

中文 DOI 系统的研究与展望

——从注册解析基础应用到基于关联数据知识服务

姚长青 李 颖 郭晓峰

(中国科学技术信息研究所, 北京 100038)

摘要: 目前Handle系统已成为下一代互联网数字内容基础设施的候选标准, 基于Handle系统的中文DOI系统也经历了从传统注册解析向知识化服务的转变。文章从发展历程、现状以及发展规划等方面, 揭示中国科学技术信息研究所中文DOI系统的研发工作, 从基础的注册解析应用到基于关联数据的知识服务, 进行不断的升级, 以促进中文DOI事业的深入发展。

关键词: 中文DOI; DOI系统; 关联数据; 知识服务; 影视资源

中图分类号: G353

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2013.06.008

Review of Research and Development of China DOI System

—From Basic Application of Registration and Resolution to Knowledge Service of Linked Data

Yao Changqing, Li Ying, Guo Xiaofeng

(Institute of Scientific & Technical Information of China, Beijing 100038)

Abstract: As the Handle System has become the infrastructure of digital content for the Next Generation Internet, China DOI based-on it have to go forward to knowledge service. ISTIC, as the institute for managing and operating China DOI, its research and development continues upgrade from the basic applications of registration and resolution to the knowledge service of Linked Data. This article, from the main aspects of development process, current situation, and development plan for China DOI, fully discloses the relevant projects of ISTIC, in order to promote the further development of China DOI in the domestic.

Keywords: China DOI, DOI system, linked data, knowledge service, movie and television resource

1 引言

数字对象唯一标识符(Digital Object Identifier, 简称DOI)最早是由美国出版协会(AAP)提出并建立的一个由国际DOI基金会(International DOI Foundation, 简称IDF)作为DOI最高管理机构、美国国家创新研究所CNRI提供Handle System技术支持、多个注册机构并行的标识体系。其目的是建立

互联网环境下数字对象标识及知识产权管理和保护的解决方案。最近, 国际电信标准制定机构正研讨将Handle系统作为互联网的下一代基础设施进行全球化标准管理^[1]。Handle系统的基本思想源于数字资源管理的通用方法, 即数字对象架构(Digital Object Architecture, DOA)^[2]。DOA具有可扩展的数字资源对象描述体系(元数据)、语义互操作等功能。DOI是Handle系统在全球最主要及其最成功的

第一作者简介: 姚长青(1974—), 男, 中国科学技术信息研究所副研究员, 博士, 研究方向: 科技评价、情报技术分析。

基金项目: 中央级公益性科研院所基本科研业务费专项资金“基于数字对象唯一标识符的知识管理建设与研究”(ZD2010-6-2); 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”(2011BAH30B01)。

收稿日期: 2013年10月29日。

应用。DOI、数字对象标识与标准对于促进网络环境下的资源管理与利用具有深远的影响力。

2007年，国际DOI基金会（International DOI Foundation，简称IDF）正式授权中国科学技术信息研究所（Institute of Scientific & Technical Information of China，简称ISTIC）联合万方数据股份有限公司成为其DOI注册机构（Registration Authority，简称RA），负责管理和运营中文DOI系统，在中文领域开展DOI注册、解析和应用等业务。中文DOI将DOI服务正式带入了中国。中文DOI系统的实施实现了资源链接的本地化。通过提供指向馆藏全文信息的永久性链接来增加已获得资源的可用性，实现自有文献资源使用效益与覆盖范围的快速增长，以及中西文文献的有效链接。由此可见，基于DOI的服务促进了中文资源的链接与共享。

近年来，中文DOI系统在中文资源注册、解析等方面取得长足的发展。截至目前，注册的DOI总数已超过1900万个，从2009年起位居全球DOI注

册机构中的第二位。CrossRef是DOI最大的DOI注册机构，于2013年提出了基于关联数据的DOI服务技术^[3]。中文DOI理应顺应时代的发展，从注册解析基础应用到基于关联数据知识服务。在我国，尽管注册和解析量逐年增长，但在数字内容知识服务等方面还需进一步提升，逐渐向资源内容的语义化、知识化服务的方向发展。图1展示了中文DOI从申请、IDF批准、系统研发、上线服务到服务内容扩展的近7年的发展历程。本文将在探讨中文DOI的发展现状以及中文DOI在线服务现状的基础上，进一步分析基于关联数据的中文DOI原型系统的研发，对中文DOI系统知识化服务进行试验研究，最后提出中文DOI系统升级改造的发展方向。

