

基于DOI RAs机制的科学数据管理与共享*

——“第一届中日韩DOI项目合作讨论会”概述

□ 李颖 / 中国科学技术信息研究所 北京 100038

刘静波 / 河北省图书馆 石家庄 050011

摘要: 第一届中日韩DOI项目合作讨论会于2013年5月30日在中国科学技术信息研究所成功举办,这也是第三届中日韩科技信息机构联席会议的专题讨论会。针对ISTIC、KISTI、JST三家的DOI RA职能,会议围绕DOI RA在科学数据管理与共享方面的共识,三家机构对各自的系统及其相关活动进行了较为深入的讨论。文章在概述本次会议的基础上,重点介绍中日韩目前的系统,即中文DOI、日本JaLC、韩国的P-CUBE,以及科学数据管理工作。最后,总结了三机构未来实施合作的意向。

关键词: ISTIC, KISTI, JST, 中文DOI, JaLC, P-CUBE, 生物科学, 科学数据

DOI: 10.3772/j.issn.1673—2286.2013.08.007

1 第一届中日韩DOI项目合作讨论会介绍

第一届中日韩DOI项目合作讨论会议于2013年5月30日在北京举办,这也是第三届中日韩科技信息机构联席会议的DOI专题讨论会^[1]。参与方为中国科学技术信息研究所(ISTIC)、韩国科技信息研究院(KISTI)、日本科学技术振兴机构(JST)。会议在ISTIC乔晓东总工

(见图1,第一排左数第5位)的指导下,三机构代表在第一、二、三届总会的基础上^[2-4],针对基于DOI RA机制的科学数据的管理与共享的议题,进行了深入的交流与讨论(图2)。会议由ISTIC信息技术支持中心朱礼军主持,JST国家生物数据中心计划管理部主任白木泽佳子等三人、KISTI科学大数据研究部李相镐首席研究员等两人、我所资源共享中心彭洁主任和信息技术支持中心朱礼军副主

任以及我所从事DOI项目的相关人员共计约20人参加了此次会议。

来自中日韩三个机构的六名代表作了发言,发表者分别介绍了DOI在中国科学数据领域的应用及中文DOI最新进展、日本生命科学数据共享以及日本DOI进展、韩国KISTI科学数据共享活动及科研数据平台(P-CUBE)。会议代表还讨论了如何加强科学数据和电子期刊元数据交换、韩国科研数据平台



图1 会场内参会者合影



图2 会场发言

*本研究获得中国科学技术信息研究所科研项目预研资金“基于DITA标准的语义知识系统框架设计及应用研究”项目的部分资助。获得国家科技支撑计划“技术创新服务平台关键技术研究与应用示范”项目的资助(编号:2011BAH30B01)。

(P-CUBE)系统在中国和日本的测试和使用等事宜。

目前,韩国科技信息研究所虽然担负着DOI注册机构(DOI RA)的职能,还没有申请DOI RA。我所是DOI理事会成员,韩国信息院所希望我所能够协调它成为DOI韩国代理,我所表示愿意积极推进此事。

此外,此次参会代表还讨论了是否扩大中日韩科技信息机构联席会议的规模,邀请三国更多的科技信息机构参加该会议,并商定将此事作为第四届中日韩科技信息机构联席会议的议题之一。

