

doi: 10. 19388/j. zgdzdc. 2018. 05. 13

引用格式: 王圣洁. 我国海洋地质知识体系 DIKW 模型及其规模估计与增长研究[J]. 中国地质调查, 2018, 5(5): 98 - 103.

我国海洋地质知识体系 DIKW 模型及其规模估计与增长研究

王圣洁

(中国地质调查局青岛海洋地质研究所, 青岛 266071)

摘要: 我国的海洋地质工作起步较晚, 缺少对学科发展的整体评价。采用 DIKW (Data - Information - Knowledge - Wisdom, DIKW) 概念模型, 阐述了我国海洋地质知识体系的构成及关联领域, 归纳并提出了可用于统一描述海洋地质知识体系的实例模型; 应用信息计量学方法, 对我国海洋地质知识体系的规模进行了初步估计, 原始数据超过 10 万件, 研究论文超过 1 万篇; 依据案例研究的结果, 分析了 1979—2017 年我国海洋地质数据 - 信息 - 知识的同步增长情况, 在经历 15 a (1980—1995 年) 的缓慢增长阶段后, 我国海洋地质工作进入了一个由基础研究向专业调查转变的快速增长时期。

关键词: 海洋地质; 知识体系; DIKW 模型; 规模估计; 增长模式

中图分类号: P736 **文献标志码:** A **文章编号:** 2095 - 8706(2018)05 - 0098 - 06

0 引言

海洋地质学是一门年轻的学科, 以 20 世纪 60—70 年代的“地学革命”达到了辉煌的顶点, 成为整个地质学变革的龙头。而我国正处在“文化大革命”的历史灾难之中, 我国科学家与这场“地学革命”失之交臂, 使我国本来起步很晚的海洋地质学与世界的差距更大^[1-2]。因此, 学界内外都有这样的思考: 经过改革开放后 40 a 的发展, 我国的海洋地质学现在处在什么样的水平?

杨子庚^[1]系统总结了我国海洋地质学在 20 世纪后期的起步和兴起: 70 年代相继开展的渤海—黄海—东海陆架地质 - 地球物理综合调查、南海地质 - 地球物理综合调查等工作, 获得了我国海洋地质最初的基本资料 and 认识; 80—90 年代, 我国在近海实施了一系列专项调查, 并在国际地学合作中开始参与大洋和极地科考。王圣洁等^[3]应用信息计量学方法, 对我国 2000—2012 年发表的论文进行了统计分析, 认为我国海洋地质研究的重点仍集中

在近海海域, 研究内容主要为盆地构造与海洋沉积等基础研究、天然气水合物等新能源及近岸环境与灾害地质等应用领域。莫杰等^[4]对 2008 年以来最近 10 a 的海洋地质调查工作进行了综述: 我国近海海洋综合调查与评价 (908 专项)、1: 100 万管辖海域基础地质调查等项目的相继实施, 获得了全覆盖、高精度的调查资料; 重点海域油气资源调查确定了潜在勘探区块, 南海天然气水合物试采获得实质性突破。经过几代人的奋斗, 我国海洋地质工作迎来了一个蓬勃发展的时期^[5]。

40 a 的海洋地质工作积累了大量的调查资料、研究成果和学术论文, 有必要从学科的整体视角进行一次系统的梳理和总结。本文引入知识体系 DIKW (Data - Information - Knowledge - Wisdom, DIKW) 概念模型, 分析了我国海洋地质工作形成的数据、信息和知识层次, 结合大数据“4V (Volume, Velocity, Variety, Value)”特征, 重点探讨了我国海洋地质知识体系的规模和增长模式, 这不但是学科研究的一次有益尝试, 对于促进海洋地质信息科学在管理中的实际应用也具有现实意义。

收稿日期: 2018 - 03 - 19; 修订日期: 2018 - 07 - 11。

基金项目: 中国地质调查局地质矿产调查评价专项“地质资料委托管理系统建设 (编号: 12120115072601)”和中国地质调查局海洋基础性公益性地质调查专项“海洋地质数据库建设 (编号: 121201005000150010)”联合资助。