2 中文DOI在线服务

依据2013年6月1日统计，中文DOI网站的累计注册总量已达1765万多条，涵盖5900种期刊的1636万多条论文数据、400所大学的学位论文127

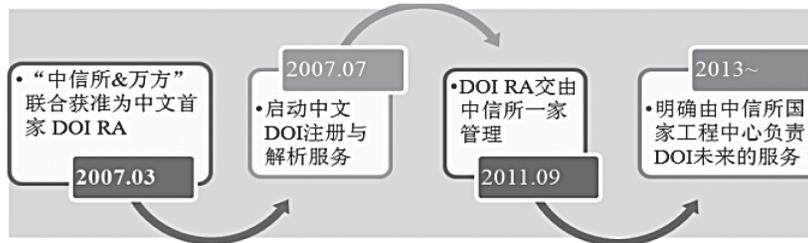


图1 中文DOI发展历程

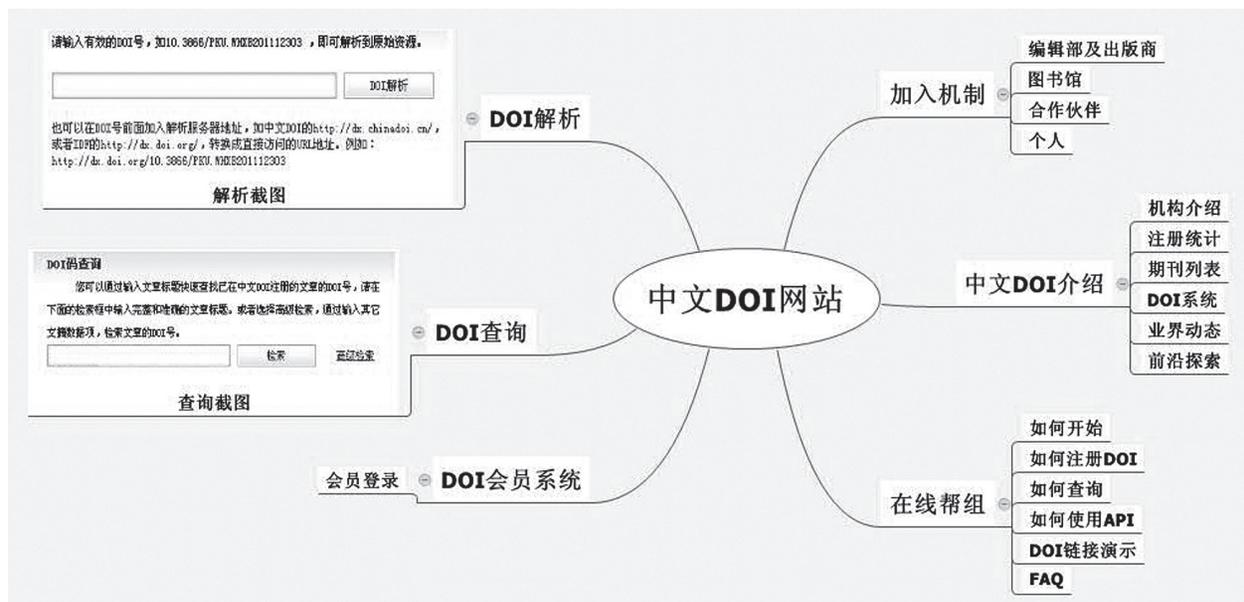


图2 中文DOI在线服务

万多篇、8857条科学数据，其他数据5714条，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。

2.1 中文DOI网站服务

中文DOI系统的开发实施已经历了两个阶段：第一阶段侧重于中文期刊论文和科学数据的DOI注册、解析及查询功能等；第二阶段侧重系统的优化和功能丰富。系统目前处于第三阶段，需要实行系统的升级，以适应环境的发展，实现更高的服务能力与水平。正在运行的中文DOI网站的结构如图2所示，主要提供期刊资源的DOI注册解析与查询功能，并具有其他的辅助功能。用户可以通过图示，获得相应的服务，并考虑申请成为中文DOI会员。

