CSpace – A More Practical and Customizable Repository Platform Serving Local Needs

Zhongming Zhu, Wangqiang Zhang, Wei Liu, Xiaona Yao, Linong Lu
{zhuzm, zhangwq, liuw, yaoxn, lunl}@llas.ac.cn

Lanzhou Branch of National Science Library, CAS
Scientific Information Center for Resources and Environment, CAS
2013.7.08-12, PEI, Canada
CSpace - A Quick Glance

- IR platform used in CAS* IR Grid
- Allowing for easy customization deployment
- Extensively extended and upgraded based on DSpace 1.4.2 since 2008
- Developed & maintained by Lanzhou Branch of National Science Library, CAS
- Open sourced officially since 2012

*CAS -- Chinese Academy of Sciences
CAS IR Grid

• An overarching repository infrastructure throughout CAS
  – Preservation & dissemination of CAS research
  – Knowledge capacity building mechanism for the institutes across CAS
  – Fostering a culture of OA in CAS and China.

• The IR for CAS as one single research organization
  – Launched in 2008

http://www.irgrid.ac.cn
CAS IR Grid
NOW!

• Total IRs: **103**, across **25 cities**
• Total Records: **430,000**
• Recs. with FT: **78%**
• **29,452,000 views**
• **4,778,000 dlds**
Simple Principles of CSpace

● Be relevant
  – User oriented
  – Service centered

● Continual enrichment of the range of value-added services
  – Respond proactively to the needs and RQMTS

● Continual imprv't of customizable capability
  – Adapt flexibly to new, different, and/or changing requirements and environments
Extended Functionalities and Services Portfolio
Collection Building Services

- Optimized self-submission
  - Two-step quick submission workflow
  - Document-type aware description form
    - display minimal required description fields
    - Folded/unfolded optional fields
    - Automatic duplication check
    - Integration of SHERPA/RoMEO query
    - Fine-grained rights control options
  - Automatic BKGD processing following submission
    - subject headings assignments based on OpenKOS
    - Pdf conversion(non-pdf docs) for online browsing
  - Remote submission via SWORD
Description ----> Confirmation

Choose templates

Description:

Choose files

Required field

recom'd field

Folded fields

RoMEO query

Access rights assignment

Confirmation:
Collection Building Services -2

- Easily used web-based bulk import
  - Content type based modifiable/self-defined XML/EXCEL templates for ingesting various types of content.
  - Predefined templates for importing formatted data from Endnote, SCI, CNKI, and other sources.
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>论文名称</td>
<td>论文作者</td>
<td>作者部门</td>
<td>作者项目组</td>
<td>通讯作者</td>
<td>内容类型</td>
<td>学科主题</td>
<td>关键词</td>
<td>分类号</td>
<td>别名</td>
</tr>
<tr>
<td>2</td>
<td>IR手册</td>
<td>manual of IR</td>
<td>zm,lul</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>IR手册</td>
<td>manual of IR</td>
<td>zm,lul</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>IR手册</td>
<td>manual of IR</td>
<td>zm,lul</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

批量导入

| 研究部门 | 中科院图书馆
| 内容类型 | 手稿 |
| 导入文件类型 | CSV, Excel |
| 文件路径 | 无 |
Collection Building Services -3

- Automated ingestion and integration
  - Pre-existing ETD database
  - ARP (Academia Resource Planning) systems, i.e. research management systems, deployed in every institute of CAS
  - Web of Science record (via web service interface)
  - BMC publications of CAS (via SWORD)
  - OAI-Harvesting (with extension of harvesting content objects), if applicable
Ingest metadata from Web of Science

<table>
<thead>
<tr>
<th>Web Service版本</th>
<th>Lite</th>
<th>Premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>认证方式：</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IP地址</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>账户</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>用户名：</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

检索条件：
- 单位名称：
- 研究部门：
- 作者：
- 发表年度：

采集控制：
- 目标数据库：
  - SCI
  - SSCI
  - AHCI
  - CPCI-S

导入控制：
- 是否查重：
  - 是
  - 否

Search by combination of:
- Institution name
- Department name
- Author
- Pub year

Choose source databases
Dissemination and Rights Management

- Multi-level & fine-grained access control
  - Item embargo: 3-6-12 months, or any time span
  - Access level: metadata, full content
  - Access scope: public, institute, community
  - Access types: online browse, watermarking, download
  - Assignment of content types related distribution policies
  - IP based full content access control

- Malicious download monitoring and blocking

- Complaints management
Author Identification & Authorship Claim

• Alias control
  – Unique author identifier
  – Names variants

• Authorship claim
  – Match/email possible authors
  – Confirmation of authorship and authorship order

- Establishing defining relationships BTW authors and articles
- Forming a reliable base for clustering related articles by authors
Multi-faceted content use and reuse

- Faceted browse/search
- Online pdf viewing
- Auto-suggest/-completion
- KOS/DDC based clustering
- Integrated connecting services
  - Recommendation
  - Recommended citation
  - Forwarding search
  - Social bookmarking
  - Export (EndNote/Word/CSV)
  - SCI citation counts
  - Usage statistics
  - Rights policies
  - comment, complaint
Researcher knowledge profile

- CV mgmt & export
  - Work/edu background
  - Research interests
  - Projects
  - .......
- Research inventories
  - Categories
  - SCI/CSCD citation counts
- Personal web site
  - Chinese version
  - English version
Usage statistics

- Combination of:
  - different content object levels
    - site, community, collection, item
  - different time interval levels
    - year, month, day, custom time period
  - different access styles
    - robot access, intranet access, repeated clicks
  - different countries or regions, etc.

