
 - 300 major academic libraries
 - Databases 250
 - e-journals 24000, e-books 130,000, ETD 133,000
 - CSDL (Chinese Science Digital Library, CAS)
 - Databases 53
 - e-journals 3300, e-proceedings 8000, ETD 133,000
 - NLC (National Library of China, MoC)
 - Databases 56, e-journals 6000
 - NSTL (National Science & Technology Library, MoST)
 - National License e-journals 93, E-proceedings 15,000 series
 - Other research, academic, and public libraries
 - Limited additional specific databases



China Archives of Digital Scientific Literature

- **1. Background**
- **1.2-2 Digital Info Resources: Domestic**
 - CALIS (Chinese Academic Info and Lib system)
 - Databases over 30
 - e-journals 12000, e-books 1,000,000, ETD 225,000
 - CSDL (Chinese Science Digital Library)
 - Databases 10
 - e-journals 12000, e-books 80,000, ETD 30,000
 - NLC (National Library of China)
 - Databases 32, e-books 120,000, e-journals 9000
 - NSTL (National Science and Technology Library)
 - E-journals 4590, e-books 200,000, ETD 130,000
 - Other research, academic, and public libraries
 - Various additional specific databases



China Archives of Digital Scientific Literature

- **1. Background**
- **1.3-1 Challenges: to sustainable use**
 - Heavy reliance on foreign information resources in S&T
 - Download 68m articles by CALIS 2005
 - Download 11 m articles by CSDL 2005
 - Heavy reliance on digital information resources in S&T
 - Mainstream resources and main means of access
 - Very limited preservation mechanisms exist
 - Many databases as network access only
 - Some mirror sites running as publisher-owned proxies
 - Some perpetual copy stored as dark archives
 - Only handful institutions with preliminary preservation policies and mechanisms

