

记录海洋科学的航行足迹： SIO Explorer——美国斯克里普斯海洋学研究所数字图书馆

刘燕权¹, 高颖²

(1.美国南康涅狄格州立大学, 纽黑文市 06515, 美国; 2.中国社会科学院哲学所, 北京 100732)

摘要: SIO Explorer海洋科学数字图书馆主要提供海洋科学数据资料的搜索服务。它集合海洋地理学、海洋地质学等海洋科学学科的研究资料, 为海洋科学领域的研究、发展提供专业信息支持。其馆藏资源丰富、数据收集全面, 其馆内数字资源的元数据描述对数字图书馆的发展有推动作用。文章对该馆的概况、现状和意义进行综合评述, 并作出评价和建议。

关键词: SIO Explorer; 数字图书馆; 海洋科学

中图分类号: G250.7

DOI: 10.3772/j.issn.1673—2286.2014.06.011

1 概述

美国斯克里普斯海洋学研究所数字图书馆(SIO Explorer)隶属于美国斯克里普斯海洋学研究所(Scripps Institution of Oceanography, 简称SIO), 是一个集合海洋地理学、海洋地质学等海洋科学数据资料的研究型数字图书馆。网址是<http://siox.sdsc.edu>(图1所示为SIO Explorer主页)。

在海洋学领域的研究中, 从SIO等组织的科学考察过程中收集到的第一手信息数据起着重要作用, 因此, 将这些数据、特别是那些几十年前存放于磁带等载体中的信息数据进行转化, 并用元数据描述、存档,

对以后的研究是有重要意义的。SIO Explorer项目始建于2001年, 由SIO、美国加利福尼亚大学圣迭戈分校(University of California at San Diego, 简称UCSD)图书馆和美国圣迭戈超级计算机中心(the San Diego Supercomputer Center, 简称SDSC)三个组织合作共建, 项目的目标是: 数字化历年来收集到的海洋学相关数据资料并进行保存^[1]。尽管SIO Explorer的工作人员并不把它看成一个数字图书馆, 而是认为SIO Explorer是一个“联合搜索引擎”, 但根据它所包含的丰富馆藏和所提供的服务, 我们认为冠以数字图书馆是名副其实的。

SIO Explorer馆藏大致分为五个资料库^[2], 大部分数据都直接来自于SIO所内历年科考收集的数据资料; 直至2005年, SIO Explorer与伍兹霍尔海洋研究所(Woods Hole Oceanographic Institution, WHOI)合作, 将WHOI的科考数据资料也收录馆藏。并且, 作为美国国家科学数字图书馆项目计划的一部分, SIO Explorer在将来终将并入美国国家科学数字图书馆这一庞大项目中, 成为其中海洋科学图书馆的重要组成部分。

自始建至今, SIO Explorer的发展离不开各类组织对其在资金和技术上的支持。网站的资金来源主要有美国国家科学基金会(the National Science Foundation, NSF)、美国国家科学数字图书馆(the National Science Digital Library, NSDL)和SIO三大组织的资助。此外, SIO Explorer的建设又不仅是网站



图1 SIO Explorer主页

的建设,其馆藏建设由许许多多其他组织和人员所资助,其所藏记录资料甚至能回溯至1953年。网站的技术支持来源广泛,该图书馆集合研究人员、计算机工程师和图书馆员等不同专业人员的共同努力,使得SIO Explorer得以实现其馆藏资源的合理揭示和获取^[1]。

2 数字资源及其组织

2.1 资源范围及种类

从2001年起,经过十年的建设,SIO Explorer组织了大量海洋科学相关的数据资料,截至2011年5月,SIO Explorer现有馆藏包括近千条航线所收集的考察数据,并且数据采集流程已流水线化。

SIO Explorer的数字资源主要包括五个数据库:

① SIO航行资料数据库,存储了超过693次出海航行或科学考察的全记录,包括每次航行中能收集和探测到的各个海域和地区的所有数据、图片和文件。

② SIO照片资料库,主要存储斯克里普斯海洋学研究所历次出海航行考察的历史照片。

③ 海山资料库,存储了历年科考中,从大概300条航线里采集到的、与太平洋海底海山相关的地球物理学和地球化学数据。

④ 海洋地质标本库,存储在SIO科学考察中收集到的地质标本,具体包括岩芯、岩石标本。

⑤ 教育资源库,主要是以SIO Explorer馆藏资源为基础制作的相关课程资料,供中学的教师、学生使用和学习。

在这些资料库中,存储的资料多元化,有高分辨率气象观测图、水资源档案、生物和地质标本、地球引力数据、地球磁场数据、海底声纳测绘图、其他文件、图像存档等多种数据类型。有了这些珍贵的数据资料,SIO Explorer每月大概有50,000访问点击量,同时下载流量超过6GB^[2]。

除了现有的五个资料库,SIO Explorer还正在筹备两个资料库,包括:

① 仪器资料库,主要收录航行中数据采集所用设备和仪器、工具等相关的各种文档;

② 深拖系统资料库,主要是深拖系统理论研究相关勘探数据,包括运用声纳系统、照相机、地磁计和其他传感器,在长达38年间,对海底地形勘探的详细数据。

2.2 资源组织

SIO Explorer不仅为海洋学领域的科学探索作出贡献,并且他们的元数据组织方式对其他数字图书馆的资源建设也有一定的启发和影响。从上面的介绍可知,SIO Explorer的资源范围集中在海洋科学,尤其是太平洋海域的勘探数据,大部分都是SIO几十年来,一代代学者和工作人员不断研究、积累得到的宝贵资料。由于原始数据总量巨大、种类非常丰富,在建设初期,有很多困难需要克服,其中最大的困难就是没有现成可利用的数据库,SIO Explorer必须从零开始把现有的非数字化资源转换成数字资源。比如,在转换SIO的约700条航次的勘探数据的过程中,不仅数据存放分散在不同的工作站等其他地点,而且数据载体的形式也多种多样,其中大部分数据的载体形式都不是数字形式,而是纸质文档、磁带和磁盘等传统载体形式。为了把这些40年积累下来的文档、资料转换为可用数据,SIO Explorer作了非常多的尝试和努力。

因此,为了组织如此庞大、种类繁多的数据,SIO Explorer主要采用元数据技术组织其馆藏数据,即对数字化对象进行描述,使数据基本属性能被揭示。在SIO Explorer中,每一个数字化对象被称作ADO(任意单个数字化对象),每个ADO都以数据库主记录和副记录两种元数据的形式存在,主记录保持不变,只对副记录进行操作来适应搜索需要。ADO以*.mif文件格式存储,即元数据互换格式(Metadata Interchange Format)。以这种格式存储,有以下两个主要优点:

① 能使SIO Explorer的资料数据与其他数据库或数字图书馆区别开来,同时能够方便地进行获取和管理。使用这样的元数据组织方式,保证了馆藏数据和数据结构的完整性,为项目今后的发展和数据的兼容提供了良好的基础;

② 每个元数据都是以ASCII格式存储为*.mif文件,每个*.mif文件相应有其元数据标准化格式,比如数据类型、港口、主要科研负责人等描述项目都是受控数据。

3 服务特征

3.1 目标用户

作为SIO的数字图书馆,SIO Explorer为用户提供海洋学及海洋勘探科学研究数据资源。从网站介绍我

们知道, SIO Explorer的目标用户大体分为三类:

① 专业用户。主要包括从事海洋学、海洋勘探学等相关工作的学者、工程师及其他海洋专业相关人员。这些用户使用SIO Explorer的各种数据资源进行研究和等工作。

② 教育用户。主要包括学校教师及学生。这些用户主要利用SIO Explorer的资料辅助教学海洋、地理的相关知识。

③ 普通用户, 即大众。SIO Explorer面向所有人开放其馆藏资源, 任何对海洋学研究感兴趣的用户均可登录网站浏览感兴趣的相关资料和数据, 并下载馆藏资源, 包括从1953年起超过693次航行日志的数据、图像和文件, 还有其他海洋研究相关资源等^[2]。

