

嵌入式文献情报聚合服务平台的探索与实践

谢靖¹ 吴振新¹ 张智雄¹ 胡吉颖¹ 许丽媛¹ 周强¹

¹中国科学院文献情报中心 北京 100190

摘要: [目的/意义]传统文献情报资源服务平台多而分散,服务平台与科研工作环境贴合不紧密,这些因素阻碍了科研文献情报服务的提升发展。本文根据科研人员期望,探索实现一种嵌入科研环境文献情报服务平台,拉近文献情报服务与科研用户的距离。[方法/过程] 本文梳理分散的文献情报中心服务,使之规范聚合,使用科技网通行证打通各个服务系统身份认证与授权壁垒;在得到科研人员身份的基础上,根据学者研究方向提供可订制的精准文献情报资源服务;并结合新一代 IT 终端嵌入技术,实现将文献情报服务嵌入到科研学者最常用的科研环境中。[结果/结论] 本文研发嵌入式文献情报聚合服务示范平台,将聚合后的应用服务嵌入院所常用科研服务系统,通过用户服务调研反馈的方式,证明新服务模式得到的科研人员的肯定。最后分析该服务模式的问题和未来期望。

关键词: 嵌入科研环境 资源服务聚合 精准服务

分类号: G251

引言

随着大数据时代的到来,资源繁多、平台服务分散成为文献情报服务面对的主要问题。基于检索行为的传统文献情报服务模式,逐渐不能满足当下科研人员对于服务方便性、准确性、及时性的需要。在庞大的科研资源下,系统间的壁垒与资源分散都成为科研人员享受文献情报服务的障碍,用户查找资源和文献已经成为科研工作的负担。

随着中科院各种优秀的文献情报资源和服务应用价值不断得到肯定,科研用户也提出了更高的要求。如:统一的认证授权;根据研究方向智能揭示文献情报信息;自主订制内容和推荐科研信息动态;将服务嵌入常用的科研平台环境等。本文在已有平台建设工作的基础上,结合用户调研,总结归纳用户的需求和当前资源服务平台矛盾,使用嵌入技术手段,探索并实现一种可嵌入科研知识环境的文献情报服务平台,以满足新知识服务时代用户的新需求^[1]。

1. 现有问题分析

本文先期在中科院十个研究所的文献情报调研工作中,总结归纳科研一线用户在文献情报服务上反馈意见,并根据已有资源服务平台的建设工作经验,对文献情报当前服务存在的问题总结为以下几个方面:

- (1) 服务系统多而分散:资源和应用分布分散在各个平台,资源服务没有有效的梳理聚合,用户需要分别访问各自不同的站点获取资源服务,服务和资源没有聚焦给用户使用带来不便。

*本文系中国科学院文献情报能力建设专项“基于统一认证的文献情报服务流整合与定制”(项目编号:科1533)研究成果之一。作者简介:谢靖(ORCID:0000-0001-6698-1786),馆员, xiej@mail.las.ac.cn; 吴振新(ORCID:0000-0003-4966-1961),研究馆员,项目负责人, wuzx@mail.las.ac.cn;

- (2) 多次重复认证：由于系统平台的分散，各平台都有单独权限控制系统，造成了用户需要分别登录认证操作，系统之间资源也不能够关联互通。使得资源服务形成孤岛，难以聚合联通。
- (3) 知识化服务诉求无法满足：传统检索服务提供全领域的数据搜索，用户需要被动的去发现知识。而当前科研学者更希望系统能够根据关心自己研究领域，让知识组织好主动“推”给自己。

2. 目标设计

根据当前科研用户对文献情报服务最迫切的需求，本文对嵌入式服务平台的目标设计归纳为：资源服务梳理聚合、精准化资源揭示服务、嵌入科研服务环境三个主要方面。

2.1. 资源服务的梳理聚合

不同资源服务平台拥有各自的认证授权规则，资源服务聚合首先需要统一身份认证系统，打破资源认证授权的壁垒。实现用户访问各个平台服务时可以无障碍的连接跳转。

文献情报资源服务内容，如文献检索、文献传递、期刊浏览、科技新闻资讯等较多优秀的服务有各自的特色资源和相应的服务职能，对资源服务系统的梳理聚合需要涉及资源梳理和服务梳理两个部分。

(1) 资源的梳理聚合

分布在不同机构、地域有同类型的数据，由于覆盖学科和资源体量有各自的不同。需要对相同类型的资源进行汇集、整合、对比去重。文献情报中心集成发现平台已经完成对资源的汇集工作，本文研究工作建立在此基础之上，不在对这方面工作叙述。

