

科学数据整合与管理开放平台P-CUBE

□ 刘静波 / 河北省图书馆采编部 石家庄 050011

李颖 / 中国科学技术信息研究所 北京 100038

摘要: 科学数据的整合与管理是国家信息资源战略方针之一, 目前还没有长期有效的方法与手段。韩国科学技术信息研究院在此领域走在了亚洲前列。其最新研制开发的“Big”科学数据整合与管理开放平台P-CUBE就是很好的解决方案。文章介绍P-CUBE的定位、架构、技术体系和应用。最后, 得出结论。

关键词: KISTI, P-CUBE, 科学数据, 管理平台

DOI: 10.3772/j.issn.1673—2286.2013.08.008

序言

科学数据是科研活动的重要产出之一。在科研活动的支撑技术手段高度发达的当今社会, 研究人员在其研究过程中产生的科学数据的量堪称“big”。然而, 大部分科学数据分散在研究者手中, 这些有价值的信息, 要么得不到重用, 要么随着时间的流逝而丢失。为此, 收集、整合、管理科学数据, 重用和共享这些数据的平台被认为是有效的解决方案。

到有效的利用, 为此, 需要管理科学数据、并重用这些数据的平台。

P-CUBE由此诞生。

本文重点介绍这一刚刚问世的开源软件的架构体系, 从P-CUBE的数据生命周期与主要角色、不同角色的作用、P-CUBE主要模块与OSS、数据模型架构、系统架构、接口, 以及P-CUBE的模型I和II等多个方面, 对P-CUBE进行描绘。

通过对P-CUBE的描述, 让研究者理解P-CUBE, 轻松地使用

P-CUBE, 从而实现全球、特别是中日韩亚洲大国之间的科学数据共享。

1 P-CUBE定位

负责P-CUBE研究开发的KISTI资深研究员SunTae Kim博士在描述“*What is P-CUBE?*”时, 给出了其定义: “大电子资源”整合研究与统一的平台。P-CUBE取自于“Platform for Convergence

P-CUBE诞生环境

◆ **设计目标:** P-CUBE是一个便于存储和重用研究人员在科研过程中产生的科学数据的安全平台, 是云环境下的科学数据的高效管理和应用系统。

◆ 由于IT硬件的发展、超高速网络的扩展及高端信息技术的出现, 科研活动产生了大量的科学数据。遴选和管理有价值的科学数据非常重要。由于数据大部分存储在科学家的PC、CD或USB中, 没有得



图1 P-CUBE与数据生命周期

research and Unification of Big E-resources”中的几个英文词汇的首字母。其关键词是：

- ◆ 平台
- ◆ 安全存储、方便的获取
- ◆ 科学数据的重用

图1从数据的视角，给出了P-CUBE在数据生命周期的定位、数据生命周期中涉及的角色及其作用。可以说，P-CUBE融合和统一了数据的全流程管理，必不可少。

2 P-CUBE架构

从图2 P-CUBE主要模块与开源软件可以看出，P-CUBE平台完全基于国际标准化体系和通用的开源系统。其数据摄取模块采用元数据收割标准OAI-PMH，数据存储管理采用开源软件FEDORA，数据发布基于DOI，而数据服务利用SOAP协议。所以说，P-CUBE是全球化的开放平台，任何科研人员、机构、本地或数据中心都可应用P-CUBE进行数据管理、数据共享和重用。

