

中国科学院机构知识库体系建设系列文档

## CAS-IR 推介简明指南

中国科学院国家科学图书馆兰州分馆

信息技术部

2008-12

## 目 录

1. 什么是机构知识库.....	3
2. 研究所建设 IR 的好处.....	3
3. 什么是 CAS-IR.....	4
4. CAS-IR 功能简介.....	4
5. CAS-IR 的运行环境和条件.....	5
6. CAS-IR 的技术支持与服务.....	6
7. CAS-IR 应用示例.....	6

## 1. 什么是机构知识库

机构知识库 (Institutional Repository, 简称 IR), 也称为机构仓储或机构知识仓储, 是科研教育机构对本单位员工所创造的各种类型有价值的知识产出进行统一收集、集中管理、长期保存和提供检索利用等增值服务的知识资产管理系统。

开放获取 (open access)、自存档 (Self-Archiving)、数字资产管理 (Digital Asset Management) 等概念和实践活动的兴起, 是 IR 发展的重要因素和推动力量。

**开放获取 (Open Access)** 是一种通过互联网让科学研究成果能够自由传播的主张和运动。它可以让用户把经过同行评议的学术论文放到互联网上, 以利于学术成果的快速传播。开放获取较少考虑版权和授权限制, 一般有两种实现模式: 开放获取期刊 (OA Journals) 和作者自存档 (Author self archiving)。

**自存档 (Self-archiving)** 是一种由作者自己将其科学研究成果提交和发布到支持开放获取的知识仓储系统 (通常为所在机构的 IR) 中供同行和公众自由下载和使用的活动, 从而有助于实现研究成果学术价值的最大化, 并有益于研究者个人和所在机构的学术声誉和影响的扩大。

国际上, IR 的发展在近年来保持稳步增长的态势, 欧美众多大学和科研机构都已逐步建立起了自己的 IR 系统, 如麻省理工学院的 DSpace@MIT、加利福尼亚大学的 eScholarship、剑桥大学的 DSpace@Cambridge、美国洛斯阿拉莫斯国家实验室的 LANL、德国马普学会的 eDocServer 等。根据 OpenDOAR 的最新统计数据, 目前全球已有近 1300 个 IR。

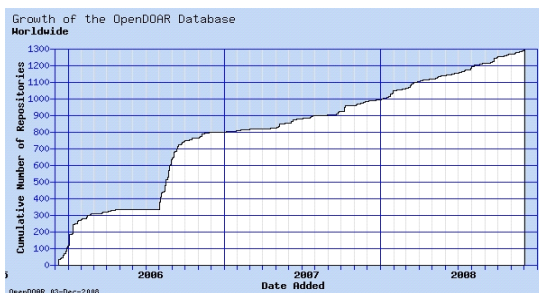


图 1 IR 增长趋势

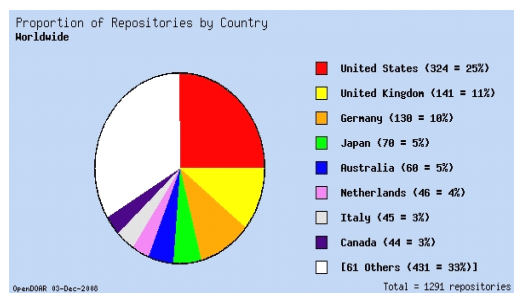


图 2 IR 国别分布图

## 2. 研究所建设 IR 的好处

- (1) 实现研究所知识产出的系统积累、长期保存和统一管理。研究所科研人员的知识产出是研究所的重要学术和知识资产, 利用 IR 可以对科研人员的知识产出和研究成果进行系统收集和长期保存, 建立集中和统一的数字资产管理系统, 从而克服各部门以及个人分散保存信息的弊端和风险。通过长期的积累, IR 将成为研究所的知识资产宝库。

- (2) **集中展现和反映研究所的整体研究实力和水平。** IR 通过对研究所分散的知识产出进行收集和整理，提供统一的展示和利用窗口，有助于研究所完整的学术形象和“品牌”的形成。
- (3) **拓宽学术成果的发布和交流渠道，增强可见性。** 研究论文以预印本方式首先在 IR 中发布，并不排斥其在期刊等传统学术媒介上的正式发布或出版，相反还能支持作者更快、更早地确定研究成果的首创地位。同时，将极大地拓展研究成果的传播范围和速度，有助于是科研人员更容易地发现同事或同行们的最新研究动向和成果。
- (4) **提高学术成果被发现和引用的几率，扩大科研人员及研究所的学术影响和声望。** 已经有研究表明，存放在 IR 中的论文，平均被引用次数是未存放在 IR 中的论文的两倍。
- (5) **支持和实现数字出版。** IR 除了可以支持对正式的研究产出的管理外，还可以将原先无法进入正式出版途径的大量知识内容（如研究报告、演示报告、技术文档、文件、数据等）以统一格式进行发布和交流。
- (6) **提供学术统计、研究评价的辅助工具。** 随着 IR 中知识内容的不断积累，可以作为对研究所知识产出和知识资产状况及其应用进行统计或审计，如按照时间、研究/试验室、课题和个人等进行统计；也可以对所保存的内容的访问利用情况进行统计和分析，支持研究评价的需要。