(2) 服务的梳理聚合

服务内容分散在不同的平台，服务质量和使用频率各有差别。首先需要通过定量统计方式，梳理筛选科研用户使用量最高的服务功能，将优质服务的遴选排序。其次使用技术手段，将服务抽取聚合，以 SAAS 的服务模式灵活的揭示服务组件。

2.2. 面向服务对象的精准揭示

资源服务聚合后，在众多的服务中给指定的用户提供其研究兴趣相关的资源和服务是提升嵌入式服务能力核心和关键因素。

(1) 用户画像：对科研用户信息的收集是精准服务的必要环节，从用户身份的基本信息，从发表文献、参加会议、申请专利等维度可以获取科研学者的丰富信息，从鉴别科研用户关注的研究方向、研究时间、资历等信息。从而粗粒度掌握用户的基本需求^[7]。

(2) 精准揭示：科研用户有自己独立的偏好特性，除了根据用户的基本信息给用户个性化资源揭示之外，根据用户主观意愿定制相应资源服务的功能也是必要补充。如定制关注的资源库、期刊群、合作的人员、相关会议、关键词语等。从而让服务系统根据用户的个人定制智能的筛选匹配，向每个用户揭示自

*本文系中国科学院文献情报能力建设专项“基于统一认证的文献情报服务流整合与定制”（项目编号：科 1533）研究成果之一。作者简介：谢靖（ORCID：0000-0001-6698-1786），馆员，xiej@mail.las.ac.cn；吴振新（ORCID：0000-0003-4966-1961），研究馆员，项目负责人，wuzx@mail.las.ac.cn；

己更感兴趣的资源服务。

2.3. 嵌入科研服务环境

资源服务分散在各个地方的网站平台或 APP，使用不便是服务的关键掣肘。科研人员希望服务可以 DIY 的融合嵌入到科研最常用的系统平台上。能否满足服务嵌入，让科研可以随时、随地方便的应用，是新的要求和挑战^[5]。

(1) 多种设备嵌入：科研人员曾提出，满足多设备环境的嵌入兼容是嵌入式服务的必要条件^[2]。即可以随意将服务嵌入到手机、Pad、网站、应用程序中。当前 HTML5/CSS3 等 Web 技术框架的逐渐成熟，为满足多型号设备、多操作系统、多尺寸屏幕的兼容打下坚实基础。

(2) 多种嵌入模式：根据嵌入的深度和服务模式的不同，将嵌入服务分为链接嵌入、页面嵌入、数据嵌入三种方式。链接嵌入服务是通过链接的方式嵌入到第三方平台的页面中；页面嵌入是服务可以作为一个独立展示的子页面镶嵌入第三方页面，尽量保持与原有页面风格一致，并自动适应页面的大小和字体，不破坏原有网站页面风格；数据嵌入是将资源服务封装为 JSON 或 XML 等 API 数据接口，让第三方应用程序可以更灵活的调用数据服务，自定义的展现方式。

3. 已有平台与技术基础

(1) 服务系统资源积累与技术标准化

文献情报中心已有的电子期刊、电子图书、论文发布、机构知识库、iAuthor、机构动态监测、原文传递服务等十多个主要的文献情报应用服务，并已经成熟运行多年，不但拥有庞大优质的资源，而且也拥有一定量的科研用户群体，形成了稳定的系统服务和规范的数据流程。

在技术架构上不断跟随国际最新技术标准，模块化结构清晰普遍具有标准数据接口，可以通过资源服务梳理形成能够支持第三方服务的中间件。良好的资源服务平台，清晰的技术架构为文献情报多个资源服务梳理聚合创造了有利的基础支持。

(2) 统一认证的成熟与广泛应用

中国科学院已经建成的中国科技网通行证，使用科研工作者邮件账户为唯一 ID，在近 100 多个研究所得以广泛应用，拥有最广大的科研用户群体。拥有通行证账号的用户可以直接认证登录任何使用科技网通行证的资源服务平台，极大地方便了各种应用服务集成。本文选用中国科技网通行证作为统一认证工具，实现文献情报聚合和应用嵌入，有助于在科学院和科学界统一身份认证标准，帮助全院用户更好地实现文献情报资源服务与其他外界服务的连接，为将来更多的资源服务聚合嵌入奠定基础。

