

Architecture and Development of CAS IR GRID

**Zhu Zhonming, Ma Jianxia
Lanzhou Branch of NSL, CAS
Zhang Xiaolin, Zhang Zhixiong
NSL, CAS**

German – Chinese Symposium
Oct. 2008

Hannover, Germany

Outline

- **Background and Some Considerations**
- **System Architecture of CAS IR Grid**
- **Two-phased Pushing-up strategies**
- **Lessons and Experiences from Implementation**

Outline

- **Background and Some Considerations**
- **System Architecture of CAS IR Grid**
- **Two-phased Pushing-up strategies**
- **Lessons and Experiences from Implementation**

1. Background and Some Considerations

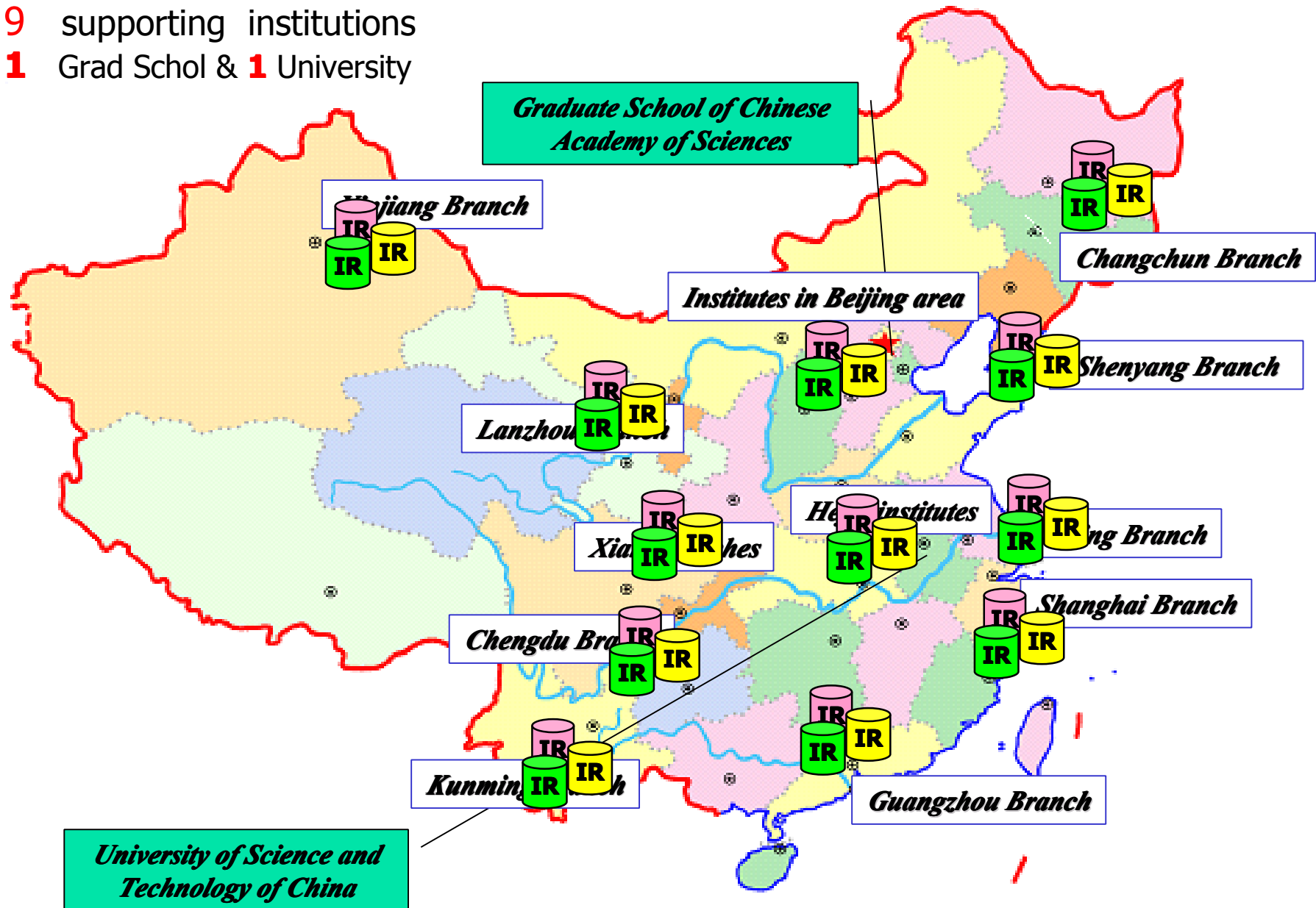
- **Scientists of CAS (Chinese Academy of Sciences)**
 - **Play a very important role in Chinese science research**
 - **Publish many high standard papers abroad**
 - **But they could not always be disclosed and accessed by Chinese scientists**
- **How to make full use of these high standard academic resources is a big problem for scientists, policy makers and librarians in CAS.**

