

从整合到聚合：国内数字资源再组织模式的变革^{*}

张云中

(上海大学图书情报档案系, 上海 200444)

摘要: 针对数字资源再组织问题, 对比经典的“数字资源整合”模式和新兴的“数字资源聚合”模式之间在背景、目标、理念、内容和方法等方面的异同, 剖析数字资源再组织模式从整合到聚合的变革路径, 并指出了数字资源再组织模式发展的精细化、体系化、可视化和立体化趋势。

关键词: 数字资源整合; 数字资源聚合; 知识组织; 发展趋势

中图分类号: G250.7

DOI: 10.3772/j.issn.1673—2286.2014.06.003

进入21世纪后, 数字信息已然成为信息资源的主流, 数字资源的开发利用已成为时代赋予的重大课题。众所周知, 数字资源组织是数字资源开发利用的前提, 然而事实上数字资源组织往往呈现出局部有序而整体无序的窘境, 具体体现为数字资源孤岛、数字资源超载等, 这大大阻碍了用户对数字资源的有效获取和共建共享。为摆脱这种窘境, 学者们都将解决该问题的途径聚焦在了数字资源的再组织上。关于数字资源的再组织, 国内学者在长期的探索实践中依次形成了“资源整合”和“资源聚合”两种模式, 本文旨在对国内数字资源再组织的两种模式进行深入的对比, 梳理出国内数字资源再组织模式的变革路径, 进而展望数字资源再组织模式发展的趋势。

1 国内数字资源再组织研究进展

1.1 数字资源整合研究进展

国内数字资源整合的相关研究开展较早, 学者马文峰于2002年最初提出数字资源整合的概念, 认为数字资源整合是数字资源优化组合的存在状态, 是把各个相对独立的数字资源结合为一个新的有机整体^[1]。随后, 数字资源整合的相关研究便如火如荼地展开。较早

时期的数字资源整合研究主要围绕数字资源整合的作用^[2]、数字资源整合的方式^[3]、数字资源整合的方法与技术^[4]、数字资源整合的实现机制^[5]等理论探讨展开。随着理论研究的深入, 图书馆数字资源整合^[6]、企业数字资源整合^[7]、医学数字资源整合^[8]等应用研究也不断被尝试。近三年来, 数字资源整合的研究又呈现出新技术、多领域的新特点, 基于本体的数字资源整合^[9]、基于Web3.0的数字资源整合^[10]、基于SFX的数字资源整合^[11]、基于关联数据的数字资源整合^[12]等成为学者们破解数字资源整合技术难题的新着眼点, 同时数字资源整合也将呈现出多领域趋势^[13]。

1.2 数字资源聚合研究进展

国内数字资源聚合研究开展稍晚一些, 其概念是由贺德方研究员在数字资源整合的基础上提出。数字资源聚合^[14]旨在发现资源之间及其内在的语义关联, 从而构建一个内容相互关联、多维度、多层次的资源体系, 形成集概念主题、学科内容和科研对象实体为一体的立体化知识网络。国内数字资源聚合的研究主要从两个思路展开, 一是从增强语义出发, 利用本体来表示或利用关联数据增强数字资源语义, 包括基于本体的数字资源语义聚合与可视化^[15,16]、基于关联数据的数字资源

^{*} 本研究得到国家自然科学基金项目“语义网络环境下数字图书馆资源多维度聚合与可视化展示研究”(编号: 71273111)、教育部人文社会科学青年基金项目“社会化标注系统资源多维度聚合与导航研究”(编号: 13YJC870032)和上海高校青年教师培养资助计划项目“概念格理论在情报分析中的应用”资助。

聚合^[17]等；另一种思路是从发现关联出发，利用计量分析来挖掘数字资源之间的引证关系，包括基于耦合关系实现数字资源聚合^[18]、基于共现与耦合实现数字资源聚合^[19]等。

2 数字资源整合与数字资源聚合的异同

数字资源整合研究方兴未艾，而数字资源聚合研究犹如雨后春笋，两个概念之间究竟差异何在、是否一脉相承已成为数字资源再组织领域不得不辩的新问题。本文认为，数字资源整合与数字资源聚合是密切相关的两个概念，两者之间既有共同的地方，也存在着明显的差异。