- Display results in a variety of forms
  - histograms, ranking lists, Excel spreadsheets, etc.
Knowledge asset audit

- Reviewing and reporting knowledge asset status in various levels and dimensions:
  - institute, community, individual
  - asset types
  - time spans
  - Subject......

- Presenting results in forms:
  - lists, histograms, line graphs, pie charts...

- Flexible personalization
Knowledge mapping

- Co-authorship network
  - Communities
  - research output types
  - year
Knowledge Mapping

- Co-subject/SKOS categories
Open Interfaces and interoperability

- OAI-PMH DP/SP
- SRU
- SWORD
- OpenSearch
- RSS
- XML Sitemaps for SEO
Customization Capabilities
Extensible Metadata Framework

- Extend metadata schema on demand via Web UI
  - Modifying/Reusing existing elements
  - Introducing new elements if needed
    - element
    - qualifier
    - label_zh
    - Lable_en
    - scope_note
    - display_on_submission
    - display_on_browse
    - display_on_stat
    - edit_allowed
    - ......

- Key to support new content types
Content Type Templates

- Create/customize content type templates via Web UI
  - Specify a list of allowable metadata terms (fields)
  - Determine fields order in submission/browse forms
  - Tailor fields behavior: display name, input style (e.g. textbox, dropdown list), default value, requiredness, repeatability, hidden....
  - Define citation format
  - create or assign distribution and rights policies

- Support content type aware submission/display
<table>
<thead>
<tr>
<th>内容类型</th>
<th>英文名称</th>
<th>引用格式</th>
<th>操作</th>
</tr>
</thead>
<tbody>
<tr>
<td>期刊论文</td>
<td>Journal</td>
<td>期刊论文</td>
<td>删除</td>
</tr>
<tr>
<td>学位论文</td>
<td>Dissertation</td>
<td>学位论文</td>
<td>删除</td>
</tr>
<tr>
<td>会议论文</td>
<td>Meeting Paper</td>
<td>会议论文</td>
<td>删除</td>
</tr>
</tbody>
</table>

**模板定制**

内容类型模版： 期刊论文  
| 显示/修改字段信息 | 添加/删除/排序字段 | 保存修改 |

**备选字段**

<table>
<thead>
<tr>
<th>题名</th>
<th>其他题名</th>
<th>作者</th>
<th>隶属作者</th>
<th>导师</th>
<th>专利发明</th>
<th>专利代理</th>
<th>编者</th>
<th>其他责任者</th>
<th>责任者</th>
<th>学科 \ 领域</th>
<th>关键词</th>
<th>专业方向</th>
<th>分类号</th>
<th>英文标题</th>
<th>摘要</th>
</tr>
</thead>
</table>

**已选字段**

<table>
<thead>
<tr>
<th>原本</th>
<th>学科</th>
<th>题名</th>
<th>其他题名</th>
<th>作者</th>
<th>隶属作者</th>
<th>导师</th>
<th>专利发明</th>
<th>专利代理</th>
<th>编者</th>
<th>其他责任者</th>
<th>责任者</th>
<th>学科 \ 领域</th>
<th>关键词</th>
<th>专业方向</th>
<th>分类号</th>
<th>英文标题</th>
<th>摘要</th>
</tr>
</thead>
</table>

**修改字段：题名**

<table>
<thead>
<tr>
<th>字段名称</th>
<th>题名</th>
<th>作者</th>
<th>隶属作者</th>
<th>导师</th>
<th>专利发明</th>
<th>专利代理</th>
<th>编者</th>
<th>其他责任者</th>
<th>责任者</th>
<th>学科 \ 领域</th>
<th>关键词</th>
<th>专业方向</th>
<th>分类号</th>
<th>英文标题</th>
<th>摘要</th>
</tr>
</thead>
</table>

**默认值**

<table>
<thead>
<tr>
<th>增加字段</th>
<th>是否必填</th>
<th>是否多值</th>
<th>是否显示</th>
</tr>
</thead>
</table>

**中文备注**

<table>
<thead>
<tr>
<th>适用范围</th>
<th>适用于统计显示</th>
<th>适用于统计目录显示</th>
<th>统计输入形式</th>
</tr>
</thead>
</table>

**英文备注**

<table>
<thead>
<tr>
<th>适用范围</th>
<th>适用于统计显示</th>
<th>适用于统计目录显示</th>
<th>统计输入形式</th>
</tr>
</thead>
</table>

**确定** 取消
Content Type Based Import/Export

- Create configurable XML import templates to import text data in any format
- Create configurable content type based EXCEL import templates to import data in EXCEL format
- Similarly, data in repository can be exported via creating content type based XML/EXCEL templates
- Import & export operations are managed via Web UI
Flexible Customization of Asset Auditing

● Overall set of audit conditions can be dynamically defined and configured
● Each time of audit process can be customized, based on prescribed audit conditions set
● Audit results manifestations can be customized to display as inventory lists, histograms, line graphs, pie charts.
● Columns of items appeared in an inventory list also can be adjusted as desired.
● Of course, all above customizations are Web-based
Web-based System Configuration

- Most of parameters or options are collected together to be adjusted or customized via Web UI
- Support simple skin change
- One-key installation package in Windows plat.
  - All neccessary resources in one executable pkg
  - Installation location can be customized
- Actually, automatic update mechnism is now in place
Future Development

- Continual enrichment /improvement of the range of, and customization capabilities of, value-added services
  - Non-textual content management
  - Automatic metadata extraction and text mining
  - Micro-services based repository infrastructure
  - Semantic enhancement services
- Contribute more and better to repo community
  - esp. DSpace community
Get CSpace

- CSpace Github Repository
  - https://github.com/cspace
  - http://sourceforge.net/projects/cspace-ir/

(Note: the latest version has not yet been uploaded)
Thank you for Listening & any questions?