以下针对中日韩DOI系统,即日本JaLC、韩国的P-CUBE,进行重点介绍,并介绍在系统之上的科学数据管理与共享工作。在总结部分,介绍三机构未来实施合作的意向。

2 中文DOI现状、科学数据工作

2.1 中文DOI现状

2007年3月,ISTIC和北京万方数据股份有限公司联合申请并取得了中文领域的DOI注册与管理

权,成为IDF授权的中国DOI注册机构(DOI RA),并于2007年7月开始正式运行中文DOI服务。这是我国在数字出版及信息内容服务行业向国际化方向迈出的重要一步。中文DOI服务从2011年9月开始交由ISTIC的DOI注册与服务中心管理和运行。2013年起,DOI注册与服务中心工作由ISTIC国家科技信息资源综合利用与公共服务中心负责,以加强DOI的知识应用研究及国际合作工作。2013年度的重点研发任务是基于关联数据的DOI数据发布、影视资源的扩展应用(详见本专题的相关文章),与ORCID(Open Research Contributor ID)、CNRI及其DONA(Digital Object Naming Authority Foundation)、CrossRef、JST和KISTI继续进行深入合作。同时,继续参与国内外相关标准的制定及活动,比如PDR(CY/T 82-2012, Press and publication digital resource identifier)、ISNI(CY/T 83-2012 China standard name identifier)、ISO 26324、ISO 27729及ISBN-A等。

目前(2013年6月1日统计)累计

注册总量已达17,657,515条。涵盖5900种期刊的16,364,969条论文数据、400所大学的学位论文1,276,986篇、8857条科学数据,其他数据5714条^[5],并启动了图书的DOI注册工作。作为实验,注册了与期刊论文和科学数据相关的图表数据。

2013年,中文DOI系统将升级为3.0版,持续推出新的服务,包括DOI引文链接、被引统计、多重解析注册等。通过系统重构,有利于扩展。重新设计门户、中文DOI成员的应用体系、统计功能、强化元数据查询功能和桌面工具、机器接口等^[6]。

2.2 科学数据工作^[7]

科学数据现状见图3。

在我国,项目结束后,数据很难继续管理和共享。在此环境下,ISTIC资源共享中心,在科学数据领域,通过DOI的应用,力图促进科学数据的管理与共享。近年开展的主要工作如图4所示。

引用DOI机制,有利于科学数据的发现、获取、跟踪、引用、重用,使其可见度增高,并可核实与跟踪其影响因子。



图3 科学数据管理存在的问题及基于DOI的解决方案

3 日本JaLC现状、科学数据工作

3.1 日本JaLC现状^[8,9]

JaLC是整合公共信息服务的日本国家项目。如图5, JaCL的主要功能如下:

- 下列资源DOI的注册机构
 - 学术期刊论文
 - 博士论文
 - 大学纪要
 - 报告

■ 利用DOIs和PubMed IDs在数据库之间提供双向链接, 比如引用及被引链接。

通过日本DOI RA机制, JaLC还将实现如下的服务:

- 提供日语DOI注册服务
- 多种数字内容的全国范围的解析(比如电子期刊、学术论文、研究报告、纪要)
- 兼容多字节(比如汉字)的元数据检索服务
- 促进DOI在日本的应用
- 促进全球用户检索日本公开的学术信息

JaLC工作的挑战如下:

- 内容的重复
- ◆ 机构仓储
(Institutional Repository, IR)与出版社之间
 - ◆ IR对多著者论文的IR
 - 处理作者不同版本的论文
 - 非数据存储用户的服务
 - 图书、科学数据的存储
 - 科学数据注册问题:
 - ◆ 范围: 由于粒度不同, 需要决定目标DOI的范围
 - ◆ 元数据政策: 不同领域元数据描述不同, 需要共识与清晰的政策