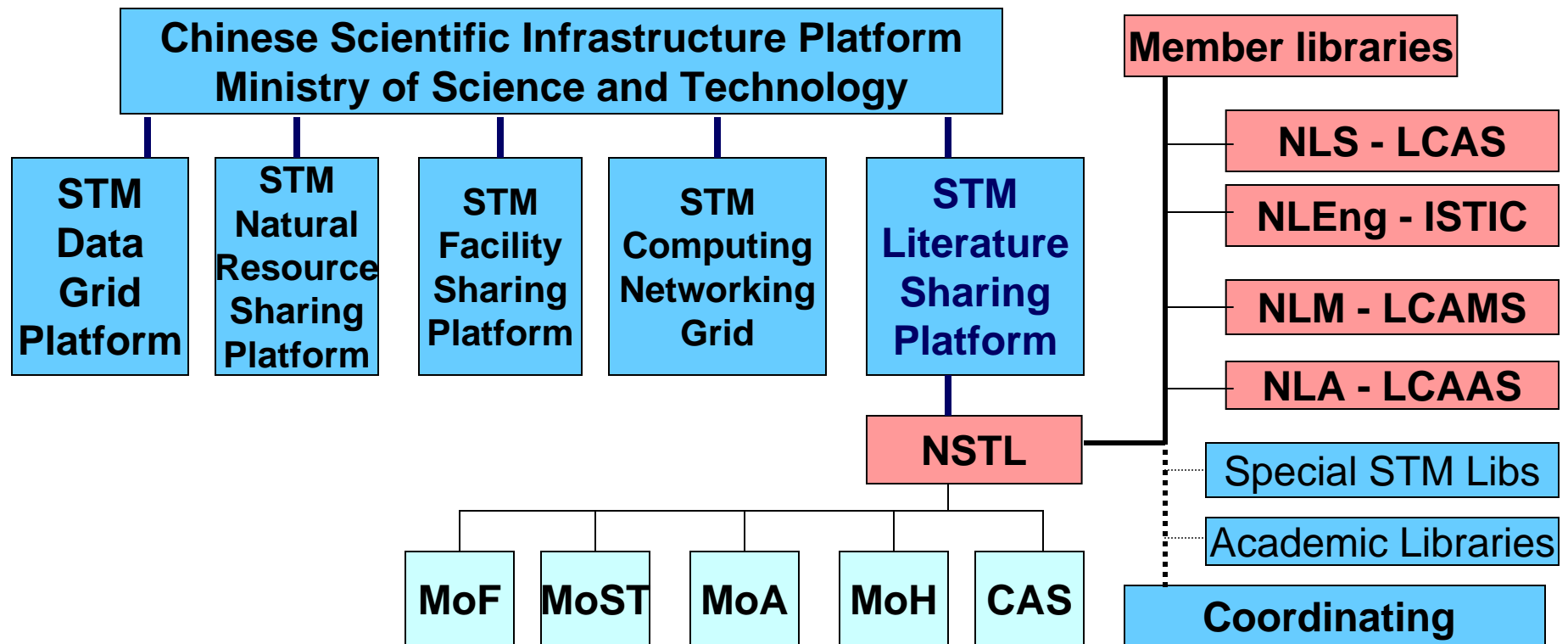


China Archives of Digital Scientific Literature

- **1. Background**
- **1.3-2 Challenges: to preservation planning**
 - Fragmented admin and funding for digital resources
 - CALIS by MoE and universities
 - CSDL by CAS and MoST
 - NSTL by MoST
 - Other digital library projects by MoC, MoH, MoA, etc., and local governments
 - Lack of systematic approach to digital preservation
 - Preservation planning, IPR management, repository certification, workflow management, service provision, technological interoperability, and costing assessment
 - Realization for the urgent need for digital preservation
 - Need for a collaborative planning
 - Need for a coordinated implementation

China Archives of Digital Scientific Literature

- 1. Background
- 1.4-1 NSTL study: NSTL as the national platform





China Archives of Digital Scientific Literature

- 1. Background
- 1.4-2 NSTL study: **the feasibility study**
 - A national approach
 - Needs and crisis analysis
 - Review of current problems and developments
 - Exploration of specific issues in
 - Digital preservation planning with distributed responsibilities
 - IPR management