参见图3，P-CUBE数据架构。

P-CUBE提供数据的存储、链接、管理和服务功能。具体如下：

P-CUBE 主要功能：

(1) 收集科学数据

- ◆ 通过研究者数据上载来收集数据
- ◆ 通过标准协议自动收集数据

(2) 管理和存储科学数据

- ◆ 利用OAIS标准系统地管理数据
- ◆ 基于数据生命周期存储管理

(3) 科学数据发布功能

- ◆ 对研究人员的数据分配全

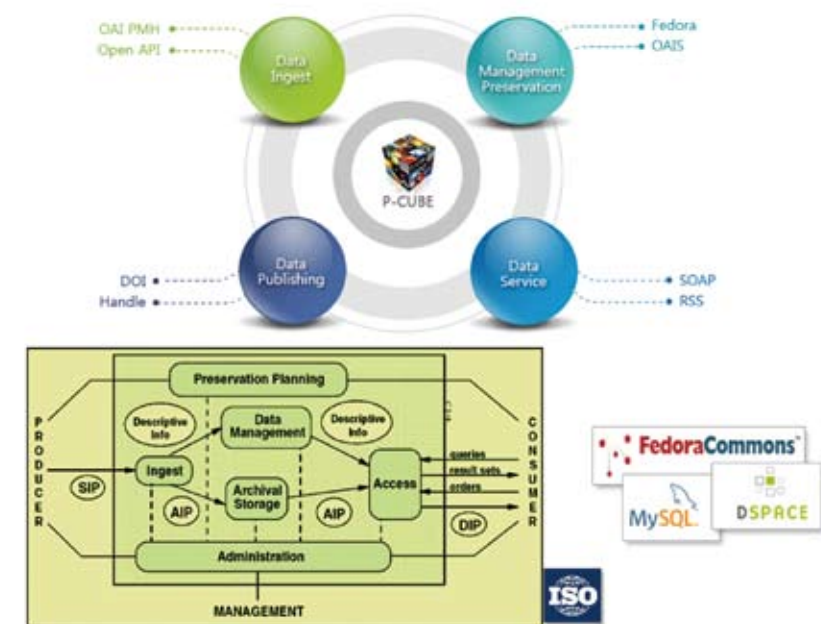


图2 P-CUBE主要模块与开源软件

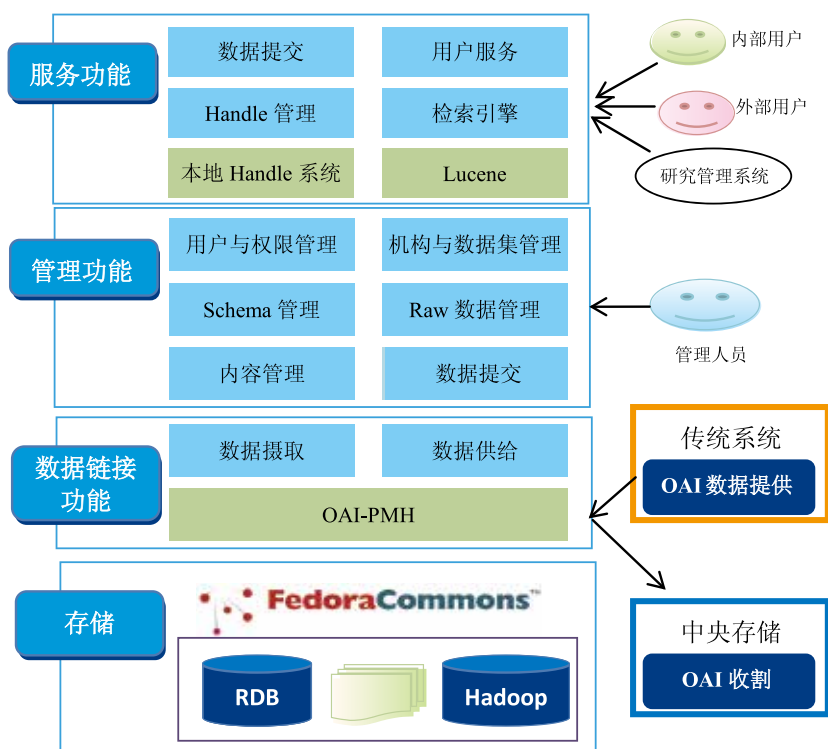


图3 P-CUBE数据架构

球标识符

采用DOI Handle机制

(4) 科学数据服务功能

- ◆ 依据组织、集合和研究者创建的标签检索数据

◆ 科学数据与学术期刊的链接服务

P-CUBE功能特点：

- ◆ 科学数据的系统管理（组织、收集）

◆ 研究人员可公开构建自身的科学数据（默认为非共享）

◆ 可构建科学数据的各种管理项目

◆ 科学数据全球发布和永久获取

◆ 基于OAI-PMH标准自动收集和发布科学数据

P-CUBE的基础:

◆ 基于世界公认的开源系统开发, 采用Fedora及MySQL数据库。

P-CUBE 技术标准:

◆ P-CUBE 遵循 ISO 14721:2003中规定的 OAIS参考模式

◆ P-CUBE的系统架构包含数据收集、数据管理、归档和检索

3 P-CUBE技术体系和应用

如图4所示, P-CUBE在技术体系上, 采用了本体描述, 基于Fedora Commons Ontology、DataCite Ontology、以及DC Ontology等主要的对象类型属性^[4-6]。

P-CUBE数据模型(参见图5):

- ◆ 组织organization
- ◆ 集合collection
- ◆ 项目item
- ◆ 文档file

P-CUBE的模式设计为图6和7两种: 论文+DOI模式和论文+数据。

P-CUBE的应用模式设想如图8所示。

目前, KISTI的P-CUBE研发队伍正基于如上的架构体系、标准规范及应用设想, 对P-CUBE进行

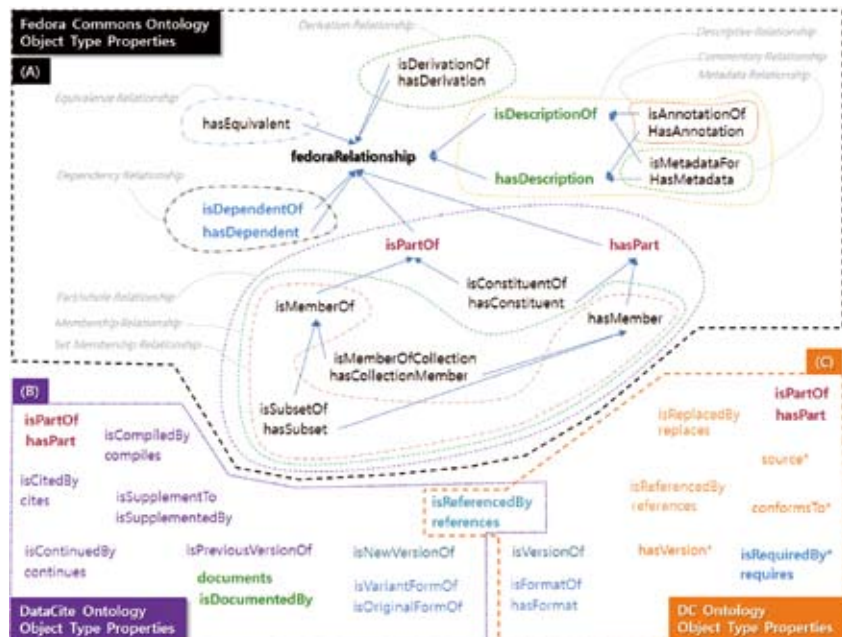


图4 P-CUBE应用的本体

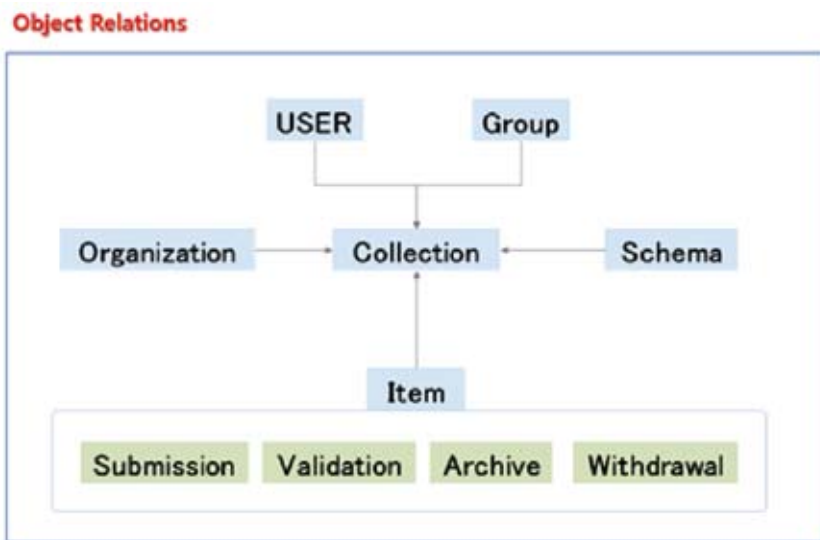


图5 P-CUBE对象关系

开放前的全面测试, 并完善技术文档, 以利于其全球化应用。预计秋季完成。

结论

P-CUBE全部模块采用国际标

准技术规范, 是通用的科学数据管理平台, 可在任何国家和任何领域应用。通过在不同领域、不同国家的安装试用, 可以实现全球化的科学数据管理与共享服务, 作为人类福祉, 被寄予厚望。