2.1 背景异同

从概念提出的背景来看，数字资源整合和数字资源聚合都是针对数字资源无序而又难以获取利用的困境提出的。前者着重关注异构数字资源之间的协同，解决数字资源存储系统的异构、资源类型的异构、资源语法的异构和资源语义的异构等问题，实现异构资源之间的互通与互操作。对比而言，后者的提出不仅契合于消除数字资源异构的大背景，还着重强调利用多种途径、多种方法和多个界面来对数字资源语义及资源关联进行深层次的揭示和展示，进而更好地满足用户对数字资源有效获取和共建共享的需求。

2.2 目标异同

数字资源整合目标是消除数字资源之间存在的各类异构，将相对分散的数字资源进行规范、科学地重组，并将它们集中整合到一个新的有机的资源体系之中，增强资源之间的有序程度，形成一个效能更好、效率更高的数字资源体系。对比而言，数字资源聚合的目标强调构建一个资源内容相互关联、多维度、多层次的资源体系，形成集概念主题、学科内容和科研对象实体为一体的立体化知识网络。后者的目标较之前者更为具体化，更具针对性，强调构建知识化、可视化的数字资源体系。

2.3 理念异同

数字资源整合一般采用自顶向下的“分类”的知识

再组织理念，根据事先设计好的数字资源整合体系结构对异构数字资源进行再组织，通过建立异构数字资源之间互通与映射的相关协议、映射规则或中介结构，将无序异构的数字资源重新纳入到新的统一的资源体系之中。相比而言，数字资源聚合则采用自底向上的“聚类”的知识再组织理念，无需事先设定资源整合体系结构，而是根据资源聚合的需求，选择合适的资源聚合维度及模型，对所选资源进行自底向上的聚类操作来生成数字资源的层次结构体系。

2.4 内容异同

数字资源整合涵盖的内容较广，一般意义上至少涵盖三方面：基于数据的数字资源整合（强调数字资源逻辑与物理上的有机集中）、基于信息的数字资源整合（强调资源链接和一步到位获取）和基于知识的数字资源整合（强调整合资源知识结构），以上三方面的内容互有交叉且逐次递进。相对而言，数字资源聚合更接近于“基于知识的数字资源整合”的范畴，其强调通过概念分析、引证关系等手段发现并用可视化手段呈现出数字资源中蕴含的知识结构，并不涉及数据整合与信息整合的范畴。

2.5 方法异同

数字资源整合已经形成了较为完善而成熟的方法体系，包括哲学方法、系统科学方法等指导性方法和知识组织方法、内容分析法等操作性方法，这些方法的综合使用推进了数字资源整合理论的蓬勃发展。相比之下，数字资源聚合尚未形成完备的方法体系，但也探寻出了两类行之有效的操作性方法：知识组织方法和计量分析方法，前者主要运用概念分析、本体、关联数据等方法增强资源语义，后者主要运用共现关系、耦合关系、社会网络分析等方法挖掘资源关联。

3 从整合到聚合：数字资源再组织模式的变革

通过对数字资源整合与数字资源聚合的异同分析，我们不难看出，数字资源聚合的概念是从数字资源整合的母体中孕育出的具有自己特定背景、目标、理念、内容和方法的新概念。更进一步地说，数字资源聚