1. Background and Some Considerations

- **In 2007, the CAS Institutional Repository Grid was brought forth**
 - **Each institute establishes its own local repositories as a node of the Grid**
 - **NSL constructs a centralized metadata repository, which could harvest the metadata of academic resources stored in many distributed institutional repositories**
 - **NSL also keeps an integrated search interface for the resources and provides other enriched services**

Future Institutional Repository Landscape CAS

- **12** branches
- **97** institutes
- **9** supporting institutions
- **1** Grad Schol & **1** University



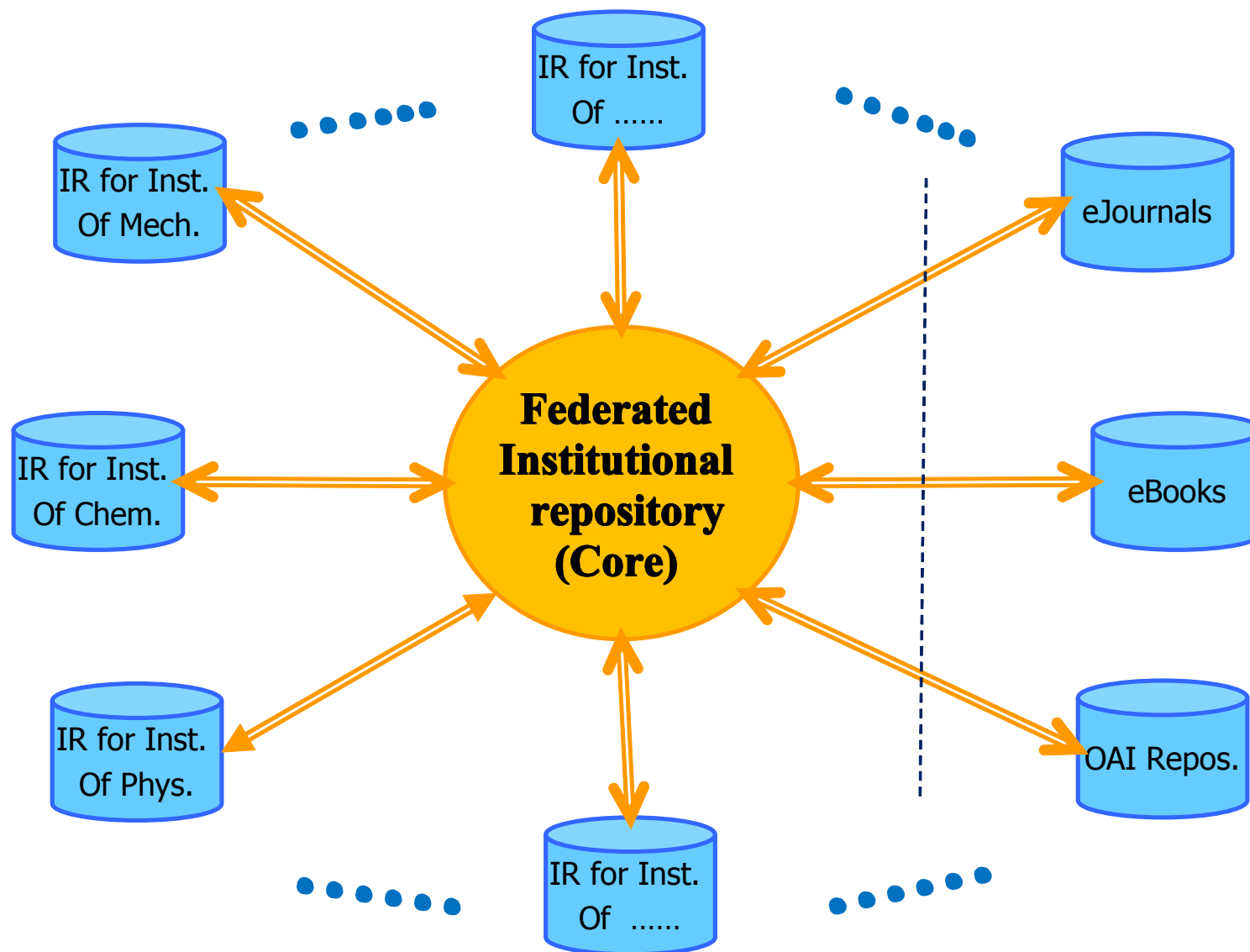
1. Background and Some Considerations cont.