 - Repository certification, evaluation, and administration
 - Preservation management including workflow, service agreement, technical interoperability, etc.
 - Funding mechanisms
 - Testing of archive systems for DP
 - Development of implementation plans
 - Development of CADSL Initiative Plan to MoST



China Archives of Digital Scientific Literature

- 1. Background
- 2. CADSL Initiative Framework
- 3. CADSL Requirements
- 4. CADSL Implementation and Future Efforts



China Archives of Digital Scientific Literature

- 2. CADSL Initiative Framework

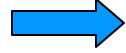
- 2.1 Research questions of the feasibility study

Preserving for whom?



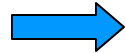
Target the user groups

Preserving what?



Prioritize the resources

Who preserves the resources?



Design the responsibility network

How to preserve the resources?



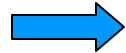
Develop the requirements (IPR, technical, workflow, service, ...)

How to manage the processes?



Implement administrative means (funding, certification, fail-safe...)

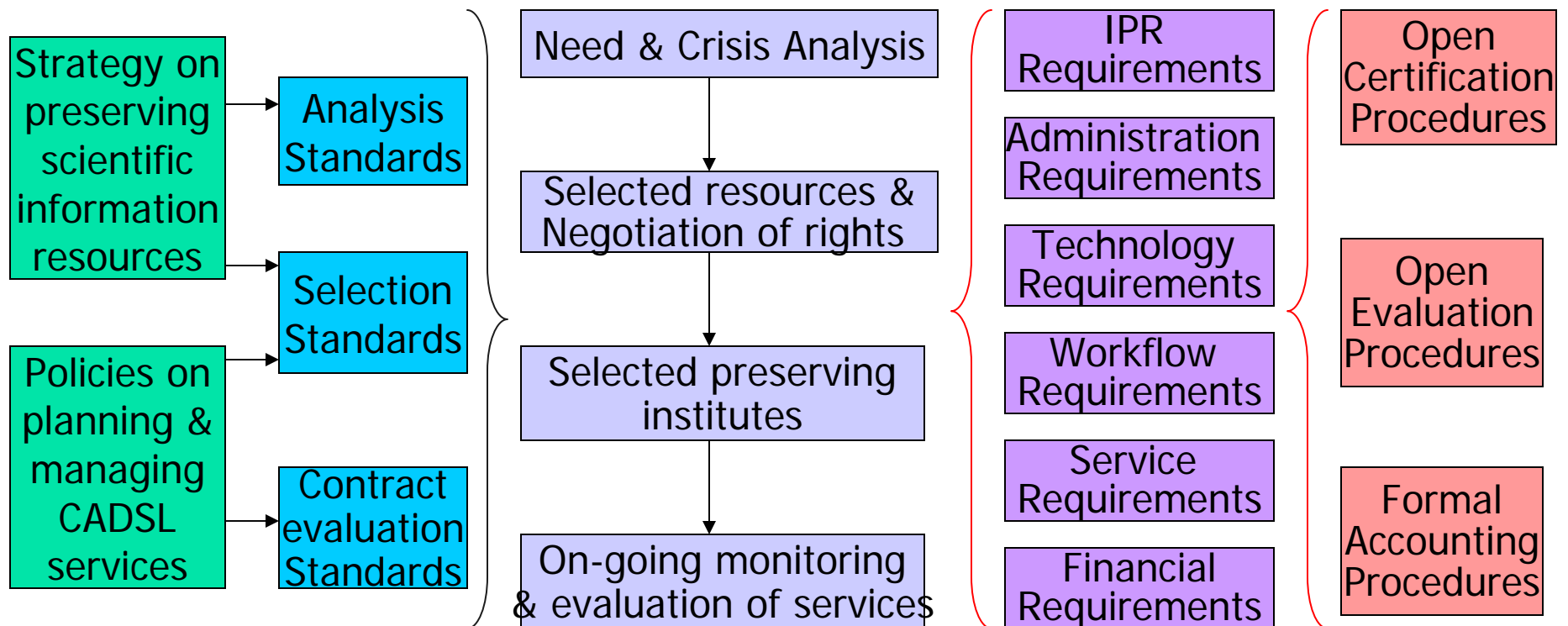
How to sustain the processes?