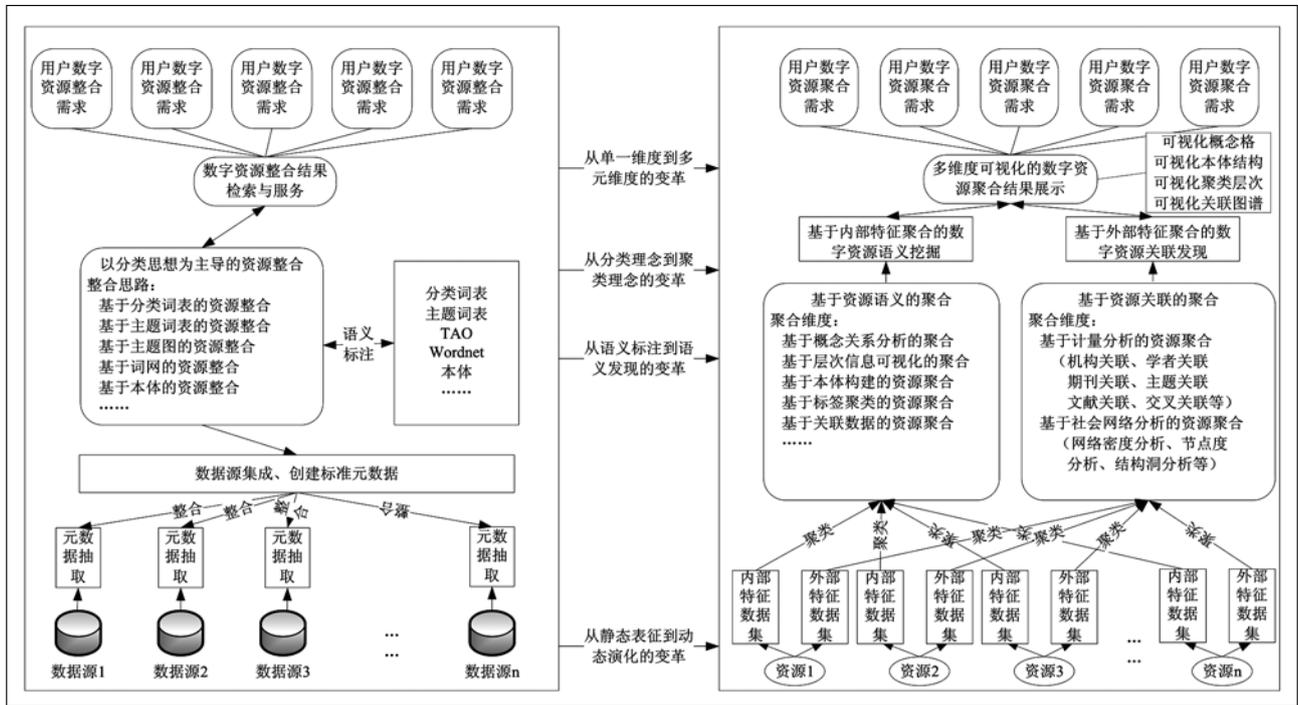


图1 数字资源再组织模式的变革

合概念的产生，意味着数字资源再组织模式已经发生了从整合到聚合的新变革。本文认为，数字资源再组织模式的变革如图1所示。

3.1 从分类理念到聚类理念的变革

从分类理念到聚类理念的变革，标志着数字资源再组织“思路”的重大转折。分类理念是指在数字资源整合中，先为待整合的数据源事先设定一个能反映其知识结构的分类体系，这种体系可以是分类词表、主题词表、主题图、本体乃至词网中的任何一种，进而根据从各个数据源中抽取和整合后的标准资源元数据将数字资源映射与集成到相应的分类体系之下，实现数字资源体系的有机统一。聚类理念是指在数字资源聚合中，并无设定的知识结构分类体系，而是从资源的内部特征数据集和外部特征数据集出发，利用资源聚合理论使得数字资源呈现出集簇状，同簇资源彼此相似和关联，不同簇资源尽量不相似不关联，进而根据数字资源呈现出的层次簇状结构为之选择一种或多种合适的表示结构，包括可视化概念格、可视化本体结构、可视化聚类层次和可视化关联图谱等。从分类理念到聚类理念的变革，使得数字资源新体系中知识结构的类和类间关系不再固定，而是根据现实因需设之，变得更为灵活。

3.2 从语义标注到语义发现的变革

从语义标注到语义发现的变革，标志着数字资源再组织“精度”的不断提升。数字资源整合中，我们能做的仅仅是根据设定的新的资源组织体系的特点，对数字资源进行相应的语义标注。在专家分类体系结构下，一般用类目等级关系进行资源语义标注；在主题法的体系结构下，一般用属分关系进行资源语义标注；在本体结构下，一般用概念关系、概念与属性关系、概念与实例关系进行资源语义标注；在主题图中，一般用主题、关联和显现三要素来进行资源语义标注，其他语义标注形式不再赘述。在数字资源聚合下，我们要做的则是利用各种形式的聚类模型和聚类技术，深入地挖掘出隐含在数字资源之间的语义关系，这些语义关系并非直接存在和轻易捕获的，而是必须通过语义发现过程才可呈现出的。在发现资源内容语义时，通常用到的方法包括形式概念分析、层次信息可视化、标签聚类等；在发现资源关联语义时，通常用到的方法包括计量分析（尤其是共现分析和耦合分析）和社会网络分析等。从语义标注到语义发现的变革使得数字资源新体系不仅能完成一般意义上的资源语义标注，更能从知识发现的角度挖掘出隐含的资源语义，大大提高了数字资源再组织的精度。