2.2 DOI网站架构

2012年9月，中文DOI系统升级为3.0版，包括DOI引文链接、被引统计、多重解析注册等。对中文DOI系统的业务逻辑进行重新设计（图3），实现系统功能的进一步扩展。重新设计了门户、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。

中文DOI系统3.0版具有以下新特性。

(1) 新增了电子书DOI的注册。为了满足出版机构对于电子书DOI服务的需求，开发了电子书DOI的注册及数据修改、解析功能。

(2) 新增了解析数据统计功能。每月定时从解

析服务器下载解析日志，统计DOI的解析情况存储到数据库中。

(3) 改善了DOI/元数据查询功能，提高了查询效率及性能，能够更人性化地显示查询结果。同时，查询结果还提供DOI元数据标准引文格式，例如“彭洁.信息资源整合技术初探[J].中国科技资源导刊, 2008,40(01):40-46.doi:10.3772/j.issn.1674-1544.2008.01.008.”

(4) 完善了客户端注册工具。

(5) 完善了第三方系统接口。

3 基于关联数据的中文DOI系统研发

文献资源的知识服务是创新时代科技信息系统的发展方向。最新中文DOI的研发目标是：与知识服务有关的多类型信息资源的元数据整合注册及其基于关联数据的知识服务。利用中文DOI拥有的数字资源对象基础设施功能，通过多种资源的元数据统一注册、提供关联数据服务，实现新型的知识系统。

DOI系统由国际DOI基金会IDF及其DOI RA 研究开发，它是可适用于任何数字对象的通用框架，提供结构化和扩展可能的标识、描述和解析方法。中文DOI是IDF下属的第二大DOI RA，目前是发展速度最快的中文学术期刊资源的应用服务系统。现有中文DOI系统的主要服务是：通过中文期刊元数



图3 中文DOI系统业务逻辑^[4]

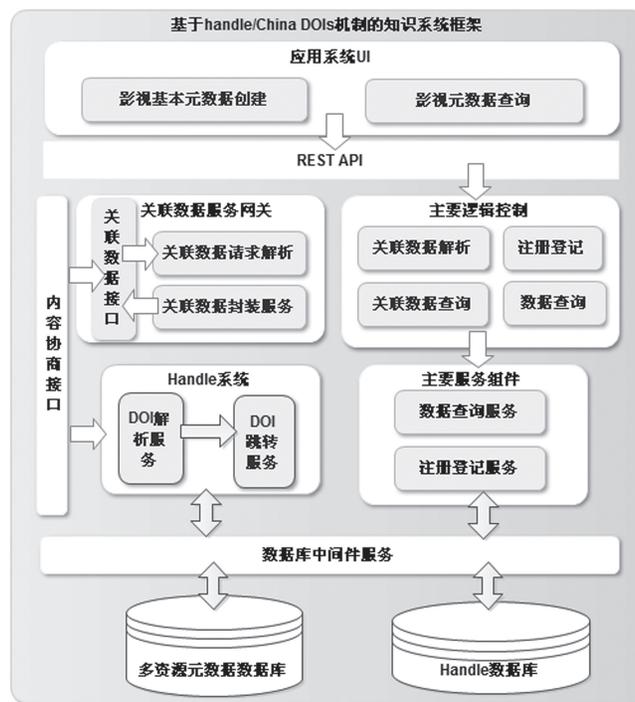


图4 中文DOI原型系统结构

据的注册与解析，提供从期刊引文到全文链接及其有关的附加价值服务，如基于数字资源权益元数据描述的DRM（Data Rights Management，数字权益管理）服务等。从原理上讲，中文DOI本身适用于任何类型的数字资源，可将资源元数据的注册，从期刊扩展到多类型资源的元数据的统一构建。同时，按照IDF的DOI与关联数据的工作机制（详见参考文献[3]），可以实现将中文DOI系统直接用于关联数据这一知识化服务。为了验证这一思想，ISTIC进行了有关原型系统的开发，其结构见图4。