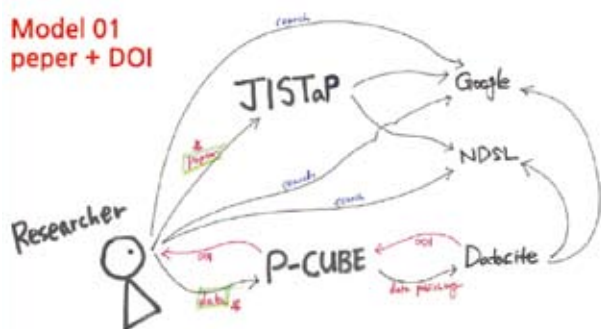


图6 论文+DOI模式

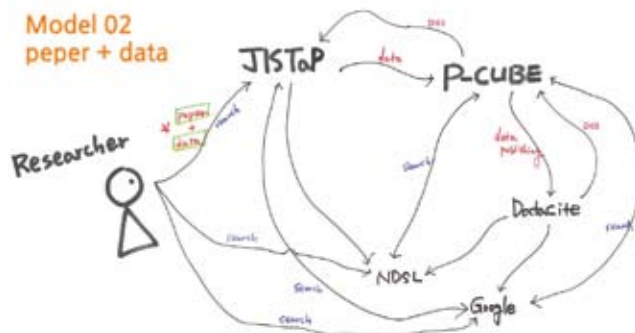


图7 论文+数据模式

(*NDSL: National Digital Science Library, 韩国国家数字科学图书馆)

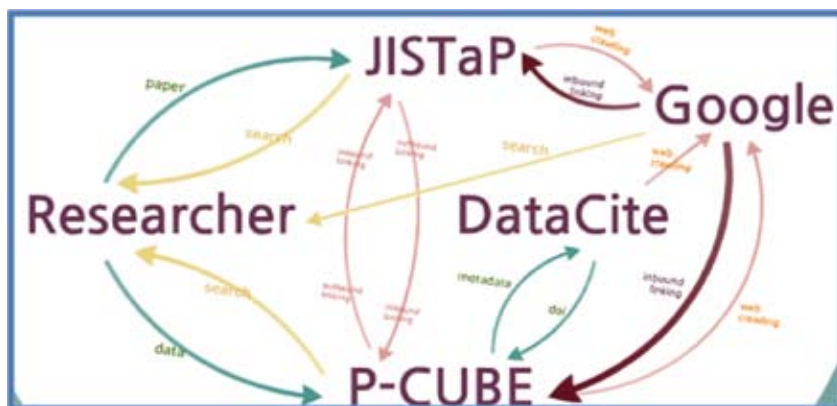


图8 P-CUBE的应用模式

参考文献

- [1] KIM S. What is P-CUBE? [OL]. [2013-06-16]. <http://or2013.net/sites/or2013.net/files/What%20is%20P-CUBE.pdf>.
- [2] KIM S. Research Data Platform and Development of DOI System [OL]. [2013-05-30].
- [3] KIM S. P-CUBE: Research Data Platform [C]//中日韩合作会议(2013.5.30)资料.
- [4] Fedora Commons官网[OL]. [2013-06-16]. <http://www.fedora-commons.org/>.
- [5] DataCite官网[OL]. [2013-06-16]. <http://www.datacite.org/>.
- [6] Dublin Core官网[OL]. [2013-06-16]. <http://dublincore.org/>.

作者简介

刘静波, 硕士, 体育方法学专业。研究方向: 教育学, 图书信息管理, 数字资源的构建等。E-mail: ryuseiha@sina.com
 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

P-CUBE: Open Platform for Convergence and Management Scientific Data

Liu Jingbo / Hebei Library, Shijiazhuang, 050011

Li Ying / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: Integration and management of scientific data is national strategy of information resources. Now, there are no long-term and effective means to handle it. Korea Institute of Science and Technology Information (KISTI) is in the top in Asia. P-CUBE, a platform for convergence research and unification of Big E-resources developed newly by KISTI, is a good solution. This article describes its concept definition, architecture, and technical systems and applications. At last, the conclusion is given.

Keywords: KISTI, P-CUBE, Scientific data, Management platform

(收稿日期: 2013-06-22)