3.3 从静态表征到动态演化的变革

从静态表征到动态演化的变革，标志着数字资源再组织“柔性”的不断提高。数字资源整合中，用于整合结果表征的新资源组织体系选定后，才可进行与之相匹配的资源语义标注，才可实现数据源的集成和标准元数据的创建。自顶向下的理念决定了新的资源组织体系一旦建立就不易修改的特点，换言之，数字资源整合的资源体系呈现出相对静止的态势，加入新的数据源或者更换数字资源再组织的方式（如从专家分类法体系转换为本体结构）对数字资源整合往往意味着一次从头到尾的新的再组织过程。而在数字资源聚合中，资源经过自底向上的聚合形成一定的聚类结构后，才会根据呈现出的聚类结构特点选择相应的资源组织体系，这不仅推迟了选择新资源组织体系的时间，更重要的是使得数字资源再组织的方式的选择与变换呈现出一定柔性。另一方面，数字资源聚合以单个资源为基本操作单位，增加或删除待聚类资源的操作愈发频繁，自底向上的资源聚合模式可以更好地支持这种不断变化的外部环境，表现出更好的适应性。从静态表征到动态演化的变革，使得数字资源新体系能够不断根据外部环境的需求对数字资源在组织体系方面作出相应的动态调整，大大提高了数字资源再组织的应变性。

3.4 从单一维度到多元维度的变革

从单一维度到多元维度的变革，标志着数字资源再组织“视角”的不断拓展。数字资源整合中，整合的视角往往都关注于单一的维度，即选用某一种资源组织体系标注不同数据源中资源间的语义关系。而在数字资源聚合中，同样是对数字资源间语义关系的表示，采用的聚合维度却呈现出多元化的形态。数字资源聚合的维度大致可分为两类，一类是基于内部特征聚合的数字资源语义挖掘，包括概念关系分析、本体构建等多种途径；另一类是基于外部特征聚合的数字资源关联发现，是数字资源整合所不具有的维度，包括机构关联发现、学者关联发现、期刊关联发现、主题关联发现、文献关联发现和交叉关联发现等多种关联。从单一维度到多元维度的变革，使得数字资源新体系能够根据用户不同的资源再组织需求设定出与之相适应的维度，拓宽了传统意义上资源再组织的视角，更易满足用户的多样化需求。

4 数字资源再组织模式发展的趋势

通过对数字资源再组织模式变革的剖析，结合当前数字资源整合和数字资源聚合研究的热点，本文认为，数字资源再组织模式发展的趋势将涵盖以下四个方面：精细化、体系化、可视化和立体化。

4.1 精细化趋势

精细化趋势是指数字资源再组织模式发展中，用语义关系揭示的概念颗粒度越来越精细。无论是数字资源整合还是数字资源聚合，两种模式最终都会采用知识组织体系来完成对数字资源的再组织，而各类知识组织体系都是通过概念实现不同的知识组织目的，只是其中的概念颗粒度有所差异。近年来运用本体来表示数字资源再组织结果的研究越来越多，其原因也在于本体的概念颗粒度较小，较之分类词、主题词、主题图等更适用于精确语义关系的揭示和语义推理。

4.2 体系化趋势

体系化趋势是指数字资源再组织模式发展中，资源再组织理论呈现出愈发完备和成熟的理论体系。数字资源整合与数字资源聚合在背景、目标、理念、内容和方法等方面虽有差异，却一脉相承，后者是在前者基础上提出的为满足新环境下数字资源再组织需求的新理论，二者统一于数字资源再组织的理论框架之下，而且，随着后续研究的不断深入，该理论体系还会被不断丰富和完善。

4.3 可视化趋势

可视化趋势是指数字资源再组织模式发展中，资源再组织的结果甚至过程逐渐都可以利用可视化方法和技术进行展示。可视化的技术不仅可以帮助用户快速和有效理解数字资源再组织结果中的语义信息，更能通过可视化分析发现隐含的语义关系进而揭示资源再组织的深层次内涵，提高结果的可理解性和可认知性，数字资源聚合中的可视化概念格、可视化本体结构、可视化关联图谱都是最好的例证。