4. 平台的架构设计

设计总体目标是建成一个有效的示范平台，将选取的 10 种成熟优质的文献情报资源服务梳理聚合，实现嵌入科研环境应用示范应用。平台框架设计主要包括 4 个层面：(1) 接口模块层：将各种优秀的资源服务系统接口梳理；(2) 中间

*本文系中国科学院文献情报能力建设专项“基于统一认证的文献情报服务流整合与定制”（项目编号：科 1533）研究成果之一。作者简介：谢靖（ORCID：0000-0001-6698-1786），馆员，xiej@mail.las.ac.cn；吴振新（ORCID：0000-0003-4966-1961），研究馆员，项目负责人，wuzx@mail.las.ac.cn；

件模块层：将接口开发成为可嵌入的可订制的中间件群，实现服务应用的灵活组合集成；（3）个性揭示层：为科研一线用户提供科技文献的资源的个性揭示与订制服务；（4）嵌入服务层：提供多种嵌入技术支撑，最终形成可以嵌入用户科研环境的知识服务能力。平台的总体建设功能设计架构如图 1 所示。



图 1 总体设计框架图

根据平台框架设计，整体建设流程与思路包括 5 个步骤：

（1）资源服务梳理集成与接口规范处理：统计并筛选使用最频繁的优质文献情报服务，将服务从已有资源服务系统梳理剥离，对服务系统中的功能剥离、改写、合并，形成可以满足服务的资源服务接口，并对服务接口进行统一技术标准处理。同时对数据资源格式进行标准化处理，统一数据交换格式。

（2）资源服务接口组织集成，开发服务中间件：将服务接口转换为可以灵活调用的中间件，每个中间件代表一种不可拆分的独立资源服务。并对资源服务中间件分类，包括：服务类型，适用学科领域、适用场景、适用科研对象、使用权限等。构建中间件管理系统，有效地支撑服务订制和嵌入科研环境的创新系统应用。

（3）基于科技网通行证认证框架，统一身份认证：在各服务系统添加“科技网通行证”认证代理模块，通过 OAuth 协议，实现委托科技网通行证认证平台验证的用户身份的功能。利用 Email 用户唯一身份标示，各服务系统根据用户身份和角色，映射在不同系统内的用户身份，从而给用户分配资源服务的使用权限。最终达到一次认证登录，即可使用多个平台授权服务与授权资源。

（4）精准揭示与订制服务：根据通行证提供的用户身份信息，结合用户的相关文献信息计算识别用户的兴趣，自动将用户研究方向相关的期刊、会议、科技资讯等内容重新组织呈现给用户。并将期刊、会议、兴趣标签等作为开放可订制内容，提供个性订制页面，便于修改勾选所需要的资源。系统根据用户修改剔除研究方实时揭示资源内容。

（5）嵌入科研环境：基于 HTML5 等 Web 框架技术和数据管道技术，提供多

*本文系中国科学院文献情报能力建设专项“基于统一认证的文献情报服务流整合与定制”（项目编号：科 1533）研究成果之一。作者简介：谢靖（ORCID：0000-0001-6698-1786），馆员，xie.j@mail.las.ac.cn；吴振新（ORCID：0000-0003-4966-1961），研究馆员，项目负责人，wuzx@mail.las.ac.cn；

环境、多设备的嵌入方式，提供多种嵌入服务方法，让科研用户可以将文献情报资源服务自由的在邮件、科研 APP、课题组讨论组、专题网站、博客等环境中使用。

5. 平台功能实现与关键技术

(1) 资源服务梳理聚合的实现

平台已经实现电子期刊、电子图书、全文数据库、科技论文预发布、机构知识库、iAuthor、知识产权集成、科技资讯、科研机构监测、图书借阅、原文传递服务、参考咨询、群组知识平台等 13 个主要的服务梳理，将优秀的服务开发成为可以独立呈现的 APPs，集成在文献情报聚合服务平台。如图 2 所示，每个文献情报服务以明确的标题和一句简短的功能描述呈现，汇集后以九宫格的方式呈现给用户，方便用户随时使用或选择嵌入相应的文献情报服务功能。



图 2 资源服务梳理组织及效果展示图

(2) 统一认证与资源权限映射技术

科技网通行证认证框架使用 OAuth 认证协议方式，实现与本地及院内资源服务系统对接。资源服务系统将原有身份认证过程跳转到通行证认证平台。用户在科技网通行证平台输入账号密码后，科技网通行证平台验证的用户身份，并将用户认证后的 Token 和身份信息通过系统间设定的安全数据通道，传送回到资源服务平台，因此原有资源服务平台能够获取用户认证状态和认证信息，并根据用户身份信息给予分配相应的资源服务使用授权。