- **CAS IR Grid, which targets at**
 - **providing the organizational and technological framework for a CAS-wide IR infrastructure for research**
 - **helping every institutes set up their own IRs to increase the visibility of, and preserve their research outputs**
 - **implementing a harvester-based cross-repository search and browse service to enhance exposure of the CAS's research output as a whole**
 - **making the GRID ready for becoming building blocks for national or international wide repository infrastructure**

Outline

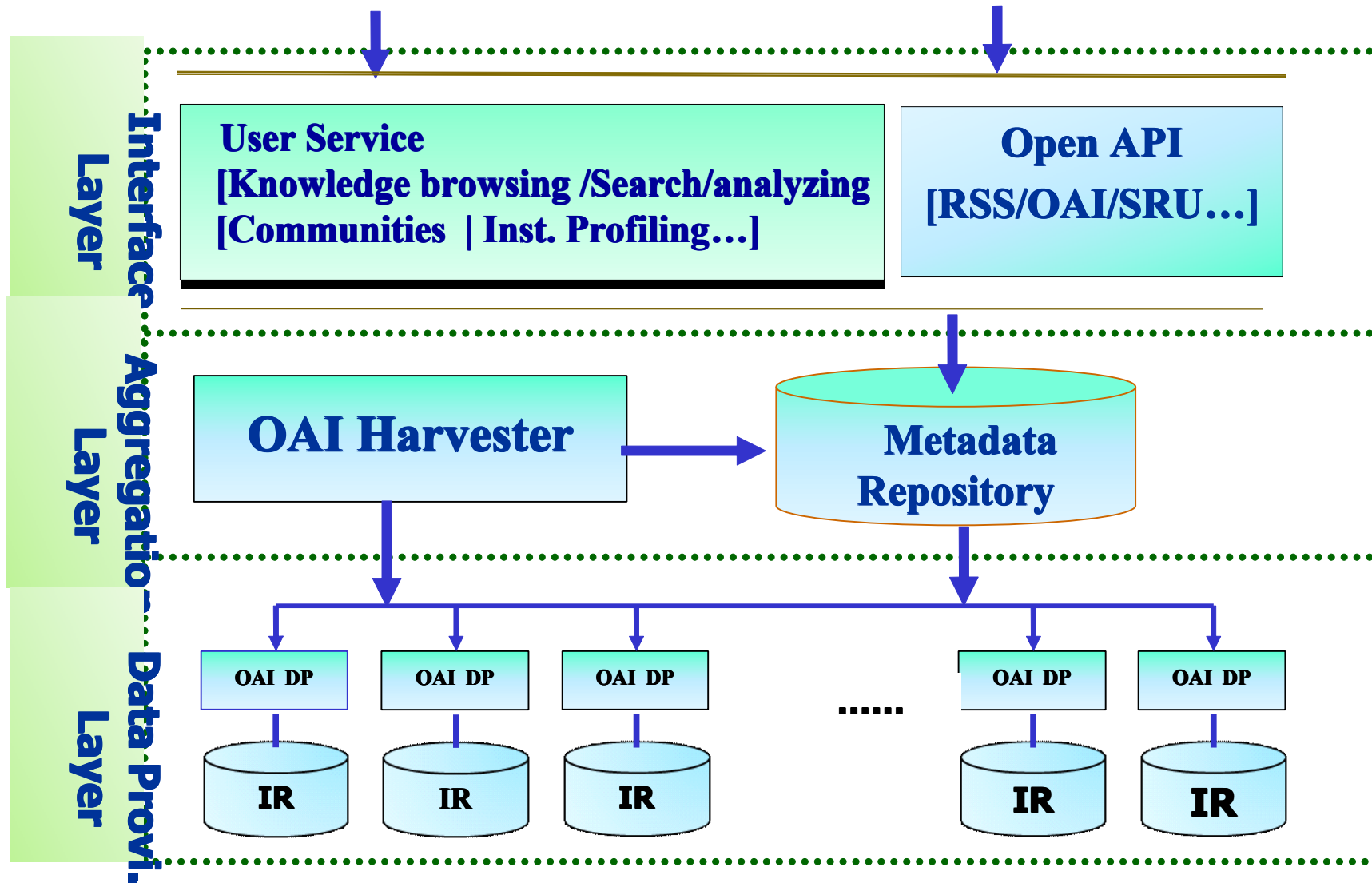
- **Background and Some Considerations**
- **System Architecture of CAS IR Grid**
- **Two-phased Pushing-up strategies**
- **Lessons and Experiences from Implementation**

2. System Architecture of CAS IR Grid



"Hub and Spoke" Model

2. System Architecture of CAS IR Grid



Illustrated Functional Structure

2. System Architecture of CAS IR Grid

- **In institute level, we focus functionalities of**
 - **depositing and publishing all sorts of knowledge outputs (Journal and conference papers, ETDs, documentary materials, research data, and etc.)**
 - **knowledge organization and sharing**
 - **long term preservation**
 - **knowledge asset auditing**
 - **knowledge relationship analysis and etc.**

2. System Architecture of CAS IR Grid

- **Whilst in the Academy level, we concern**
 - **knowledge discovery and scheduling**
 - **organization of Communities**
 - **cross-disciplinary or cross boundary content integration**
 - **institutional profiling**
 - **knowledge capacity analysis and etc.**

2. System Architecture of CAS IR Grid

- **Also, Federated Institutional Repository should be embedded into the services of Library, and be a component of research infrastructure of CAS**
- **We have two steps to go to build our IR Grid**
 - **Step One: Building Institutional Repositories in Institutes**
 - **Step Two: Building the Federated Repository for the whole Academy**

Outline

- **Background and Some Considerations**
- **System Architecture of CAS IR Grid**
- **Two-phased Pushing-up strategies**
- **Lessons and Experiences from Implementation**

3. Phase I: Building Institutional Repositories in Institutes

- **Four tasks we have done**
 - **Select the institute and set up examples for other institutes to follow**
 - **Develop IR based on DSpace**
 - **Integrate with the existing system (such as ARP)**
 - **Help the institute plan and implement IR**

3.1 Select the institute and set up examples for other institutes to follow

- **Best practice is very important for implementation of IR in institutes**
- **We are very careful in choosing the institute for setup a example**

