

## 中国科学院国家科学图书馆组织建立 全院机构知识库网格服务系统

机构知识库 (Institutional Repository, 简称 IR) 是科研机构实现知识资产管理和机构知识成果传播的平台, 是实现开放获取的主要手段之一。研究机构通过建设 IR 可以: (1) 实现对机构知识成果的系统收集和管理, 籍此随时了解和掌握机构自身的知识资产“家底”; (2) 支持机构对其知识资产进行长期保存, 避免各部门和个人分散保存及技术变迁等造成的知识资产流失风险; (3) 促进机构知识成果的广泛交流共享, 扩大机构及其科研人员学术影响力和声望; (4) 为知识产出和知识能力分析评估提供基础等。因此, IR 的研究和建设得到了国际上众多大学和科研机构的重视, 全球已有 1440 多家科研机构建立了 IR, 并逐步发展成为所属机构普遍性的信息基础设施。在此背景下, 中国科学院从 2007 年即开始筹划在中国科学院范围内开展研究所机构知识库的建设, 并提出了构建中国科学院机构知识库网格 (Chinese Academy of Sciences Institutional Repositories Grid, 简称 CAS IR Grid) 的建设框架: 在全院各研究所开展 IR 的推广建设, 形成 IR Grid 的节点系统; 在研究所 IR 建设基础上, 通过元数据自动采集技术, 建立全院集成的机构知识成果检索服务, 集中揭示和传播全院知识产出的集成服务平台。

CAS IR Grid 的建设采取了试点示范、规模推广、逐步完善的策略。在 2008 年少数研究机构试点的基础上, 2009 年开始, CAS IR Grid 的建设进入规模推广阶段。截至目前, 全院已有分布于全国 22 个城市的 63 家研究所启动了 IR 建设。其中, 有 40 家支持公开访问, 23 家暂时处于建设启动或所内访问状态。部分研究所已经建立了与年度考核和学术奖励关联的机构知识库内容存缴管理政策, 为 IR 的内容采集和传播形成了有利的制度和政策保障。在数据规模方面, 全院研究所 IR 累计可公开访问的期刊论文、会议论文等已达 85000 条, 其中 67% 带有全文内容。

随着研究所 IR 建设进入规模推进阶段和有了一定规模的数据积累, CAS IR Grid 集成服务门户开始同步进行平台的开发和数据的采集。从 2010 年下半年开始, 集成服务门户系统正式开始对支持公开访问的研究所 IR 进行数据的自动采集和集成服务, 支持从研究所、作者、内容类型、时间、关键词等多种途径对全院知识成果进行浏览、检索或统计利用, 支持与国际上著名知识库或知识库网络的数据交换和共享集成。目前, 集成服务门户已实现对 34 家有一定可公开访问数据规模的研究所 IR 进行数据的自动采集, 累计完成 85840 条数据的采集和发布。

CAS IR Grid 未来发展的重点包括四个方面: 一是推动研究所建立完善的 IR 建设制度和政策框架, 建立公共资金项目所发表的研究论文的开放存缴制度, 推动存放内容从论文、专利等逐步扩展到科学数据、教学课件、科技报告、会议演示文档、管理文件以及其它开源信息集等, 实现对研究所创造的有重要保存和传播共享价值的知识内容的全谱段可靠捕获、广泛宣传 and 长期保存。二是继续扩展研究所 IR 建设的规模, 逐步使全院绝大部分乃至全部研究所都加入到 CAS IR Grid 体系中来。三是加强 CAS IR 软件在复杂类型内容对象及其语义关系管理的功能扩展, 积极遵循开源软件的精神, 尽快实现将 CAS IR 软件以开源方式发布, 为推动国内机构知识库的发展积极贡献力量。四是加强 CAS IR Grid 与国际著名数字知识库网络 COAR (Coalition of Open Access Repositories) 等的合作, 积极支持和促进中国科学院的知识成果在国际国内的开放获取和共享传播。

中国科学院国家科学图书馆