由于科技网通行证使用科学院 Email 系统进行身份唯一标示，对原有资源服务平台已有的注册用户使用自己独立的 ID 进行唯一身份标示，如何绑定和映射科技网通行证账号和原有用户，实现身份认证的无缝对接成为解决这方面的关键技术。系统采用两种技术方法处理权限映射问题：(1) 对院内或已经绑定 Email 原始账户，根据 Email 地址进行账号匹配映射，将 Email 地址相同的科技网通行证用户赋予原先账户相同的权限。(2) 对于新注册科技网通行证的用户，或使用院外未验证 Email 地址的用户，需要通过原有账户绑定验证的操作。即首次使用科技网访问资源服务系统，用户需要验证原有系统账户，通过验证后，可以将原有注册账号绑定到科技网通行证账号下，当用户再次访问时便可以自动映射到已绑定的账户，从而实现无缝对接。

当前已完成文献情报主要资源服务的科技网通行证的嵌入，通过通行证账号登录任何一个资源服务平台后，当用户再次访问其他资源服务平台，通过安全数

*本文系中国科学院文献情报能力建设专项“基于统一认证的文献情报服务流整合与定制”（项目编号：科 1533）研究成果之一。作者简介：谢靖 (ORCID: 0000-0001-6698-1786)，馆员，xie.j@mail.las.ac.cn；吴振新 (ORCID: 0000-0003-4966-1961)，研究馆员，项目负责人，wuzx@mail.las.ac.cn；

据通道该平台从科技网通行证自动获取到用户当前登录状态和身份，实现跳转自动登录认证，从而实现一次认证登录，多个平台互通的功能，打破多平台间用户身份认证的壁垒。

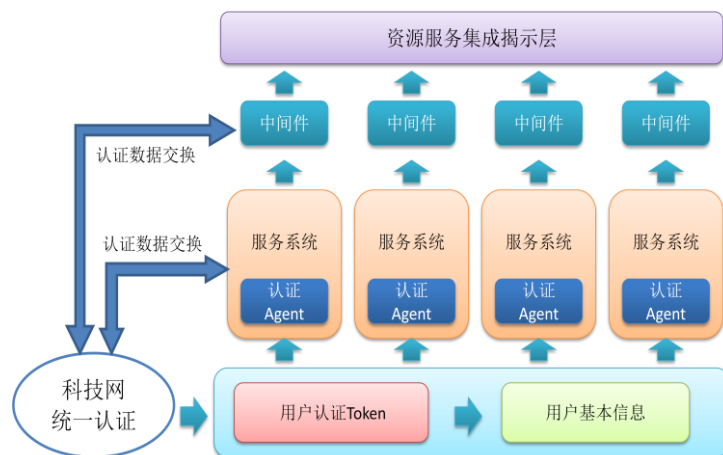


图3 基于院统一认证的资源服务梳理组织流程图

（3）精准揭示与资源订制技术

平台从统一认证系统获取登录用户的基本信息、所在机构以及用户发表的文献等信息，综合判断用户基本研究方向，计算并筛选出满足该用户兴趣的内容，将精确的期刊、电子书、会议论文等科技文献情报信息展示给用户。推荐该学科重要的数据资源、领域研究热点、相关研究机构最新科研动态等内容。结合用户历史访问数据、访问量和影响力因子等综合计算进行 Rank 排序，通过更权威的数据，有效的帮助用户从海量资源中找到自己研究方向最贴切的信息，增强用户使用体验^[4]。