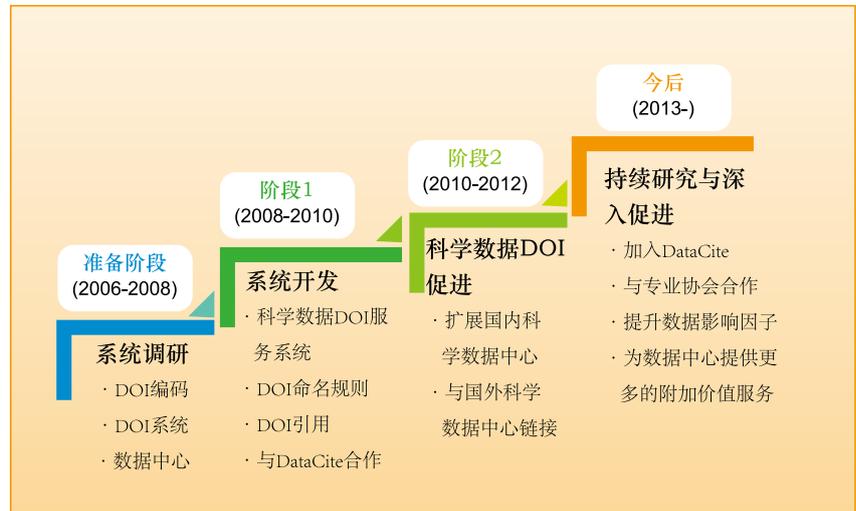


图4 ISTIC资源共享中心工作

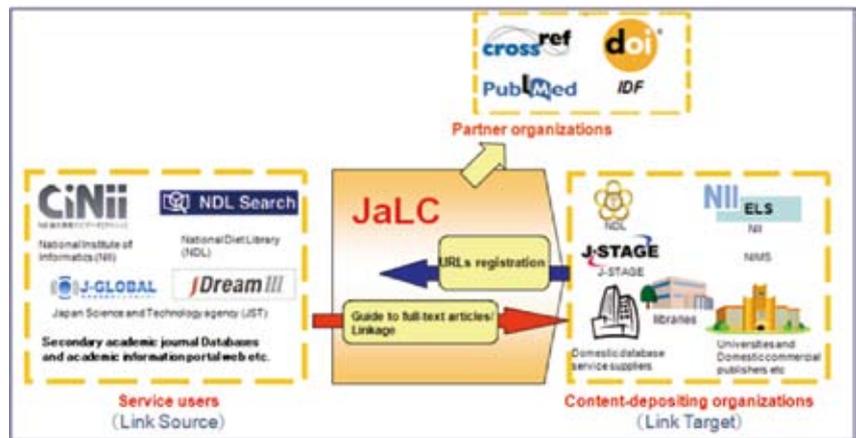


图5 JaLC的服务功能

◆ 可扩展性: 需要可扩展的系统来适应元数据的修改

小结:

- 2013年3月, JST被批准为IDF第9家RA
- 国立国会图书馆NDL、日本国立情报学研究所NII、国立材料科学研究所NIMS及JST共同设立了JaLC的RA指导委员会, 总部在JST, 是JaLC项目的管理实体。
- DOI是日本全国范围元数据整合的关键基础设施
- 通过日语资源的定位来充分利用DOI

3.2 科学数据的应用^[10-12]

如图6所示, JST国家生物数据库中心NBDC的主要业务是:

- 1) 方针制定
 - 2) 门户的构建与维护
 - 3) 数据库整合基础技术的研究与开发
 - 4) 生物数据库的整合促进
- 通过NBDC的门户网站, 用户可以检索日本构建的各种生命科学数据库。

NBDC实施公共资助下的科学数据共享活动。实施数据共享的理由是:

- 增强研究的透明性与再现性
- 加速研究
- 减少研究过程的重复与资助的重复
- 为数据发现导入新的方法
- 促进发达与发展中国家之间的教育与合作研究

2013年4月, NBDC公布了人类数据共享及其数据安全处理的指导方案。人类数据将通过NBDC的人类数据库共享, 包括DNA序列、人脑影像及临床信息, 见图7。

JST的NBDC建立数据共享的动机:

- 强化和拓宽基金机构及政府所有的数据共享政策
- 对数据共享的研究者给予适当的信誉
 - ◇ 建立引用标准
 - ◇ 建立数据应用和重用的评价指标
- 可持续和永久标识
 - ◇ 永久标识符 (DOI, 或PURLs)
 - ◇ 持续的资助

4 韩国P-CUBE系统、科学数据活动

4.1 科学数据管理与共享系统P-CUBE^[13,14]