3.1 Select the institute and set up examples for other institutes to follow

- **We laid out the principles for choosing institutes for set up examples**
 - **Urgent needs of IR from institute**
 - **Support from the head of institute**
 - **Clear understanding of IR from the library of the institute**
 - **Necessary input to the IR implementation**
 - **Experienced in implementation of information systems**

3.1 Select the institute and set up examples for other institutes to follow

- **After evaluation, Institute of Mechanics, CAS become our choice**
- **In addition to the evaluation criteria we laid out, the institute have some other features we prefer:**
 - **The head of library also is one of the head of institute**
 - **The library of the institute is one of the most active libraries in CAS**

3.2 Develop IR based on DSpace

- **To reduce the complexity and TOC of the system, we choose to use DSpace as fundamental institutional repository system**
- **To make DSpace adapt to the Chinese information environment, we have localized the DSpace and built a Chinese institutional repository system based on it**

3.2 Develop IR based on DSpace

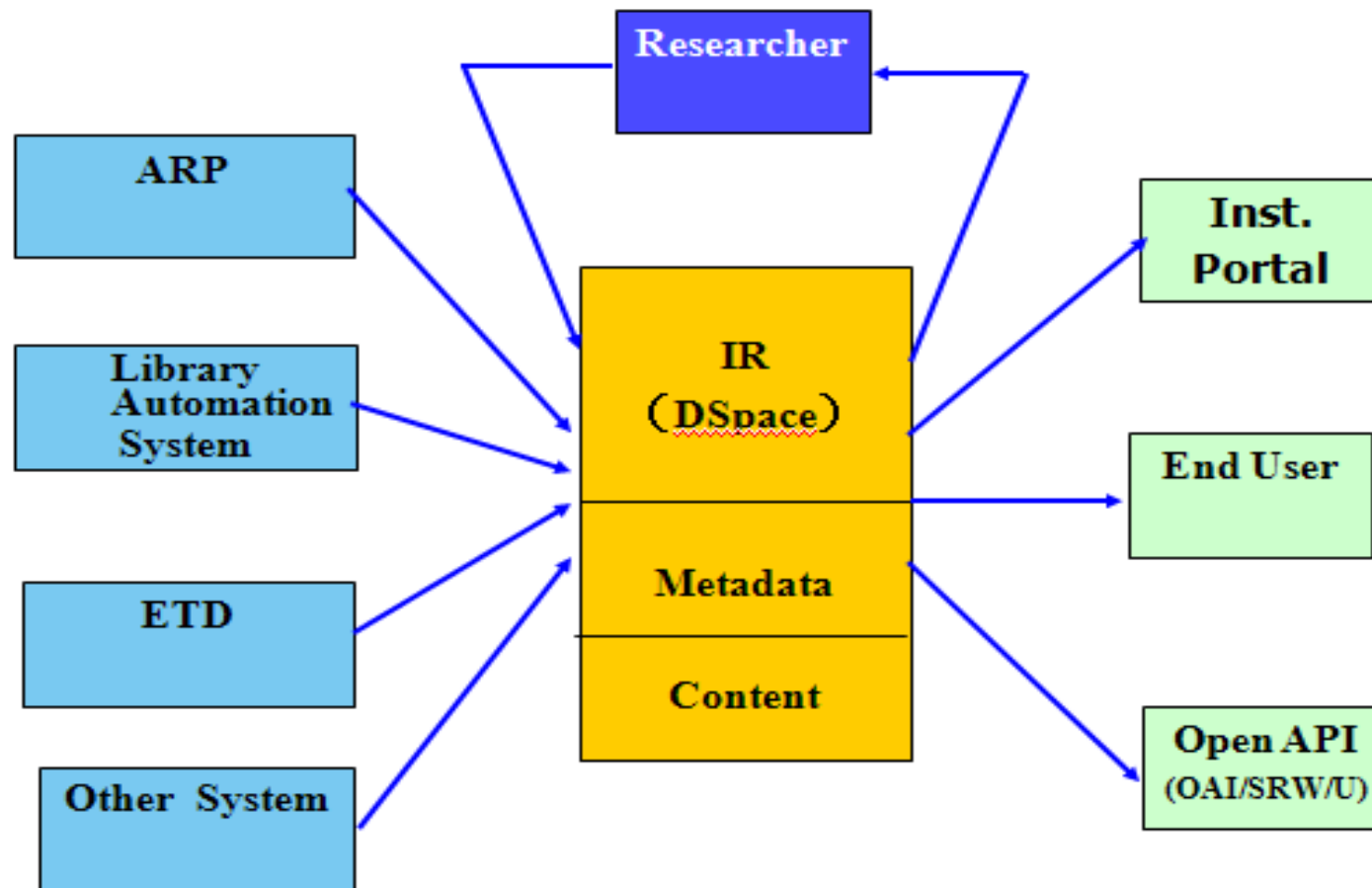
- **What we have done to localize the DSpace**
 - **Make DSpace support Chinese**
 - **Extend the metadata schema of DSpace**
 - **Improve the workflow and interface of DSpace**
 - **Enhance the import/export functions of DSpace**
 - **Enhance access control of DSpace**
 - **Optimized OAI-PMH data provider, especially for exposing customized Metadata record**
 - **Support research statistics**
 - **Provide open interface for DSpace, for example SRU (Search and Retrieval from URL)**