具体地，中文DOI知识系统框架的概念设计包括：（1）将现有中文DOI中文期刊元数据数据库向影视资源扩展，实现多类型资源元数据的统一构建。DOI系统在影视行业的扩展应用被认为是最有经济价值且最有可能实施的工作。以影视资源元数据的构建来验证基于handle/China DOIs机制的多类型资源的元数据统一构建具有实际意义。本研究在原有期刊DOI元数据的构建与服务的基础上，通过原型系统实现了影视资源元数据与期刊元数据的整合，并实现了基于元数据查找影视资源DOI的功能^[5]。（2）将已注册中文期刊资源的中文DOI系统用于关联数据，即将DOI作为关联数据的HTTP URI定位与关联数字对象，实现关联数据服务。以期刊论文的关联数据服务为例，关联数据服务网关接收到转

换请求及头信息中包含application/rdf+xml的URI请求后，对请求信息进行解析，如果请求的是期刊论文的资源信息，就根据请求的DOI，从中文DOI系统的元数据库中查询相应的期刊论文元数据信息，包括所在的期刊母体URI，分类号、外部链接等信息，并将这部分元数据信息通过关联数据封装服务，封装成RDF/XML文件，由关联数据服务接口返回。通过语义浏览器Disco-Hyperdata Browser获得更直观的该论文的RDF信息展示^[6]。

4 基于中文DOI的知识管理原型研究

目前，ISTIC围绕自建资源的保存、管理、共享已建设了多个信息系统，如机构文献库、科研报告管理系统、科研资源共享库等，以满足不同应用的需要。这些应用系统存储管理了某些种类的信息资源，并具体提供相关的应用功能。同时，这些应用系统都有独立的用户身份和用户权限管理功能，以满足系统安全、应用等方面的需要。但在实际使用中，由于各系统中的信息资源、用户身份、数字权益等未实行统一的管理，很难统一实现各系统的信息互通，给信息资源的长期保存、动态发现、合理利用都带来了诸多不便^[7]。