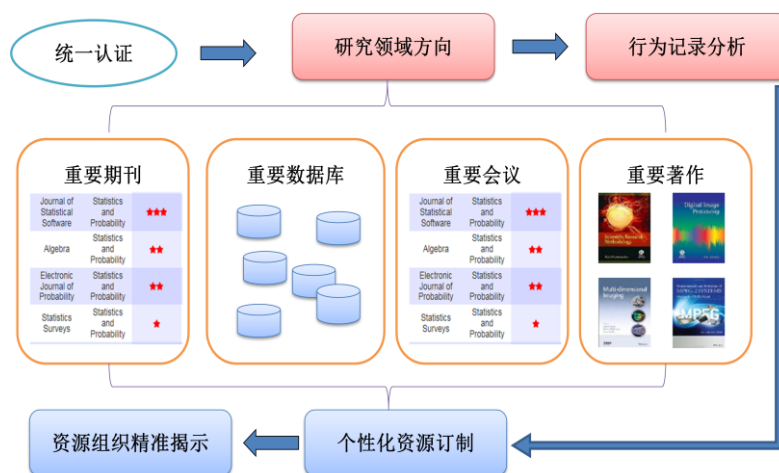


图4 个性化资源精准揭示

平台提供个性订制功能以满足每个用户的个体需求，在现有默认揭示的基础上，用户可以根据兴趣修改和订制个人关注的关注主题、期刊、图书、会议、数据集、科研机构动态消息等。在订制修改后，平台记录用户订制兴趣，从而智

*本文系中国科学院文献情报能力建设专项“基于统一认证的文献情报服务流整合与定制”（项目编号：科1533）研究成果之一。作者简介：谢靖（ORCID：0000-0001-6698-1786），馆员，xiejj@mail.las.ac.cn；吴振新（ORCID：0000-0003-4966-1961），研究馆员，项目负责人，wuzx@mail.las.ac.cn；

能修改计算需要揭示的内容, 实现满足个人研究兴趣的信息展示^[7]。

此外, 平台从行为日志自动计算用户的兴趣^[8], 反馈调整优化个人兴趣模型, 增强精准文献情报服务, 从而减轻用户获取有效文献情报资料的负担。

(4) 嵌入科研知识环境服务的实现

为方便用户在常用的科研环境中使用文献情报服务, 嵌入式知识服务平台设计了三种文献情报嵌入服务包括: 链接嵌入, 页面嵌入, 数据 API 嵌入^[5]。实现在自己的平台上无缝访问多个文献情报服务。

链接嵌入: 实现将文献情报服务封装成为一个单个小应用, 展示应用服务的名称、图表及功能描述, 从而以链接的方式嵌入到具体的科研服务系统中。

页面嵌入: 使用 HTML5 响应式页面的方式, 让页面可以根据具体放入的科研网站或 APP 的位置自适应的调整宽度大小、颜色风格, 从而实现笔记本电脑、pad、手机等多屏幕的自适应。

数据 API 嵌入: 提供 API 接口给科研用户最常用的第三方服务平台, 由平台技术人员调用文献情报服务的 API, 从中获取最新的数据, 并自定义的将数据放在第三方科研平台需要的位置。平台仅作为数据的提供方, 真正的应用可以由第三方科研环境平台灵活设计。

平台分别以链接方式嵌入院邮件系统, 科研在线; 以数据 API 方式嵌入 ARP3.0; 以页面方式嵌入国科大与研究所群组平台示范应用。

6. 结论与评价

嵌入式文献情报聚合服务平台实现了主要的文献情报服务产品嵌入到科研用户环境, 给科研人员提供精准化的文献情报信息嵌入揭示服务。在中国科学院声学研究所、电子学研究所、北京基因组研究所、心理研究所、物理研究所。进行所内平台试用、调查问卷的方式进行调研。根据调查问卷统计, 平台服务受到调研用户的普遍欢迎, 其中 88.8% 的用户愿意嵌入邮件系统, 33.3% 的用户选择嵌入所主页, 22.2% 的用户选择嵌入课题组主页, 33.3% 的用户选择嵌入个人学术网站。调研期间也反馈一些未来需要完善的内容。总结评价为:

(1) 用户期待汇集更多资源服务应用, 如科技查新、专利检索等优秀的应用。需要扩大服务的范围, 在多个研究学科更多掌握用户的嵌入需求, 常用应用可以融合组合嵌入到相应的环境, 提供更多更丰富的嵌入式文献情报服务。

(2) 根据研究所机构研究方向数据组织和精准揭示的方式, 让科研用户看到研究方向匹配的精炼的信息, 得到用户的普遍支持和欢迎。个性数据订制需要将数据进行更精确细粒度的分类和排序, 如期刊的二级学科、排名区间等, 便于用户快速查找到定位符合需求的订制内容, 提高精准服务能力, 提高数据的应用价值。基于用户行为日志的分析的精准推送, 目前由于日志积累较少推荐的效果还有待于进一步改进和提高。

(3) 多种灵活嵌入方式已经在已有服务环境试点应用, 提高用户使用体验效果。链接嵌入和 API 数据嵌入方式较为稳定。使用 HTML5 的方式响应式页面, 由于原始平台并不是完全良好支持 HTML5 框架, 使得用户深入到 2-3 级功能页面后会出现嵌入显示兼容问题。因此还需要与资源服务的实际应用平台同步升级为 HTML5 技术框架才能更好的展示^[9]。