3.3 Integrated with the existing system

- **An important problem is how to integrate the existing system**
- **For example, since the institutes of CAS have implemented ARP (Academic Resources Plan) system which stored metadata of research outputs of the institutes. It is quite important for the institutes to use those existing metadata in the ARP systems to avoid duplicate input work**

3.3 Integrated with the existing system

- In fact, lots of existing systems to be integrated



3.3 Integrated with the existing system

- **What we have done**
 - **Standard metadata and data package Format (for import and export in bulk)**
 - **Open search interface via SRU**
 - **Try to make good use of OpenURL functions provided by DSpace**

3.4 Help the institute plan and implement IR

- **In order to make Institute of Mechanics, CAS set up the IR and make IR function normally, we try to**
 - **Help the institute lay out a strategies**
 - **Help the institute organize the deployment team**
 - **Provide Training program for**
 - **the staff of library**
 - **the researchers of the institute**
 - **the students of the institute**

3.4 Help the institute plan and implement IR

- **We have also provided reference materials on implementation of IR:**
 - **Guidelines of planning and implementation of IR in CAS**
 - **The manual of IR in CAS**
 - **The system management guide of IR in CAS**
 - **Best practices of implementation of IR in CAS**
 - **.....**

3.4 Help the institute plan and implement IR

■ IR sited in institutes of Mechanics, CAS

The screenshot shows a Microsoft Internet Explorer browser window displaying the website <http://dspace.imech.ac.cn/>. The website header features the logo of the Institute of Mechanics, Chinese Academy of Sciences (IMECH), and the title "中国科学院力学所机构知识库" (IMECH-KNOW). The main content area is titled "NEWS" and features a news article:

美国国家科学基金会K. P. Chong教授来访
7月12日, 美国国家科学基金会力学与材料工程部 (Mechanics and Materials, Directorate for Engineering, National Science Foundation) 主任K. P. Chong教授访问我所, 在非线性力学国家重点实验室 (以下简称LNM) 副主任戴兰宏、宋凡的陪同下参观了LNM实验室。
参观结束后, 郑哲敏院士、何国威主任、魏悦广研究员等与Chong教授座谈。Chong教授介绍了美国国家基金会在材料与力学方面的资助情况, 双方就该领域当前研究发展态势交换看法, 并希望今后推进相关合作。

The website also includes a search bar, a navigation menu on the left, and a "最新提交" (Latest Submissions) section on the right with a list of research papers. The browser's address bar shows the URL <http://dspace.imech.ac.cn/>.