Institutionalize long-term policy & Planning mechanism

China Archives of Digital Scientific Literature

- 2. CADSL Initiative Framework
- 2.2 General Framework of CADSL



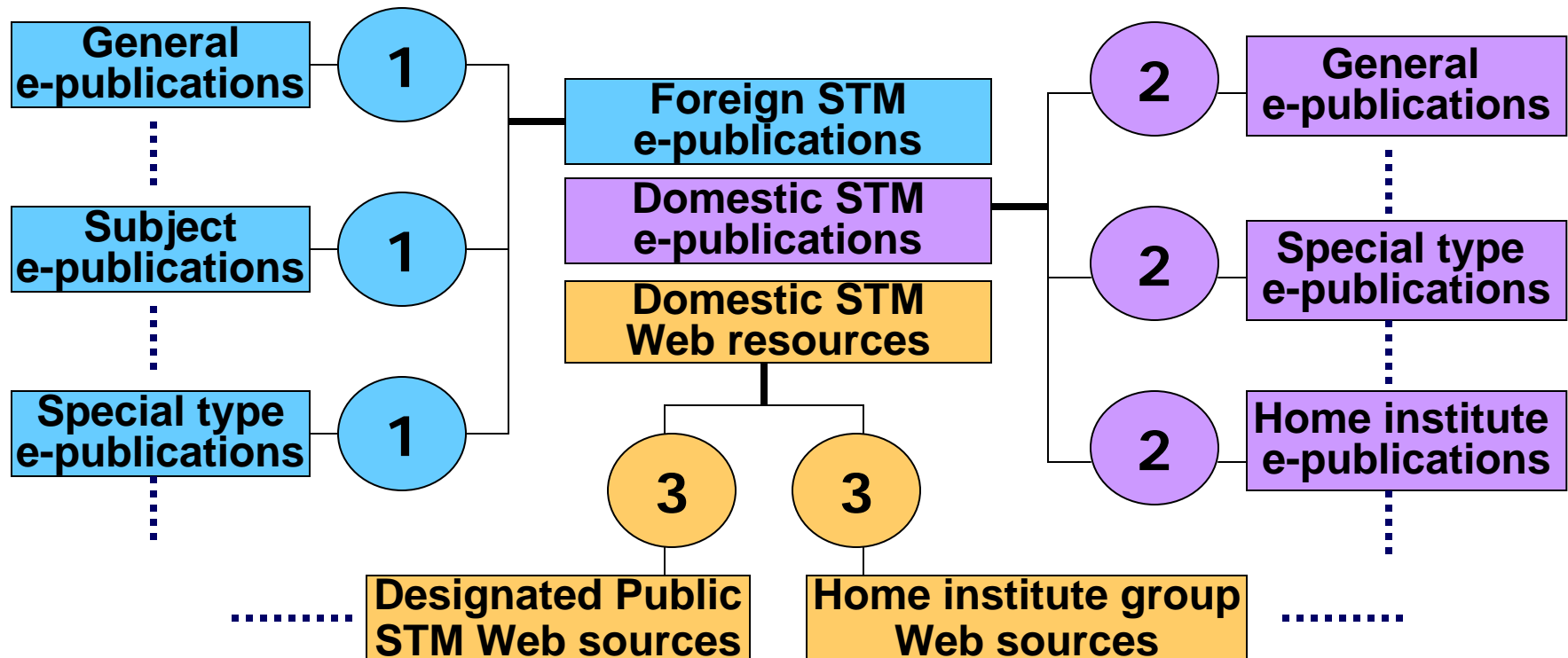


China Archives of Digital Scientific Literature

- **2. CADSL Initiative Framework**
- 2.3-1 Basic targets of CADSL
- CADSL target user groups
 - scientific research and educational communities
- CADSL prioritized information resources
 - Formal scientific scholarly publications and key web resources
 - Foreign scientific scholarly R&D publications
 - E-journals, e-proceedings, ETDs
 - Chinese scientific scholarly R&D publications
 - E-journals, e-proceedings, ETDs
 - Key Chinese scientific scholarly web resources

China Archives of Digital Scientific Literature

- **2. CADSL Initiative Framework**
- 2.3-2 Basic structure of CADSL
- An extensible network of trusted archives





China Archives of Digital Scientific Literature

- **2. CADSL Initiative Framework**
- **2.3-3 Distributed and shared responsibilities**
 - A system of shared responsibilities by major STM libraries
 - Designated by NSTL as national archives
 - National libraries or library systems for general STM e-publications
 - Statutory or historical responsibilities
 - Current mirror site responsibilities
 - National/major research libraries for subject-based or special-type STM e-publications
 - Statutory or historical responsibility
 - Acquisition and service of the resources
 - Service responsibility to primary user groups
 - STM libraries for formal home e-publications (e-journals and ETDs)
 - Some will act as proxies for other institutes' e-publications
 - Flexible structures to suit operational realities
 - Accommodation of distributed nature of library governance
 - Utilization of current system and administrative structure
 - Built-in fail-safe backups with overlapping responsibilities



China Archives of Digital Scientific Literature

- **2. CADSL Initiative Framework**
- 2.3-4 Administrative Mechanisms of CADSL
 - Supported by MoST
 - As a part of National Scientific Infrastructure Platform and NSTL
 - With collaboration from NCFC, MoE, MoC, CAS, MoA, MoH, CAST
 - Organized by NSTL
 - Status: as a standing program of funding and management
 - Administration:
 - **A planning committee** with major stakeholders
 - **An expert group** to be consulted with policy and op issues
 - **A CADSL Center** within NSTL to oversee the operation
 - Participation:
 - **A competitively selected group of member institutions** designated as trusted operators of archives within the network
 - Management:
 - **Standard procedures** to select, certify, fund, oversee, and evaluate trusted archive operator institutions



China Archives of Digital Scientific Literature

- **2. CADSL Initiative Framework**
- 2.3-5 Examples of CADSL archives
 - CALIS
 - For major STM commercial/general publisher databases, and in the future for major learning resources in STM
 - Designated by NSTL as national archives
 - Operated by Beijing University, Tsinghua University, etc.
 - CAS
 - For major societal publisher databases in basic sciences & geo-environmental sciences
 - CAMS
 - For major societal publisher databases in biomedical sciences
 - CAAS
 - For major societal publisher databases in agricultural sciences
 - ISTIC
 - For major societal publisher databases in biomedical sciences
 - Chinese STM e-journals



China Archives of Digital Scientific Literature

- 1. Background
- 2. CADSL Initiative Framework
- 3. CADSL Requirements
- 4. CADSL Implementation and Future Efforts



China Archives of Digital Scientific Literature

- **3. CADSL Requirements**
- **3.1 Administrative requirements**
 - Institutional commitment to the responsibility and input of resources
 - Written strategy and policy statements from the institution
 - Public service orientation
 - Shared and collaborative efforts
 - Long-term and sustainable operations
 - Coordinated with institutional preservation strategies & policies
 - Formal contracts with CADSL/NSTL
 - Certification
 - Periodic evaluation
 - Institutional administration for preservation operations
 - Formal administrative structure with high level authority involvement and overseeing mechanism within the institution
 - Formal and “independent” staffing and budgeting
 - Collaborating mechanisms with CADSL system
 - Open institutional management procedures for the operations