已诞生和发展了20余年的DOA以及DOI方面的理论和实践，能够为解决上述问题，提升信息系

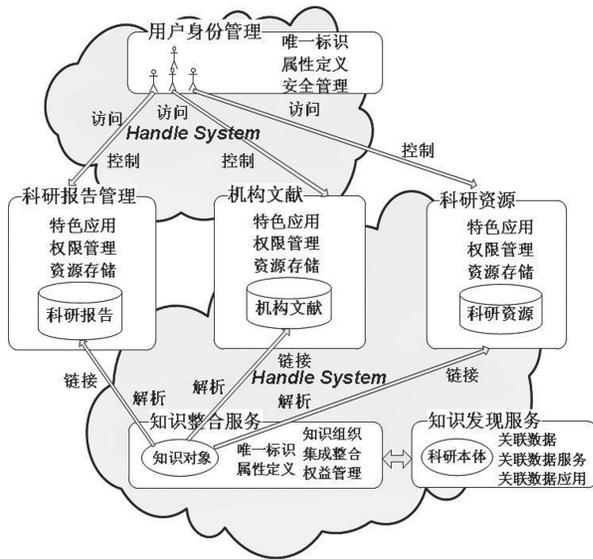


图5 基于DOI知识管理项目技术路线

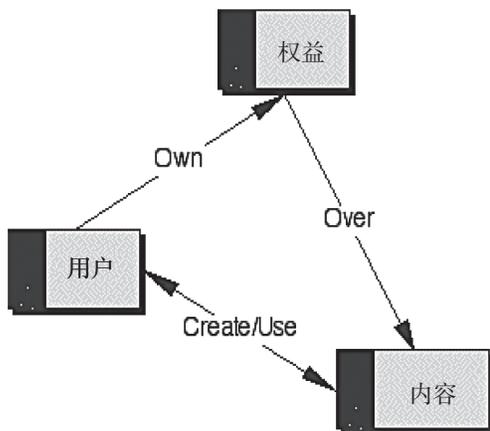


图6 DRM信息架构——核心实体模型

统建设的综合效能，提高信息资源的管理和利用效率，从而从根本上解决各相关系统的信息互通提供统一框架。应用中文DOI的多种资源的元数据统一注册及其关联数据服务，可以研发新型的知识服务体系。遵循这个思路，我们基于中文DOI实现机构知识库管理系统进行了研发试验，设计了基于DOI知识管理项目技术路线（图5），为ISTIC信息资源的长期保存、动态发现、合理利用提供基础服务。基于DOI知识管理项目的技术路线包括以下几条。

(1) 通过知识对象实现信息资源的集成和相互链接。在知识管理系统中建立知识对象，并在Handle系统中为知识对象注册唯一标识符、定义知识对象的属性，依托Handle系统的解析功能，灵活、易扩展的属性定义和维护功能以及安全管理功能等，方便地实现知识对象到信息资源的解析链接

及知识对象的相互关联。

(2) 通过建立统一的用户身份管理服务简化权限管理。建立用户身份对象，在Handle系统中为所职工建立统一的用户身份标识，并统一管理用户身份对象的属性和安全性等，不仅可以使用户的身份认证管理作为一种公共服务，与各应用系统的授权管理相对分离，从而减轻各应用系统管理用户身份负担，消除用户需要分别在各应用系统注册账户的负担，同时也可以保证更高的安全性和隐私保护。更重要的是，可以将用户身份作为公共数字对象在各应用系统中使用，方便地定义和维护用户身份与其他数字对象的关系，这样就为开放式的数字权益管理提供了基础。

(3) 通过建立开放式数字权益管理（DRM）框架实现知识的开放式管理。在图6所示的数字权益管理（DRM）核心实体模型中，用户、内容（信息资源）、权益三类实体都需要加以确认和描述，并且应该通过系统中每个实体的开放标准机制（如URI、DOI/Handle、ISTC等）来完成确认工作。因此，上述两个步骤为权益实体的描述和确认提供了基础。实质上，定义权益实体的属性在很大程度上即是定义用户实体和内容实体的关系。在2007-2009年ISTIC与美国全国研究创新联合会（CNRI）合作实施的国际科技合作项目“建立中国数字对象唯一标识符体系的研究与应用”中，开展了DRM方面的实验，将信息资源的权益元数据在Handle系统中注册和存储，各应用系统可以通过统一接口获得资源的权益元数据，并在应用系统中按照权益元数据的定义对资源的访问进行控制。本研究可以继续深化研究和实验，建立权益对象，对权益实体进行更加细致、灵活的管理。以开放式的权益管理为基础，就可以对知识的加工、集成、访问、利用等环节进行开放式管理，实现可控制、有隐私的知识共享。

(4) 通过关联数据实现知识发现服务和应用。利用关联数据工具和服务，可以将已有的数据变成机器可以理解的形式，并将异构的数据源连接起来，构建计算机可以理解的语义数据网络，为构建智能应用提供基础。数字资源的标识与描述是关联数据的核心原则。因此，ISTIC的各类数字资源，经过统一标识与注册，可使用关联数据工具进行整合，基于唯一标识建立科研资源的关联关系，应用语义技术，生成关联数据及应用接口并进行发布。其意义在于以统一的数据对象或实体为中心，将来