作者简介: 王方数据 (64—), 男, 研究员, 主要从事海洋地质信息分析与研究。Email: wshengjie0532@163.com。

1 我国海洋地质知识体系实例模型

1.1 知识体系概述及 DIKW 概念模型

知识体系又称知识体系金字塔,其概念涉及“数据”“信息”“知识”“智慧”等关键词,这些词汇最早可追溯到西班牙诗人 T. S. Eliot 的一首诗(The Rock, 1934),其大意是:“生命迷失在生活里,智慧迷失在知识里,知识迷失在信息里,信息迷失在数据里”。现在一般认为,美国 Russell L. Ackoff 是较早提出知识体系的学者,在其“From Data To Wisdom”的论文中,给出了数据、信息、知识、智慧的定义及其关系(图1)^[6]。后来的学者在此概念模型的基础上有了进一步的论述,使得数据、信息、知识、智慧的内涵更加明确:①“数据”是反映客观事物及其性质和状态的记录,是人的感官或通过仪器进行观测获得的文本、数字、事实或图像,数据是最原始的记录,在孤立和分散状态时,通常并不具备特定意义;②“信息”来自数据之间的联系,从而可以更好地反映“因果关系”,回答“who(谁)”“what(什么)”“where(哪里)”“when(何时)”等问题;③“知识”是经过筛选的信息,并具有明确的目的和作用,可以回答“how(怎样)”“why(为什么)”这样的高阶问题;④“智慧”是人类所具有的一种特有能能力,是对知识的深入理解和灵活运用,以及对事物发展的前瞻性看法。

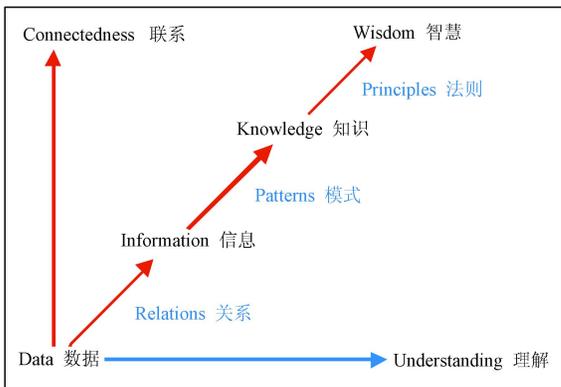


图1 DIKW 概念模型^[6]

Fig. 1 DIKW conceptual model^[6]

知识体系概念的提出,虽然有助于人们理解从数据到智慧的层次结构和相互关系,但在实际应用中仍遇到不少困难:①数据、信息、知识之间的界限并不像定义中那样分明,一个过程的“信息”可能

是另一个过程的“数据”,有些学者试图给出判别准则,但仍偏重哲学思考的范畴^[7];②对于每个具体学科,专业研究者更加关注细分领域的获取与分析,知识体系的整体探讨较为少见,以本文讨论的海洋地质学科为例,由于其仅仅只有70a的发展历史,尚缺少对其知识体系的探讨和论述。

1.2 实例模型的引入及其层次特征

知识体系概念模型包括4个层次,这里重点选取其中的数据、信息和知识3个层次,以建立关于海洋地质知识体系的实例模型。对于模型中的每个层次,从计量单位、数据描述、资源汇聚体系和碎片化与移动服务4个方面给出其主要特征(表1),结合当前业务工作的实际,指出与海洋地质相关的一些特点。

表1 海洋地质知识体系实例模型
Tab. 1 Instance model of marine geological knowledge hierarchy

层次	地质资料	地质数据库	知识库
知识层			期刊论文总库 (卷-期-篇; 目录;大数据; CNKI,实时;碎片化与移动应用:是)
信息层	成果地质资料 (档-案卷-件;目录;大数据;汇交;碎片化与移动应用:否)	成果数据库 (包-数据集-记录;元数据;大数据:部分、滞后;碎片化与移动应用:是)	
数据层	原始地质资料 (档-案卷-件;目录;大数据;归档;碎片化与移动应用:否)	原始数据库 (包-数据集-记录;元数据;大数据:部分、滞后;碎片化与移动应用:是)	

注:地质资料、地质数据库、知识库包含内容类型有计量单位、数据描述、大数据汇聚体系和碎片化与移动服务。

我国海洋地质知识体系的实例模型涵盖了3个传统的业务领域,分别为地质资料档案管理、地质数据库建设和知识库管理。一般认为,资料、数据和知识的管理与传统上的图书、情报及档案3个学科密不可分^[8],但随着信息技术的进步,3个领域的运行模式各自发生了很大的变化,并对知识体系的形成和管理产生了不同的影响。