3.2 服务方式

SIO Explorer为用户提供多种搜索服务, 使其能够查询并获取数据库中存储的数据和元数据。在数据获取页面里, 我们能够发现, 有两种不同的数据搜索接口供用户选择: 一种是网页检索界面, 另一种是基于Java技术的Cruise浏览器。

3.2.1 网页检索界面

网页检索界面是通过网页的方式访问数据库, 类似于普通的搜索引擎界面。如图2所示。

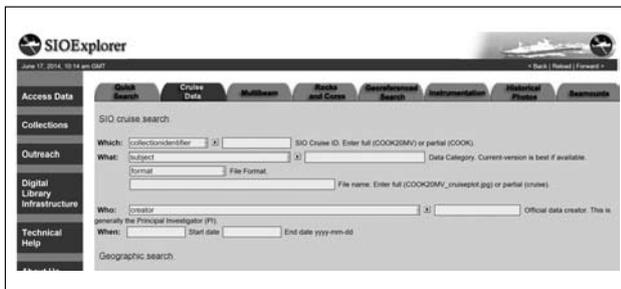


图2 网页检索界面

从SIO Explorer的网页检索界面我们可以看到, 页面上方有八个标签, 除了一般性检索即快速检索之外, 其余有七个资料检索标识库。其中“多波束测探(Multibeam)”标识库, 是“航行资料数据库”的一个分支, 有可能是出于专业检索需求而单列出来; “岩石和岩芯(Rocks and Cores)”标识库对应的是“海洋地

质标本库”。标识库基本与馆藏五大数据库相关对应。

由于各个库所存资料类型不同, 内容千差万别, 因此, 不同于一般的论文、期刊等文本型数据库资料, 对SIO Explorer的资料库分别进行检索能更好地满足检索要求。

在快速检索中, 检索前, 用户可在标识库一栏里选择检索所需资源库, 主题栏里选择需要的数据类型, 即可进行检索。

其他七个检索库的搜索项, 比快速检索有更多检索点可选择。这是因为检索库的资料种类各异, 根据资料数据的特点进行不同特征揭示, 比如: 在航线资料库中, 用户能够通过检索船队名称、勘探设备、科考时间等特定检索点进行检索。值得一提的是, 在航线资料库、岩石和岩芯库、地理参考库三者的检索页面中, 均提供了经纬度精确检索点, 用户只需输入关注区域的精确经纬度, 就能检索出在这一经纬度附近的所有海洋学科考察数据信息。

3.2.2 Cruise浏览器

Cruise浏览器是基于Java技术开发的一个搜索浏览界面, 要求用户终端安装Java Web Start方能使用, 这对于操作系统不是Windows的用户来说在使用上有一定障碍。

Cruise浏览器是在网页检索界面的基础上, 结合Google Earth等地图软件, 使检索结果能更直观地展现给用户, 对用户来说, 这无疑是一种非常方便且直观的检索界面。

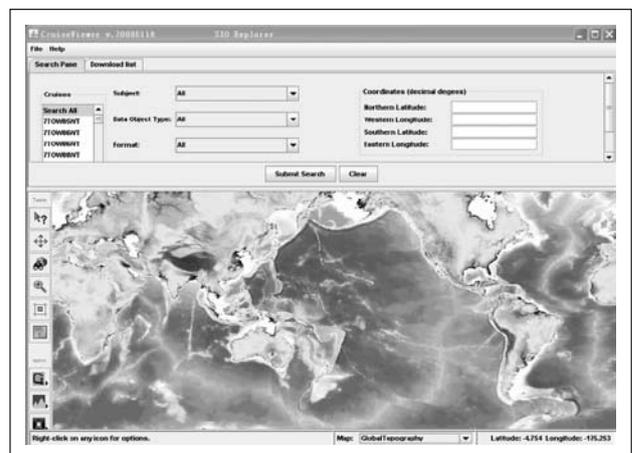


图3 Cruise浏览器

3.3 检索特色

在SIO Explorer的众多检索功能中,有些是其独有的。数据库返回检索结果的同时,还有一些可扩展的附加信息。

① 在安装了Google Earth软件的终端上,用户可以从SIO Explorer下载KML格式文件,在Google Earth地图上显示科考航次的全程路线信息图,直观地将信息呈现给用户;