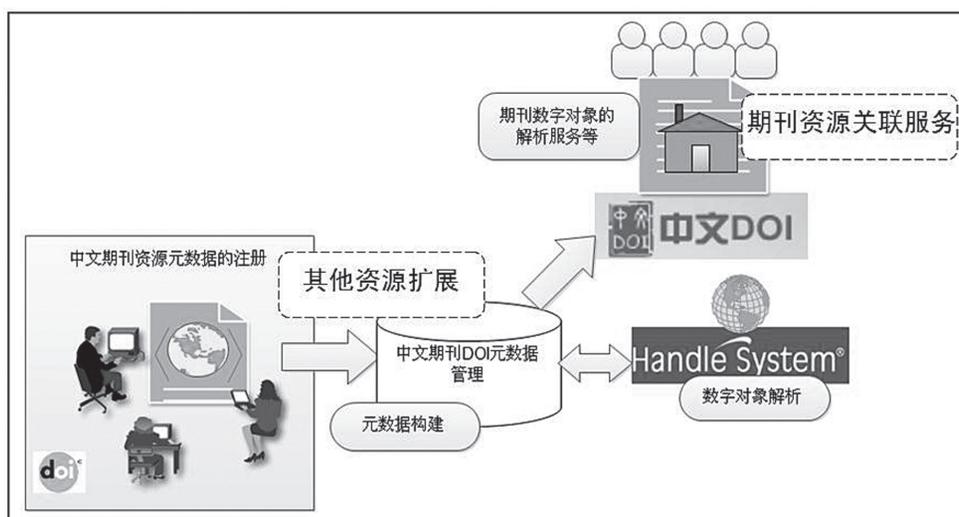


图7 面向知识服务的中文 DOI 框架

自不同科研管理系统的数据库，以统一的视图呈现给用户，从而促进 ISTIC 信息资源在该机构内外的发现和利用。

5 结论与展望

截至目前，中文 DOI 系统进行了具有前瞻性的知识系统的探索，开发了多资源元数据的统一构建与资源的关联数据服务，为学术信息资源有效的知识化服务奠定了基础。然而，要想实现真正意义上的多资源元数据的统一构建，还需要结合其他信息资源，如科学数据、学位论文、科技报告等，来验证元数据统一构建的有效性。通过关联数据应用，提升了中文 DOI 系统的知识化服务能力，但由于中文 DOI 元数据的结构化及语义化程度有限，因此，利用关联数据的知识化服务水平也亟待提高。这些都将是中文 DOI 系统面临的挑战。

ISTIC 将在未来与 DOI 其他 RA 的交流合作中不断汲取同行经验，同时，还将考虑将该机构拥有的语义元数据及关联数据的研究成果，有效地整合到中文 DOI 系统中，从而实现实用化的基于 DOI 的知识系统。具体的，基于 ISTIC 的现状以及原型研究开发结果，从知识化服务的视角出发，ISTIC 将从以下 3 个方面，对中文 DOI 系统进行了升级改造，并将在下一期的服务中实现。

(1) 各种科技信息资源的 DOI 元数据的整合服务。对各种科技信息资源的 DOI 元数据进行整合服务，实现元数据的结构化、模块化、语义化的统一存储与管理(图 7)。

(2) 融合科学数据等重要的数字对象资源。科学数据作为一种科技资源，在科技文献的知识化服务中的作用日益凸显。国外相关机构已经开展了 DOI 在科学数据上的应用，并取得了一定的成效。中文 DOI 在中国科学数据领域中的应用已经做了实验性的尝试，下一步将实现科学数据注册解析的实际应用^[8-9]。

(3) 知识服务技术的实用化——关联数据、本体技术。在目前开发的原型基础上，实现各种类型 DOI 资源关联数据的量化发布，提供 DOI 系统的语义化、知识化服务。同时，考虑将本体技术用于 DOI 系统，比如将科研本体、领域本体应用到 DOI 架构中。在更高的层面上，探索 DOI 系统的语义化、知识化服务。

参考文献

- [1] DONA [EB/OL]. [2013-09-21]. http://www.doi.org/doi_handbook/7_IDF.html.
- [2] CNRI. Overview of the Digital Object Architecture [EB/OL]. [2013-09-20]. <http://www.cnri.reston.va.us/papers/OverviewDigitalObjectArchitecture.pdf>.
- [3] DOIs and Linked Data: Some Concrete Proposals [EB/OL]. [2013-09-21]. http://www.crossref.org/Cross-Tech/2010/03/dois_and_linked_data_some_conc.html.
- [4] 李颖,乔晓东,杨兴兵.基于 DOI 系统的中日韩三国之间科技信息共享[J].中国科技资源导刊,2012,44(6): 81-85.
- [5] 李颖,时其迪,杨兴兵.中文 DOI 系统在影视行业的扩展应用——基于 EIDR 体系应用原型的研发[J].数字