(1)地质资料档案管理体系是海洋地质知识体系资源汇聚的基础。新中国建立以来,我国就形成了国家-地方两级地质资料管理体系且一直延续至今^[9],并在地质资料档案资源的汇聚过程中发挥

了保障作用。目前,已有的部省两级地质资料馆藏机构包括全国地质资料馆、国土资源实物地质资料中心和31个省(区、市)地质资料馆,馆藏地质资料47.55万种^[10-11]。此外,随着油气、海洋等地质资料委托保管机制在2012年的建立和运行,行业地质资料也逐步纳入国家地质资料管理体系^[12]。值得注意的是,我国地质资料档案管理虽然属于一个独立的学科且自行发展,但有关原始地质资料和成果地质资料的划分与定义,与知识体系中的“数据层”“信息层”对应良好。

国务院《地质资料管理条例》中规定,地质资料是指在地质工作中形成的文字、图表、声像、电磁介质等形式的原始地质资料、成果地质资料和岩矿心、各类标本、光薄片、样品等实物地质资料。《地质资料管理条例实施细则》则进一步明确了原始地质资料和成果地质资料的范围,将原始地质资料细分为8个专业合计35类实际资料,将成果地质资料细分为10个专业合计58类地质图件和报告。

随着地质资料工作的重点由档案管理向信息服务转移,一些传统管理的局限性开始凸显:一是档案“档-案卷-件”的计量单位较大,这种“数据-信息粒度”不适合进行重组和分析;二是档案的保存和使用多数采用纸质媒介,不利于信息碎片化处理及在移动应用中使用。此外,我国海洋地质资料采取了委托保管方式,原始地质资料在各个海洋地质调查机构分别归档保存,全国性的大数据汇聚机制尚不完善。

(2)地质数据库建设推动了海洋地质知识体系向数字化形态发展。我国基础地质数据库建设是过去近20a的一项重要工作。自1998年开始,中国地质调查局组织开展了各省(市、自治区)地质数据库试点建设,截至2015年底,相继完成了1:250万、1:50万、1:25万、1:20万数字地质图控件数据库,特别是已经完成约4600幅1:5万地质图空间数据库,形成了海量的中大比例尺的区域地质图空间数据,包括地质图、矿产图、第四纪地质图或基岩图以及有关的地质报告、说明书、专题报告等,涵盖了全国近60a积累的地质成果^[13]。我国陆域地质数据库建设还包括全国重要矿种矿产地数据库、全国区域地球化学数据库、1:20万全国重砂数据库等专业内容,但这些数据库均为成果地质资料数字化所建,原始地质资料并未入库。

我国区域海洋地质数据库建设则是始于2002

年,分为区域海洋地质调查资料数据库和区域海洋地质成果数据库,两者同时建设,因此能够与原始地质资料和成果地质资料一一对应。区域海洋地质调查资料数据库包括了地球物理调查资料(站位与测线资料、海洋重力、海洋磁力、地震资料和浅层剖面调查资料等)、海洋水深地貌调查资料(水深、多波束和旁侧声纳调查资料)、区域地质调查资料(地层资料、表层沉积物样品、浅柱状样、钻孔资料等以及相关样品的测试数据),区域海洋地质成果数据库则主要包括重力异常系列图、磁力异常系列图、水深图、地貌图、环境系列图、灾害系列图、构造图、地质图、盆地分布及矿产分布等地质图件^[14],可以满足集中管理、重复使用的实际需求。

此外,区域海洋地质数据库以归档调查资料和成果报告为素材,其建设过程滞后于海洋地质调查的实际进度,加之海洋地质资料委托保管的原因,原始调查资料分散保存在各个调查机构,难以构建全国性的海洋地质大数据。2017年以来,随着“地质云1.0”的建成和上线运行,海洋地质数据共享机制将得到一定程度的改善,海洋地质数据库建设、更新和服务会逐渐走向规范^[15-16],其固有的“细粒度”“碎片化”“实时性”的优势也会逐步发挥作用。