*本文系中国科学院文献情报能力建设专项“基于统一认证的文献情报服务流整合与定制”(项目编号: 科 1533) 研究成果之一。作者简介: 谢靖 (ORCID: 0000-0001-6698-1786), 馆员, xiej@mail.las.ac.cn; 吴振新 (ORCID: 0000-0003-4966-1961), 研究馆员, 项目负责人, wuzx@mail.las.ac.cn;

参考文献:

- [1] 初景利. 嵌入式图书馆服务的理论突破[J]. 大学图书馆学报, 2013, 06
- [2] 李文文, 陈雅. 图书馆嵌入式服务模式研究[J]. 2011, 29(1):90-92
- [3] User Authentication with OAuth 2.0 [EB/OL]. [2015-10].
<https://oauth.net/articles/authentication/>.
- [4] 张玉霞. 改进的个性化智能文献推送方法在数字图书馆中的应用研究 [J]. 情报理论与实践, 2012, 35 (7): 92-95.
- [5] 彭文虎. 面向用户需求的高校图书馆嵌入式服务模式研究. 图书馆学刊, 2015(5):66-69
- [6] 刘凯, 王伟军, 黄英辉等. 个性化推荐系统理论探索:从系统向用户为中心的演进. 情报理论与实践, 2016, 39(3):52-56
- [7] 洪亮, 钱晨, 樊星. 移动数字图书馆资源的情境感知个性化推荐方法研究. 现代图书情报技术, 2016(z1):110-119
- [8] 王智圣, 李琪, 汪静等. 基于隐式用户反馈数据流的实时个性化推荐. 计算机学报, 2016(1):52-64
- [9] 弗里曼. HTML5 权威指南[M]. 北京: 人民邮电出版社, 2014

作者贡献说明: 谢靖: 嵌入式服务模式设计, 平台框架设计及示范系统研发工作; 吴振新: 项目负责人, 整体目标设计, 项目和团队协调管理; 张智雄: 嵌入式知识服务的思想与建议; 胡吉颖: 示范平台技术研发; 许丽媛, 精准服务数据提供及用户调研; 周强: 数据的汇集加工。并感谢科技网技术团队的技术支持。

*本文系中国科学院文献情报能力建设专项“基于统一认证的文献情报服务流整合与定制”(项目编号: 科 1533)研究成果之一。作者简介: 谢靖 (ORCID: 0000-0001-6698-1786), 馆员, xiej@mail.las.ac.cn; 吴振新 (ORCID: 0000-0003-4966-1961), 研究馆员, 项目负责人, wuzx@mail.las.ac.cn;

The Exploration and Practice of the Aggregation Embedded Service Platform of Literature and Information

Xie Jing, Wu Zhenxin, Zhang Zhixiong, Hu Jiying, Xu Liyuan, Zhou Qiang
National Science Library, CAS Beijing 100190

Abstract: [Purpose/Significance] The traditional document information resources service platform is scattered. And the service platform is not closely linked with the scientific research work environment. All these factors hinder the development of the scientific research literature information service. In this paper, according to the scientific research personnel, we hope to develop a kind of embedded information service platform of scientific research environment, and to narrow the distance between the literature information service and scientific research. [Method/Process] These dispersed Documentation and Information services been sorted and regulated. With science and technology network passport, users can get through each service system for identity authentication and authorization to break the barriers. Scientists can be provided with customizable accurate document information resources service according to the research field based on research personnel identity. With a new generation of embedded IT terminal technology, realize the literature information the service is embedded into the most commonly used scientific research environment. [Result/Conclusion] The demonstration platform embedded aggregated information service into scientific research environment. Through the feedback of user service research, it proved that the new service mode of scientific research is succeeding. Finally, this paper analyzes the problems and future expectations of the service model.

Keywords: Embedded Service in Research Environment, Precision Service, Resource Service Aggregation

*本文系中国科学院文献情报能力建设专项“基于统一认证的文献情报服务流整合与定制”（项目编号：科 1533）研究成果之一。作者简介：谢靖（ORCID：0000-0001-6698-1786），馆员，xiejj@mail.las.ac.cn；吴振新（ORCID：0000-0003-4966-1961），研究馆员，项目负责人，wuzx@mail.las.ac.cn；