China Archives of Digital Scientific Literature

- **3. CADSL Requirements**
- **3.2-1 IPR requirements**
 - Complicated relationships among CADSL
 - Content owners, Preserving archives, User communities
 - Rights to ingest
 - Requiring transfer of perpetual copies with full data with intervals frequently enough
 - Requiring full descriptions of data and publisher's support in ingesting and managing the data
 - By contracts with content providers specifying the rights of perpetual copies for national archives
 - By consortia negotiations fully using market power
 - By Instituting public investment requirements by perspective ministries



China Archives of Digital Scientific Literature

- **3. CADSL Requirements**
- **3.2-2 IPR requirements**
 - Rights to process for preservation
 - Necessary re-formatting, re-organizing, transformation, etc., of the data during ingest, archiving, and servicing
 - Necessary handling for preservation of copyright management information and IPR technical measures
 - Rights to serve
 - When to serve and serving whom
 - By contracts with content providers for consortium or public services when situations warrant
 - By contracts with archiving institutions for service level agreements



China Archives of Digital Scientific Literature

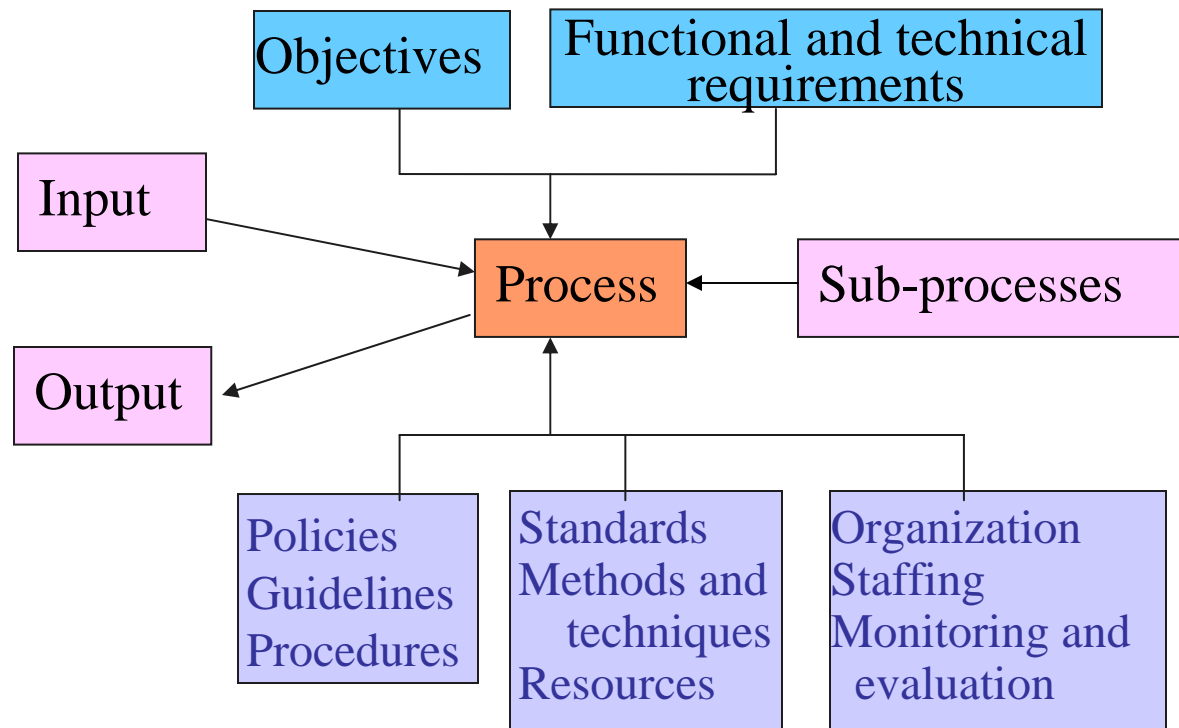
- 3. CADSL Requirements
- 3.3 General technical requirements
 - Ensuring serviceable archives
 - Not necessarily parallel, but equivalent in main content & functions
 - Full scale serviceability within limited time periods
 - Functionability for the whole life-cycle of preservation
 - Open to all ingested materials and ingesting environment
 - Ensuring effectiveness, trustability, and interoperability of the archives
 - Following the OAIS model
 - Using OSS-based and well-tested, well populated systems
 - Adopting standard AIP formats and metadata formats
 - Supporting well-formed workflows of ingesting, archiving, long-term managing, and servicing
 - Utilizing standard S&R interfaces
 - Implementing standard modules for archive operation monitoring and management