(3)中文文献数据库是我国海洋地质知识体系的主体组成。1998年,世界银行提出了国家知识基础设施(National Knowledge Infrastructure, NKI)的概念,旨在建立一个可共享、可操作的多学科知识群;此后,我国三大中文学术期刊数据库相继创立并得到了快速的发展^[17]。以中国知网为例,在不到20a的时间,CNKI《中国学术期刊(网络版)》已经发展成为世界上最大的连续动态更新的中文学术期刊全文数据库,内容覆盖自然科学、工程技术、农业、哲学、医学、人文社会科学等多个领域,收录国内学术期刊约8000种,全文文献总量超过5000万篇。

CNKI知识库具备大数据应用的所有特征。这类知识库不仅期刊论文收录完整,而且具有近乎实时的汇聚能力。2017年,部分期刊论文已经实现了“网络首发”,提高了期刊出版、发行的时效。海洋地质是知识库的学科内容之一,因此可以通过检索和大数据分析开展相关知识体系的研究。

2013年,王圣洁等^[3]对我国海洋地质论文进行了计量分析,利用CNKI知识库对海洋地质论文进行

了检索和统计,从载文的作者分布、学科分布、资助来源、研究层次和研究内容等5个方面进行了深入讨论,揭示了海洋地质研究在我国的最新现状以及存在的问题,讨论了海洋地质学科的发展方向。

2 我国海洋地质知识体系的规模估计与增长模式

2.1 海洋地质知识体系的规模估计

由于CNKI等知识库的存在和便利使用,了解海洋地质“知识层”的规模相对简单,但由于目前的检索方式仍不能准确得到需要的检索结果,“数据清洗”成为其中最烦琐的环节。2013年,王圣洁等^[3]对我国海洋地质论文进行了主题词检索,获得了2000—2012年的有效论文数量共162篇。其中,中国地质调查局所属的青岛海洋地质研究所论文数量29篇,广州海洋地质调查局论文数量18篇,国家海洋局两研究机构论文数量合计15篇,中国科学院两研究机构论文数量合计17篇,4所高校论文数量合计40篇,其他研究机构合计43篇。2015年,王圣洁等^[5]又按作者单位对青岛海洋地质研究所发表的所有论文进行了检索,获得了1979—2013年的有效论文数量1804篇,其中2000—2012年的论文数量为1239篇。

对比以上2种检索方式和结果可以看出,使用主题词检索只能检出部分论文,其检出率约为2%。因此,通过检出率和抽样检索获得的论文数量将可以大致估算出我国海洋地质论文的总体规模,1979—2013年形成的海洋地质文献约为10077篇(表2)。

表2 我国海洋地质论文数量和馆藏地质资料的规模
Tab.2 Article counts and sum of geological data in our marine geological field

研究机构	主题词检索量(2000—2012年)/篇	作者单位检索量(1979—2013年)/篇	馆藏地质资料统计量(1979—2013年)/件
青岛海洋地质研究所	29	1804	15109
广州海洋地质调查局	18	1119 ^①	66388
国家海洋局	15	933 ^①	7775 ^②
中国科学院	17	1057 ^①	4404 ^②
高校研究机构	40	2488 ^①	6911 ^②
总计	162	10077 ^①	100587 ^③

注:①根据2%的检出率进行估算;②根据产出率进行估算,其中包括:国家海洋局研究机构设定为12%,中国科学院研究机构设定为24%,高校研究机构设定为36%;③馆藏地质资料总计中未包含其他研究机构。

海洋地质资料长期分散在各个调查研究机构,进行准确的统计较为困难。2012—2016年,中国地质调查局海洋地质资料委托保管系统建设项目得以实施,完成了馆藏海洋地质资料的整理和统计,获得了1979—2013年馆藏地质资料的准确数据,其中,中国地质调查局所属的青岛海洋地质研究所计15109件,广州海洋地质调查局计66388件。仍然采用对论文数量的估计结果,可以得到2个研究机构的资料—论文产出率为12%和2%。

可以看出,不同海洋地质调查研究机构的论文产出与地质资料比率存在很大差异。考虑到中国科学院以及高校等研究机构的产出率一般更高,在给出其适当的经验值后就可以通过论文数量对其可能拥有的原始资料进行估算,计算结果总计超过10万件,其中中国地质调查局所属调查研究机构形成的地质资料约占80%(表2)。

2.2 海洋地质知识体系的增长模式与影响因素

由于目前尚未形成海洋地质知识体系的完整大数据,对其整体的增长研究缺乏基础。选择青岛海洋地质研究所开展了案例研究,尝试通过“解剖麻雀”反映我国海洋地质发展的总体趋势,其获得的结论有一定的参考价值。