- 图书馆论坛,2013(8):10-17.
- [6] 白海燕.关联数据在DOI系统中的应用与实现[J].数字图书馆论坛,2013(8):2-8.
- [7] 姚长青,乔晓东,赵蕴华,等.DOI在精品科技期刊全文数据库中的应用[J].中国科技资源导刊,2008(1):13-15.
- [8] 李颖,刘静波.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[J].数字图书馆论坛,2013(8):37-41.
- [9] 刘静波,李颖.科学数据整合与管理开放平台P-CUBE[J].数字图书馆论坛,2013(8):42-45.

(上接第33页)

共识与依靠专家共识判断之间的矛盾。CNRS设立了“快速与探索计划”，并将其分为两个部分：快速项目和早期概念性探索研究。后者专门用于支持处于初级阶段的、尚未验证的、具有潜在变革性研究思想或方法的探索性工作。

(3) 支持科学仪器设备

现代科技的进步越来越依靠前沿科学仪器的发展，科学技术重大成就的获得和科学研究新领域的开辟，往往是以检测仪器和技术方法上的突破为先导的。

目前，世界各国都把研发先进的大型科学仪器和实验实施，构建世界级实验设施基础平台，上升为创造世界一流科研成果、培养和吸引优秀人才的一项战略措施。CNRS设有对科学仪器和基础设施进行支持的专项基金，并提供仪器和设施共享的平台，对研究和教育设备、基础设施和装备、极地研究设施给予资助。其提供的精良仪器设备，极大地提高了研究和教育的整体能力。这些设施包括实验室设备和仪器、多用户研究设施、分布式网络设施、加速器、天文望远镜等。

(4) 重视国际合作与交流

法国科研基金将国际合作与交流作为重要的目标和任务之一。CNRS在2006年推出国际合作项目“国际研究和教育伙伴计划”，旨在鼓励法国科研机构与科学家与国外伙伴开展研究和人才培养方面进行长期合作与交流。项目要求申请者应有明确和突出的研究目标，必须在双方科研和教育合作交流的

基础上实施，且在项目申报中说明双方在该项目执行中的互补优势。法国国际科研中心资助的国际合作方式和项目种类有数十种，例如交换科研人员并进行合作研究，建立联合实验室，成立合作团队，为发展中国家的青年学者提供专项奖学金计划等。

参考文献

- [1] 徐治立.论政府对科技活动干涉的作用与限度[J].科学学与科学技术管理,2004(11):90-93.
- [2] 何斌,孙树栋.当代国际科技合作的经验和启示[J].科技管理研究,2008(12):78-80.
- [3] 王晓阳.美国联邦政府资助大学科研的政策与机制[J].高等教育研究,2009(3):105-109.
- [4] 张公一,卢艳秋.跨国技术联盟合作创新绩效机理研究[J].求是学刊,2011(6):63-69.
- [5] 张公一,吕文婷.国际科技合作创新博弈分析[J].哈尔滨工业大学学报:社会科学版,2012(1):132-136.
- [6] 孟太生.科学基金组织的变革管理及其经验比较[J].科研管理,2008,29(4):180-184.
- [7] 刘云,邹立尧.国外主要科学基金组织国际合作战略与资助特点及其对我国的启示[J].中国基础科学,2006,8(5):37-42.
- [8] 孙承晟.法国国家科研中心及其合作制度[J].科学文化评论,2008,5(5):46-59.
- [9] 朱卫东,周光中.国外科学基金绩效评估对我国的启示[J].中国科技论坛,2009(7):134-138.
- [10] 吕蓓蕾.法国国家科研中心国际合作战略和资助格局[J].中国基础科学,2004(6):45-47.