### 3. 什么是 CAS-IR

CAS-IR 是由中国科学院国家科学图书馆组织开发的 IR 软件系统，免费向全院各研究所推广使用，支持研究所依托该平台快速构建和部署 IR 应用，促进研究所知识资产管理活动的开展。

### 4. CAS-IR 功能简介

CAS-IR 具备常见 IR 系统的所有功能，包括知识内容的存缴、组织、传播利用、保存、统计和互操作等多种功能和服务。

(1) **存缴方面**，CAS-IR 提供了快速简便的存缴接口，提供了针对主要知识内容类型的元数据描述模板，支持提交过程中，作者可以根据自己的意愿，选择简单的或相对复杂的元数据描述，完成提交和文件内容的上传。同时，CAS-IR 也提供了以批量方式装载外部数据（如来自 ARP 系统的产出物数据）的方法和接口。

CAS-IR 目前支持的主要知识内容类型包括：期刊论文、会议论文、专著、学位论文、研究报告、演示报告、多媒体资料等，对科学数据的支持还不成熟。

(2) **组织方面**，提供了按照研究/试验室、研究方向或课题组、知识内容类型、学科主题分类、时间、题名、作者等多种组织方式。

(3) **传播利用方面**，支持按照以上组织方式的浏览和导航，支持从题名、作者、时间、关键词等多种途径和条件的简单检索或组合检索，并支持全字段及全文档的全文检索。支持基于 OAI 的元数据开放收割接口和基于 SRU 的开发检索接口，支持其他系统开放地检索和方便与相关系统的开放集成和互操作应用。在知识内容的传播保护方面，提供了基于 IP 地址、特殊用户组、作品发布时限等多种方式对全文性内容实施限制存取。

(4) **长期保存方面**，支持对存缴对象的技术信息、权属信息、源流信息等与长期保存有关的信息的记录，对已存缴作品进行撤出管理，定期备份保存内容，以及对存缴对象以 METS 格式进行封装和输出迁移等。

(5) **统计方面**，CAS-IR 提供了基于知识库内容的统计和访问利用统计。内容统计，支持按照研究所、研究/实验室、职称、身份、作者或用户、时间段等多条件组合统计，统计结果可以 EXCEL 格式导出利用；访问统计，提供 IR 整体或单篇作品在预定义时间段内的浏览、下载利用情况。

(6) **其他功能和服务**。除了上述主要功能和服务，CAS-IR 还提供基于电子邮件的最新内容订阅、RSS 订阅、浏览和检索结果的保存及输出、关联检索、推荐等。需要说明的是，基于电子邮件和 RSS 方式内容订阅，默认情况下为关闭状态，需根据情况决定是否打开。

## 5. CAS-IR 的运行环境和条件

### (1) 硬件平台

CAS-IR 对硬件平台无特殊要求，任何支持 Unix/Linux、Windows XP/2000 系列操作系统运行的硬件环境，如各种 Unix/Linux 服务器、以及大部分主流品牌的 PC 服务器都可以作为安装和运行 CAS-IR 的基本硬件平台。硬件平台的具体配置要求，视拟建设 IR 系统在数据规模、保存、备份、性能等方面的不同要求而有所不同。通常情况下，推荐按照如下条件配置运行 IR 的硬件平台：

**服务器：**主流双 Intel Xeon 处理器，2~4GB 内存，500G 以上外存。若需要做数据的离线备份和保存，可以考虑配置内置或外置磁带机（库）。

投入预计：3 万元左右。

### (2) 软件环境

#### ◆ 操作系统：

- **首选：**Linux/Unix 服务器操作系统；
- **推荐：**Ubuntu 8 服务器版（免费），或 Red Hat Enterprise Linux（商业）。
- **可选：**Windows XP/2000 系列操作系统（商业）。需要说明的是，目前 CAS-IR 在 Windows 平台运行时部分功能（如访问统计、定时备份）等还不稳定。

#### ◆ 数据库管理系统：Postgre SQL 8.1（免费）

#### ◆ Web 应用服务器：Tomcat 5.5（免费）

◆ Java 平台: J2SE 1.5 (免费)

## 6. CAS-IR 的技术支持与服务

国科图将在学科馆员团队和 CAS-IR 项目组为核心的基础上, 形成稳定支持和保障研究所开展 IR 应用的工作机制, 以多种方式为研究所提供必要的实施咨询、安装和部署支持(现场或远程)、应用培训、系统升级等技术支持与服务。

## 7. CAS-IR 应用示例

The screenshot displays the homepage of the Knowledge Repository of the National Science Library, CAS. The interface is in Chinese and features a blue header with the library's name and logo. On the left, there is a navigation menu with options like '登录/注册', '研究社群', and '用户服务'. The main content area is divided into three sections: '最新提交' (Recent Additions), '内容类型' (Types of Content), and '研究社群' (Communities & Collections). The '最新提交' section lists various documents and includes a text box with a mission statement about developing institutional knowledge management capabilities. The '内容类型' section provides a grid of content categories such as '期刊论文', '会议论文', and '学位论文'. The '研究社群' section lists various research groups and their respective counts. On the right side, there are sections for '下载排行' (Download Ranking) and '相关链接' (Related Links), listing popular documents and external search systems.