从补充统计的数据来看(表3),大致以1995—2000年为衔接,之前的15a里,海洋地质论文与资料增长表现为平缓线性增长模式,不仅年度增长量较少,年度变化也较小;之后的15a里,论文与资料的累计增长转为典型的快速指数增长模式,年度增长量持续大幅度提高,显示我国海洋地质工作处在一个快速发展的阶段,其增长趋势预计仍将持续15~20a的时间^[3,5]。需要注意的是,2010年之后,论文与资料的增长出现了相反的趋势,这可能与海洋地质调查与研究的阶段性调整不无相关。

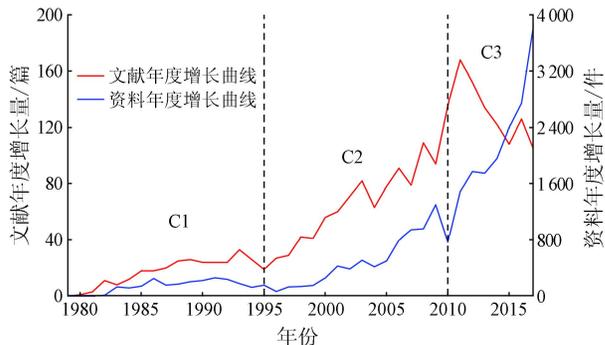
海洋地质论文与资料的阶段增长还与国家财政投入密切相关。2000年前后,国土资源部的成立以及中国地质调查局的重新组建,事实上增加了海洋地质工作的投入,一系列国家海洋地质调查专项的实施推动了海洋地质资料和论文的快速增长^[3-5]。此外,地质调查与研究的关系一直是地质工作部署中的一个重要问题,海洋地质论文和资料的同步增长以及趋势变化都会带给我们有益的启示,当海洋地质论文与资料增长显著相关时,表明地质调查与研究的关系密不可分,当论文数量与资料增长出现阶段差异时,则表明地质调查与研

表3 海洋地质资料与发表论文对比统计^[5]

Tab.3 Comparison between data size and article counts in marine geological field^[5]

年度	资料数量	论文数量	年度	资料数量	论文数量
1979	0	0	1999	151	41
1980	0	1	2000	259	56
1981	0	3	2001	428	60
1982	12	11	2002	386	71
1983	130	8	2003	510	82
1984	115	12	2004	417	63
1985	141	18	2005	502	78
1986	250	18	2006	798	91
1987	156	20	2007	943	79
1988	170	25	2008	955	109
1989	204	26	2009	1 299	94
1990	222	24	2010	764	136
1991	260	24	2011	1 488	168
1992	238	24	2012	1 772	152
1993	178	33	2013	1 749	134
1994	124	26	2014	1 958	122
1995	155	19	2015	2 399	108
1996	65	27	2016	2 745	126
1997	131	29	2017	3 871	104
1998	137	42	合计	26 082	2 264

究在其不同的发展阶段各有侧重：2010年前的青岛海洋地质研究所，其论文增长快于资料增长，具有所谓“研究型”的特征；2010年以来，论文增长慢于资料增长的速度，开始显示出“调查与研究并重”的特点^[5]（图2）。



数据分别来自 CNKI 文献数据库和馆藏海洋地质资料目录数据库

图2 海洋地质论文与资料年度增长曲线对比^[5]

Fig.2 Annual growth curve comparison between the data size and article counts in marine geological field^[5]

3 结论

引入 DIKW 概念模型并结合我国地质资料档案管理、海洋地质数据库建设和知识分析，提出了我国海洋地质知识体系的实例模型，对知识体系的

整体规模进行了估计，对知识体系的增长模式进行了阶段划分，形成以下主要认识：

(1) 我国海洋地质知识体系可将传统地质资料档案管理、现代地质数据库建设和正在兴起的知识管理 3 个领域结合为一个整体，其实例模型的提出有助于在一个更高的理论层次上从整体的视角审视 3 个方面的工作，有助于合理布局相关业务，建立和完善大数据汇聚体系和机制，发挥海洋地质信息科学与技术的支撑作用。

(2) 改革开放 40 a，迎来了我国海洋地质调查研究的快速增长时期，与数据层和信息层对应的原始地质资料和成果地质资料累计达到 10 万件，与知识层对应的科研论文超过 1 万篇。其中，中国地质调查局占有的海洋地质资料超过 80% 以上，以此为基础可以快速构建我国的海洋地质核心大数据。