China Archives of Digital Scientific Literature

- **3. CADSL Requirements**
- 3.4 Workflow management requirements
 - Trustable workflow management for the whole system and for each of the processes

What, in what format, how much and often, what/how to control, what possible problems

What, in what format, how much and often, what/how to control, what possible problems, to what process





China Archives of Digital Scientific Literature

- **3. CADSL Requirements**
- **3.5 Service management requirements**
 - Sustainability of the shared responsibility relies on well-performed services to other members and the public
 - Contract with CADSL/NSTL for service level agreement
 - Specifications for acceptable service levels
 - Performance standards for archiving
 - Performance standards for services
 - Response time for setting-up of the service when needed
 - Service functionality
 - Service reliability
 - Administrative, financial and legal liabilities of the archiving institutions
 - Open management
 - Open evaluation



China Archives of Digital Scientific Literature

- **3. CADSL Requirements**
- 3.6 Certification and evaluation requirements
 - A certification process as a prerequisite for CADSL
 - An Independent certification process authorized by MoST/NSTZL with outside experts
 - A comprehensive set of certification standards for organizational, administrative, technological, operational, legal, and financial aspects, utilizing international experiences
 - A combination of certification and evaluation processes
 - A standing procedure and budget for certification and evaluation
 - Open certification and evaluation standards
 - Open certification and evaluation processes
 - Enforcement mechanisms for evaluation results



China Archives of Digital Scientific Literature

- **3. CADSL Requirements**
- 3.7 Funding management requirements
 - Public responsibility of the government to preserve
 - MoST funding for major initial development of CADSL, continuing extension and improvement, and part of continuing operation
 - Shared responsibility of the participating institutions
 - Institutional funding for part of the initial development of the assigned archives and for continuing maintenance (staff, facilities, administration, ...)
 - Multiple incentives for supporting the participation
 - Shared preservation by others for resources needed
 - Combined with MoST support of resource acquisition
 - Investment by MoST on systems, personnel, expertise, etc.
 - Specific funding management at NSTL and archiving institution levels



China Archives of Digital Scientific Literature

- 1. Background
- 2. CADSL Initiative Framework
- 3. CADSL Requirements
- 4. CADSL Implementation and Future Efforts



China Archives of Digital Scientific Literature

- 4. CADSL Implementation and Future Efforts
- 4.1 Phased implementation
 - Phase 1 (6 months):
 - Limited scale experimental systems with
 - 1 general foreign full-text database
 - 2-3 subject-based foreign full-text database
 - 2-3 special type foreign databases
 - 1 Chinese full text database
 - Development of standards, procedures, and best practices
 - Development of certification and evaluation procedures
 - Phase 2 (6m-1y):
 - Full scale operations with the experimental systems
 - Limited scale experiment systems with other special and key resources
 - Full scale development of CADSL management
 - Phase 3 (2-3yrs)
 - Inclusion of all important STM resources with fully operational CADSL management



China Archives of Digital Scientific Literature

- 4. CADSL Implementation and Future Efforts
- 4.2 Mobilizing scientific communities
 - Promotion of the need and crisis for preservation
 - Advocacy of shared responsibility
 - Education of trusted preservation services and preservation management
 - Development of funding formulas and incentives
 - Promotion of star experimental archiving institutions
 - Working together with major library systems
 -



China Archives of Digital Scientific Literature

- 4. CADSL Implementation and Future Efforts
- 4.3 Future efforts
 - Collaborations
 - With scientific data preservation
 - With humanities and social sciences resource preservation
 - With cultural heritage preservation
 - With institutional knowledge assets management
 - With international partners in scientific communities
 - Policies
 - Development of Institutional policies
 - Coordination of institutional policies with national and CADSL policies
 - Incorporation of preservation into the organizational structure and management process of information institutions
 - Services
 - Improvement on technologies
 - Development of integration with other services and processes
 -



China Archives of Digital Scientific Literature

Thanks!

谢谢！