4.4 立体化趋势

立体化趋势是指数字资源再组织模式发展中,用于刻画数字资源再组织的维度越发多元,呈现出立体化的知识网络。发展至今,数字资源再组织的维度已不限于从资源内部特征出发标注和发现数字资源语义关系一个维度,而是拓展到从内部特征和外部特征两个大维度及其包含的若干小维度共同刻画数字资源语义关系,使得数字资源体系呈现出丰满、立体的语义形态。

结语

数字资源再组织是解决数字资源孤岛、数字资源超载等问题,促进数字资源有效获取和共建共享的有效途径。本文通过对国内学者在长期的探索实践中逐渐形成的“数字资源整合”和“数字资源聚合”两种模式的异同,探寻了从整合到聚合的数字资源再组织模式变革路径,并指出了数字资源再组织模式发展的趋势。

参考文献

- [1] 马文峰.数字资源整合研究[J].中国图书馆学报,2002(4):63-66.
- [2] 李希明,土丽艳,金科.从信息孤岛的形成谈数字资源整合的作用[J].图书馆论坛,2003(6):121-122,61.
- [3] 黄晓斌,夏明春.数字资源整合方式的比较与选择[J].情报科学,2005(5):690-695.
- [4] 马文峰,杜小勇.论数字资源整合的方法与技术基础[J].中国图书馆学报,2005(6):49-53.
- [5] 李广建,汪语宇,张丽.数字资源整合的实现机制及关键技术:对国外数字资源整合系统的实证研究[J].中国图书馆学报,2007(2):75-80.
- [6] 黄传慧,李娟.我国图书馆数字资源整合研究[J].图书与情报,2009(4):66-69,82.
- [7] 谷俊,严明.企业数字资源整合系统的设计与实现[J].情报杂志,2010(5):183-187.
- [8] 白海燕.医学数字资源整合技术的研究与应用[J].中华医学图书情报杂志,2008(1):56-58.
- [9] 马小军,李广建.基于本体的数字资源整合方法与技术[J].情报科学,2010(10):1541-1546.
- [10] 胡翠红.基于Web3.0的数字资源整合研究[J].图书馆学研究,2012(6):62-65,87.
- [11] 田子德,毕强,王雨.SFX在数字资源整合中的应用[J].情报科学,2012(12):1858-1861.
- [12] 郑燃,唐义,戴艳清.基于关联数据的图书馆、档案馆和博物馆数字资源整合研究[J].图书与情报,2012(1):71-76.
- [13] 肖希明,唐义.国外多领域数字资源整合研究进展[J].中国图书馆学报,2013(4):26-35.
- [14] 贺德方,曾建勋.基于语义的馆藏资源深度聚合研究[J].中国图书馆学报,2012(4):79-87.
- [15] 何超,张玉峰.基于本体的馆藏数字资源语义聚合与可视化研究[J].情报理论与实践,2013(10):73-76,39.
- [16] 李劲,程秀峰,宋红文.基于语义的馆藏资源深度聚合方法研究[J].情报科学,2013(11):100-103.
- [17] 游毅,成全.试论基于关联数据的馆藏资源聚合模式[J].情报理论与实践,2013(1):109-114.
- [18] 杜晖.基于耦合关系的学术信息资源深度聚合研究[D].武汉大学,2013.
- [19] 邱均平,王菲菲.基于共现与耦合的馆藏文献资源深度聚合研究探析[J].中国图书馆学报,2013(3):25-33.

作者简介

张云中,男,1985年生,博士,上海大学图书情报档案系讲师,研究方向:知识组织。E-mail: zhang-yun-zhong@126.com。

From Integration to Aggregation: The Change of Digital Resources Re-organization Pattern in China

ZHANG YunZhong

(Department of Library, Information and Archives, Shanghai University, Shanghai 200444, China)

Abstract: For the problems of digital resources re-organization, the classic "digital resource integration" pattern are compared with the emerging "digital resource aggregation" pattern from several aspects, such as the background, the objectives, the ideas, the contents and the methods, then the change path of digital resources reorganization pattern from integration to aggregation are analyzed, and finally the developing trend of digital resources re-organization pattern is pointed out that it will be more meticulous, systematic, visual and three-dimensional.

Keywords: Integration of digital resource; Aggregation of digital resources; Knowledge organization; Developing trends

(收稿日期: 2014-04-30)