② 检索结果的显示方式多样,包括显示详细信息、显示简要信息和批量下载等三种方式。如果用户选择批量下载方式,即可一次性批量下载检索航线的日志及其他文件;但批量下载是通过基于Java的DOTS软件来运行的,因此这项功能要求用户在终端设备上安装Java才能够正常使用。

4 技术特征

4.1 技术架构

在文章开头,我们已经说明,SIO Explorer数字图书馆是由SIO、UCSD和SDSC三个组织合作建设,其整体架构也是三者共同搭建。馆藏资源由SIO整理并提供;UCSD负责将资源进行元数据描述和揭示;而SDSC负责数据的存储及网站的运行支持。如图4所示。

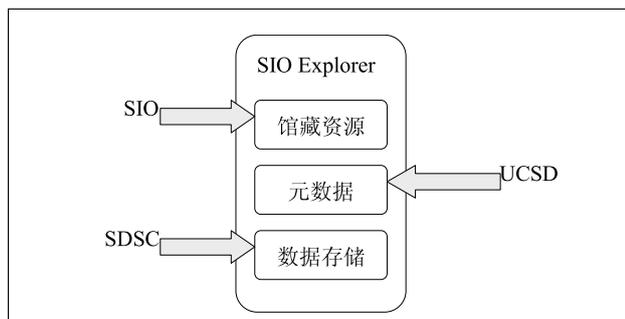


图4 SIO Explorer数字图书馆技术架构

网站大量使用开源软件,主要有两种:W3C Markup Validation version 1.2软件和“批量下载”功能所用的基于Java Web Start的DOTS软件。

SIO Explorer数字图书馆能够从专业的角度解析馆藏资源,挖掘适合研究的检索点,整合三大组织的技术支持,通过网络为海洋学研究及其他用户提供了专业

可靠的科学数据,为科研提供高效辅助。

4.2 界面设计

SIO Explorer海洋科学数字图书馆的界面采用代表海洋的蓝色为主色调,界面简单清晰。主页显示SIO Explorer的主要服务——资料搜索服务入口,其中快速搜索居上,网页搜索接口和Java搜索接口位于其下,从整体来看,三者位于网站主页最醒目位置,用户能够直接点击选择所需搜索入口进行搜索。这样简单的方式,把SIO Explorer最重要的组成部分和提供的主要服务直接呈现给用户,虽简单但实用。

在主页的左侧纵向列表排列着网站的导航栏,包括“数据获取”、“馆藏”、“延伸服务”、“馆内架构”、“技术帮助”以及“关于我们”六个栏目。用户在任意子页面都能由左侧导航栏转链到感兴趣的页面,在各个栏目中,也有相应的文字简要介绍栏目相关信息。

5 评价和建议

SIO Explorer海洋科学数字图书馆馆藏丰富,海洋科学专业方面的资源在质量和数量上都让人惊叹,其特色资源——科考航行数据库尤为珍贵,数据保存完整、年代跨度大并且兼有免费获取的优点,是别的数字图书馆无法企及的。SIO Explorer对专业数据的发掘,启用元数据对资源进行描述,保证了馆藏数据和数据结构的完整性,为项目今后的发展和数据的兼容提供良好的基础,这为其他数字图书馆的建立树立了良好的标杆。同时,对于数据的组织,SIO Explorer对专业数据的揭示也尽量根据其专业特点和要求进行调整,以适应特殊数据的揭示,如对航行记录中航线数据,结合Google Earth等地图软件,直观显示航线,这和现在的地理信息系统(GIS)有异曲同工之处,也体现了SIO Explorer的前瞻性和优秀兼容性。

但是,从用户的角度,SIO Explorer也存在以下几点可以改进的地方:

① 馆藏范围略显狭窄。其馆藏专注于海洋科学,大部分数据库资料都来源于SIO,而SIO的主要研究领域专注于太平洋海域,因此,与之相比,其他海域的数据略显单薄,数据库的内容范围仍显得有些狭窄。

② 资料库名称不统一,搜索帮助不清晰,导致

用户友好度不高。在搜索引擎界面里,除了“快速检索”之外,还有七种资源标识库,而在馆藏资源介绍里,我们只能看到五个数据库。比如“多波束探测(Multibeam)”一项,资源介绍里没有说明,而它实质上是“航行资料数据库中”的一个分支,有可能是出于专业检索需求而单列;而另一个“岩石和岩芯(Rocks and Cores)”也让笔者有些困惑,实际上它就是“海洋地质标本库”。

③ 针对普通大众的服务较薄弱。SIO Explorer搜索服务功能强大,能很好地揭示库中海洋学相关数据及元数据,但是,服务的角度都是从专业用户出发,如果没有经过专业知识训练,需要获取海洋学的一些基础信息的一般用户,使用起来有一定障碍,因此,这对SIO Explorer服务对象的普通民众来说,搜索引擎不太友好。

④ 技术兼容性有待升级。出于对馆藏学科资料的专业性考虑,SIO Explorer的资料,尤其是图片资料必须在Java的基础上才能显示,这对于Windows用户来说,安装Java Web Start不是难事,但对于其他操作系统的用户来说,却存在不兼容的问题。

SIO Explorer海洋科学数字图书馆建立已有十年,它在数字图书馆的发展中不断地探索和创新,为海洋科学的研究提供丰富详实的资料,为环境的保护、人类研究的进程作出了特殊的贡献。

参考文献

- [1] SIO. Geological Data Center at Scripps Institution of Oceanography [EB/OL]. [2011-06-07]. <http://gdc.ucsd.edu/>.
- [2] SIO EXPLORER. SIO Explorer [EB/OL]. [2011-06-07]. <http://siox.sdsc.edu/>.
- [3] MILLER S P, KOPPERS A, HELLY J, et al. SIO Explorer: Digital Library Project [EB/OL]. [2011-06-07]. www.peterbrueggeman.com/cv/oceans2001.pdf.
- [4] MILLER S P. Scripps Explorer [EB/OL]. [2010-02-20]. <http://siox.sdsc.edu/>.
- [5] National Science Foundation. Award Abstract #0911838: Oceanographic Instrumentation for SIO [EB/OL]. [2009-07-24]. <http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0911838>.
- [6] National Science Foundation. Award Abstract #0723024: MRI Acquisition of a Primary Cluster for the SIO COMPAS Shared Computer Facility [EB/OL]. [2007-08-08]. <http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0723024>.
- [7] National Science Foundation. Award Abstract #0824992: Operation of the Clivar and Carbon Hydrographic Data Office at UCSD/SIO, 2009-2013 [EB/OL]. [2008-08-25]. <http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0824992>.

作者简介

刘燕权,男,博士,美国南康涅狄格州立大学终身教授,研究方向:数字图书馆、数据挖掘等。E-mail: liuscsu@gmail.com。
高颖,女,硕士,中国社会科学院哲学所馆员,研究方向:图书馆服务、数字图书馆等。

SIO Explorer——A Digital Library to Oceanography

Yan Quan LIU¹, GAO Ying²

(1. Southern Connecticut State University, New Haven, CT 06515, USA; 2. Institute of Philosophy, Chinese Academy of Social Sciences, Beijing 100732, China)

Abstract: As part of the National Science Digital Library (NSDL), the SIO Explorer provides access to more than fifty years of digital oceanographic cruise data, historical information, links and publications. Utilizing the most recent tools in database and digital library creation, its goal is to take the various forms of data, documents, and images from well over "693 cruise logs" along with various other oceanographic related materials and make them available to researchers, students and the general public who are interested in the study of deep sea exploration. This paper provides an overview of the major development of its digital collection and organization, service provided and interface designed. Authors' comments and suggestions are also given.

Keywords: SIO Explorer; Digital library; Ocean sciences

(收稿日期: 2014-03-31)