(3) 海洋地质知识体系和大数据应用的目的还是在于“发现价值”，这方面的研究和实践依旧薄弱：一是需要加强知识体系的基础研究，充分发挥地学统计、数据挖掘等技术优势，促进海洋地球信息科学的创新和突破；二是大力推进数据、信息的开放与共享，努力在社会大数据的多样性中占据关键位置，切实发挥海洋地质调查成果支撑和保障自然资源管理的先导作用。

参考文献：

- [1] 杨子庚. 海洋地质学[M]. 济南: 山东教育出版社, 2004.
- [2] 许清华. 搏击沧海地质革命风云录[M]. 2 版. 北京: 地质出版社, 2006.
- [3] 王圣洁, 周永青, 张兆代, 等. 我国海洋地质论文的科学计量分析[J]. 海洋地质与第四纪地质, 2012, 32(6): 94, 112, 128.
- [4] 莫杰, 王文海, 彭娜娜, 等. 我国海洋地质调查研究新进展[J]. 中国地质调查, 2017, 4(4): 1-8.
- [5] 王圣洁, 周永青, 张兆代, 等. 我国海洋地质文献及地质资料增长规律的案例研究与对比[J]. 海洋地质与第四纪地质, 2015, 35(6): 189-195.
- [6] Ackoff R L. From data to wisdom[J]. Journal of Applied Systems Analysis, 1989, 16: 3-9.
- [7] 梁战平, 张新民. 区分数据、信息和知识的质疑理论[J]. 图书情报工作, 2003(11): 32-35, 104.
- [8] 张莉娅, 邓勇. 图书情报档案三学科交叉融合研究——从交叉著者和知识网络的角度[J]. 图书馆杂志, 2014, 33(12): 62-68.
- [9] 廖寿郎. 全国地质档案资料管理工作三十年[J]. 档案工作, 1985(9): 7-8.
- [10] 王黔驹, 颜世强, 王永生, 等. 全国地质资料馆藏机构现状、问题与对策建议[J]. 资源产业经济, 2011(1): 18-21, 31, 54.

- [11] 国土资源部办公厅. 2015 年度全国地质资料管理与服务情况通报[EB]. 2016.
- [12] 王圣洁, 张兆代, 王黔驹, 等. 我国地质资料研究论文的增长与学科发展[J]. 中国矿业, 2016, 25(11): 76-82.
- [13] 赵林林, 刘荣梅, 邓勇. 区域地质数据库建设进展[J]. 地质论评, 2017, 63(增刊1): 389-390.
- [14] 魏合龙, 戴勤奋, 林峰, 等. 区域海洋地质数据库建设纲要[J]. 海洋地质与第四纪地质, 2004, 24(1): 139-142.
- [15] 魏合龙, 孙纪红, 苏国辉, 等. 数字海洋地质工程建设进展[J]. 海洋地质前沿, 2018, 34(3): 1-7.
- [16] 谭永杰. 地质大数据体系建设的总体框架研究[J]. 中国地质调查, 2016, 3(3): 1-6.
- [17] 谭捷, 张李义, 饶立君. 中文学术期刊数据库的比较研究[J]. 图书情报知识, 2010(4): 4-13.

Study on DIKW instance model and size of the marine geological knowledge hierarchy in China

WANG Shengjie

(Qingdao Institute of Marine Geology, China Geological Survey, Qingdao 266071, China)

Abstract: In China, the marine geological survey has started late and lacked an overall evaluation for the discipline development. Utilizing DIKW (Data - Information - Knowledge - Wisdom, DIKW) conceptual model, the author expounded the composition and associated fields of the marine geological knowledge hierarchy in China, and put forward an example model for describing the system. Through the informetrics method, the author evaluated preliminarily the size of the marine geological knowledge system, and the results show that there are the 100,000 raw data and more than 10,000 research articles in the system. According to the study results of this example, the author analyzed the simultaneous growth of the marine geological data - information - knowledge from 1979 to 2017. After 15 years' (1980 to 1995) sluggish growth, the marine geological work in china has entered a rapid growth period of the transformation from fundamental research to professional investigation.

Key words: marine geology; knowledge hierarchy; DIKW model; volume evaluation; growth mode

(责任编辑: 刘永权)