P-CUBE是KISTI开发的全球科学数据管理与共享的开放平台, 完全基于国际标准和开源软件, 见图8所示。

有关P-CUBE诞生背景、P-CUBE功能特点、P-CUBE架构体系、P-CUBE技术标准、P-CUBE主要功能, 本专题有专门的介绍, 在此不再赘述。

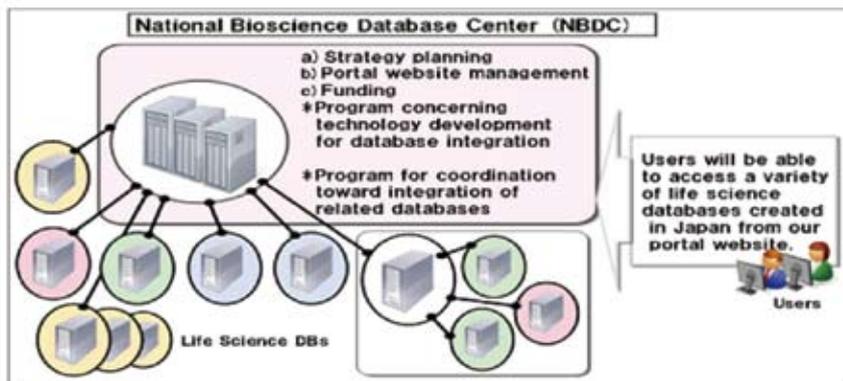


图6 NBDC业务



图7 NBDC人类数据共享

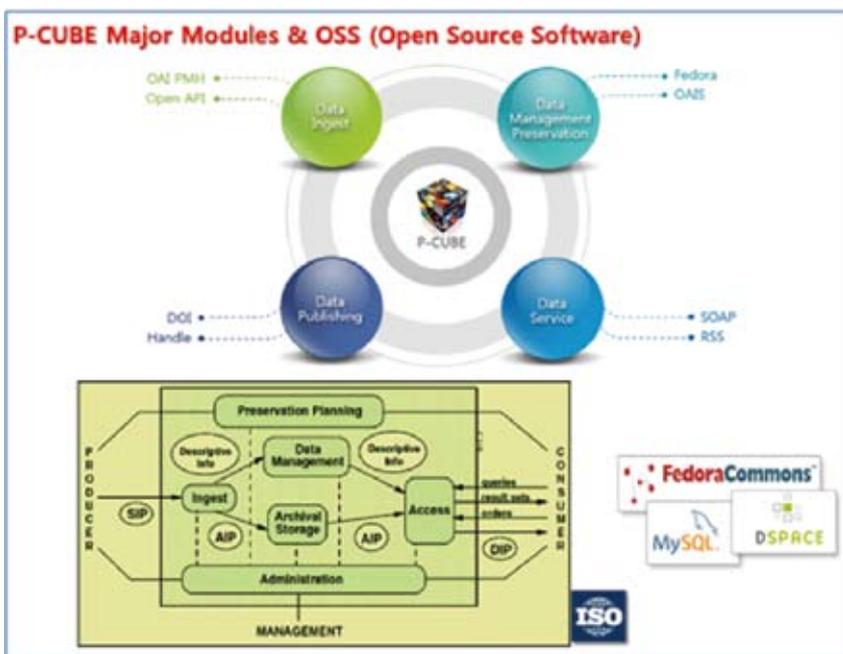


图8 P-CUBE主要模块和开源软件

4.2 科学数据工作^[15]

KISTI目前的科学数据项目有:

- CODATA活动: 国家代表
- GBIF: 生物多样性数据和

GBIF韩国节点 (<http://www.kbif.re.kr/index.do>)

- 事实数据库的开发
 - ◇ 医药图像、3D、形状测量数据

◇ 政府和研究机构建立的其他事实数据

与科学数据有关的活动:

■ 改变战略

◇ 构建科学数据库

◇ 收集科学数据

◆ 只要政府支持研究项目,科学数据将继续产出

◆ KISTI实施政府层面的研究项目的管理系统NTIS

◆ 所有资助机构都加入了NTIS项目

■ 支持机构数据的收集

◇ 开发和发布数据收集工具P-CUBE

◇ 支持研究机构、学会及大学构建数据仓储

◇ 支持研究者数据活动

◇ 支持政府层面的资助活动

◇ 与资金机构、学会、研究机构、大学合作,促进数据收集活动

5 中日韩三机构科学数据管理与共享合作意向

通过本次交流,中日韩在科学数据管理与共享方面的初步合作设想为:

■ 通过中日韩国家科技信息研究机构的会议,公共努力,促进

合作

■ 经验交流(问题、成功案例等)

■ 共享软件基础,比如P-CUBE

■ 交换元数据

■ 采用共同的标准,比如OAI-PMH或元数据格式

■ 设立新的合作项目

三国合作的基本路线是: KISTI在秋季前,将完善P-CUBE平台,交由ISTIC和JST进行测试和试用,三机构基于P-CUBE尝试进行生物科学数据管理与共享,并向其他领域扩展。

参考文献

- [1] 首届日中韩DOI会议报道.JST北京办事处[OL]. [2013-06-16]. http://www.jst.org.cn/activity_201306_03.html.
- [2] 李颖,徐硕,朱礼军,等.面向技术创新服务与DOI应用的国际合作综述:“第三届中国科技信息机构联合研讨会”概述[J].数字图书馆论坛,2012(8):50-60.
- [3] 李颖,梁冰,乔晓东.“第二届中国科技信息机构联合会议”综述[J].数字图书馆论坛,2011(6).
- [4] 李颖,乔晓东,梁冰.科技创新信息支持门户的对比研究:中日韩三国信息系统的介绍与展望(第三届中国科技信息机构联合研讨会)[J].数字图书馆论坛,2010(5):2-8.
- [5] 中文DOI官网[OL]. [2013-06-16]. <http://www.chinadoi.cn>.
- [6] GUO X F. Updates on Chinese DOI. [2013-05-30].
- [7] TU Y. DOI application for scientific data in China [OL]. [2013-05-30].
- [8] JaLC官方网站[OL]. [2013-06-16]. <http://japanlinkcenter.org/jalc/>.
- [9] KATO A. Review of JaLC project and the updates [OL]. [2013-05-30].
- [10] KOSO A. Sharing Publicly Funded Scientific Data - Activities at NBDC [OL]. [2013-05-30].
- [11] JST国立生物科学数据库中心官网[OL]. [2013-05-30]. <http://biosciencedbc.jp/>.
- [12] 大波純一,佐藤恵子,白木澤佳子,等. JST バイオサイエンスデータベースセンターにおけるデータベース統合への取り組みと医学分野での利活用の紹介[J].情報科学技術,2013,63(5):179-185.
- [13] KIM S. Research Data Platform and Development of DOI System [OL]. [2013-05-30].
- [14] KIM S. What is P-CUBE? [OL]. [2013-06-16]. <http://or2013.net/sites/or2013.net/files/What%20is%20P-CUBE.pdf>.
- [15] LEE S H. Recent Scientific Data Project in KISTI [OL]. [2013-05-30].

作者简介

李颖, 信息系统专业博士。研究方向: 语义知识组织, 基于主题的知识组织技术的应用等。E-mail: liying@istic.ac.cn
刘静波, 体育教育学专业硕士。研究方向: 资源管理, 数字资源构建。E-mail: ryuseiha@sina.com

Management and Sharing of Scientific Data Based on the DOI RA Mechanism — Review for "the 1st Korea-China-Japan Cooperation Seminar of DOI-related Projects"

Li Ying / Institute of Scientific and Technical Information of China, Beijing, 100038
Liu Jingbo / Hebei Library, Shijiazhuang, 050011

Abstract: The 1st China-Japan-Korea Joint Seminar on DOI Project Cooperation was successfully held on May 30, 2013 in ISTIC, China. This is also a special workshop of the 3rd China-Japan-Korea Joint Seminar on Science and Technology Information. With three DOI RA functions of ISTIC, KISTI, JST, around the consensus of DOI RA application in management and sharing of scientific data, the three agencies conducted a more thorough discussion on their system and their related activities. In this paper, overview of the meeting is presented, focused on China, Japan and Korea's current systems, namely China DOI, P-CUBE and JaLC, as well as activities for scientific data management. Finally, it summarizes the cooperation intention of the three agencies in the future.

Keywords: ISTIC, KISTI, JST, China DOI, JaLC, P-CUBE, Bioscience, Scientific data

(收稿日期: 2013-06-22)