3.4 Help the institute plan and implement IR

Two test IR sites for NSL and OPT institute

The image displays two screenshots of institutional repositories (IR) for the Chinese Academy of Sciences (CAS).

The top screenshot shows the "中国科学院国家科学图书馆机构知识库" (Knowledge Repository of National Science Library, CAS). It features a search bar, filters for "作品访问排行" (Work Access Ranking) and "统计概要" (Summary Statistics), and a list of works. A red box highlights the search and filter area.

The bottom screenshot shows the "中国科学院西安光学精密机械研究所机构知识库" (Knowledge Repository of Xi'an Institute of Optics and Precision Mechanics, CAS). It features a search bar, a login/register section, and a list of "最新提交" (Recent Additions) and "内容类型" (Types of Content). A red box highlights the entire page content.

The "Recent Additions" list includes:

- 干涉成像光谱技术研究
- 俘精酸酐光致变色特性及应用研究
- 飞秒激光在透明介质中的三维光存储研究
- 飞秒激光与透明介质相互作用的非线性及应用
- 飞秒激光诱导叶绿素、类胡萝卜素分子在光系统II中能量传递研究
- 飞秒激光引发PSII色素分子间的传能动力学
- 二极管侧面泵浦倍频固体激光技术研究
- 多通道光学成像系统数据处理技术的研究
- 短波长与固体浸没透镜高密度光存储技术研究
- 第三代像增强器研究

The "Types of Content" section includes:

- 期刊论文
- 会议论文
- 学位论文
- 专著
- 研究报告
- 演示报告
- 多媒体资料
- 其他

The "Communities & Collections" section includes:

- 研究社群

The right sidebar of the bottom screenshot lists various services and links:

- 下载排行
- 无可显示数据..
- 相关链接
- 跨库检索
- Cross Search System
- 跨界集成检索
- Crossdomain Search System
- 联合目录
- Union Catalogue
- ScienceChina
- 中国科学文献服务系统
- 学位论文检索
- E-Thesis&Dissertation
- 原文传递
- Document Delivery
- 问图书馆员
- Ask Librarian
- 科技新闻聚合服务
- S&T News Syndication

Other 4~6 institutes are queued to be launched their IR program this year

4. Phase II: Building the Federated Repository for the whole Academy

- **Three tasks need to be done**
 - **Deploy the IRs and Spread the IRs to institutes across the country**
 - **Develop OAI metadata harvester system to harvest metadata from IRs in institutes to NSL center repository**
 - **Develop value-added and augmented services**

4.1 Deploy IRs and Spread the IRs to institutes across the country

- **Provide training program for**
 - **Librarian of the institutes**
 - **Management officer of the institutes**
- **Provide detailed materials for implementing IR**
 - **Manual, guide, best practices**
- **Provide strong technologies support for installation and deployment of IR**
 - **Install on spot**
 - **Remote install**
 - **Remote help system**

4.2 Develop OAI metadata harvester system

- **Now we consider the FIR are to be based on DSpace as well**
- **OCLC open source software OAI Harvester 2.0 are selected to be extended and integrated into DSpace via a way of plugin**
- **DRIVER's open-sourced D-Net toolset are also considered whether it will be a better option for us, though**

4.3 Develop augmented & value-added service

- **Extend browsing and search system to provide content in a variety of logically organized dimensions**
- **Develop knowledge capacity and capability analysis functionalities**
- **Support OAI and SRW/U interface that it can be easily integrated with other systems**
-

Outline

- **Background and Some Considerations**
- **System Architecture of CAS IR Grid**
- **Two-phased Pushing-up strategies**
- **Lessons and Experiences from Implementation**

5 Lessons and Experiences from Implementation

- **The requirement of institutes is a factor more important than technologies**
- **Integrated with the existing system and other libraries services is very important**
- **Planning ahead will lead to successful deployment**
- **IR must be an open system provide necessary access interfaces**

Thanks!

zzm@lzb.ac.cn