

SIOExplorer: IT for Research and Education

Stephen P. Miller, Geological Data Center, SIO
John Helly, San Diego Supercomputer Center
Hubert Staudigel, IGPP, SIO

Deborah Day, UCSD Libraries
Catherine Johnson, IGPP, SIO

University of California, San Diego
9500 Gilman Dr.
La Jolla, California 92093-0220 USA
spmiller@ucsd.edu

Abstract- SIOExplorer (<http://SIOExplorer.ucsd.edu>) is now in operation as a web portal to the discoveries of the Scripps Institution of Oceanography. Data, images and documents from 822 SIO cruises are being placed online, integrated with a growing seamount catalog, global databases, and historic archives. More than just a web site or online archive, SIOExplorer is a fully-searchable collection in the NSF-sponsored National Science Digital Library (NSDL; www.nsdl.org).

was to stabilize the valuable collection of multibeam seafloor swath mapping data, which was at risk from several factors. Much of the data was stored on magnetic tape, and needed migration to new media. Technologies, formats and processing practices have varied significantly since 1981. The SIO collection includes data from 822 cruises since the 1950's, including 281 multibeam cruises, now growing at about 200 GB per year.

I. INTRODUCTION

The SIOExplorer project was announced at the Oceans 2001 meeting in Hawaii as an initiative to improve access to Scripps Institution of Oceanography expedition data [1]. Our first thought

While the initial goal was to make data available for the researcher, it soon became apparent that the efforts could have significant impact on education. Rather than just create an online archive or a website, the decision was made to build a fully searchable

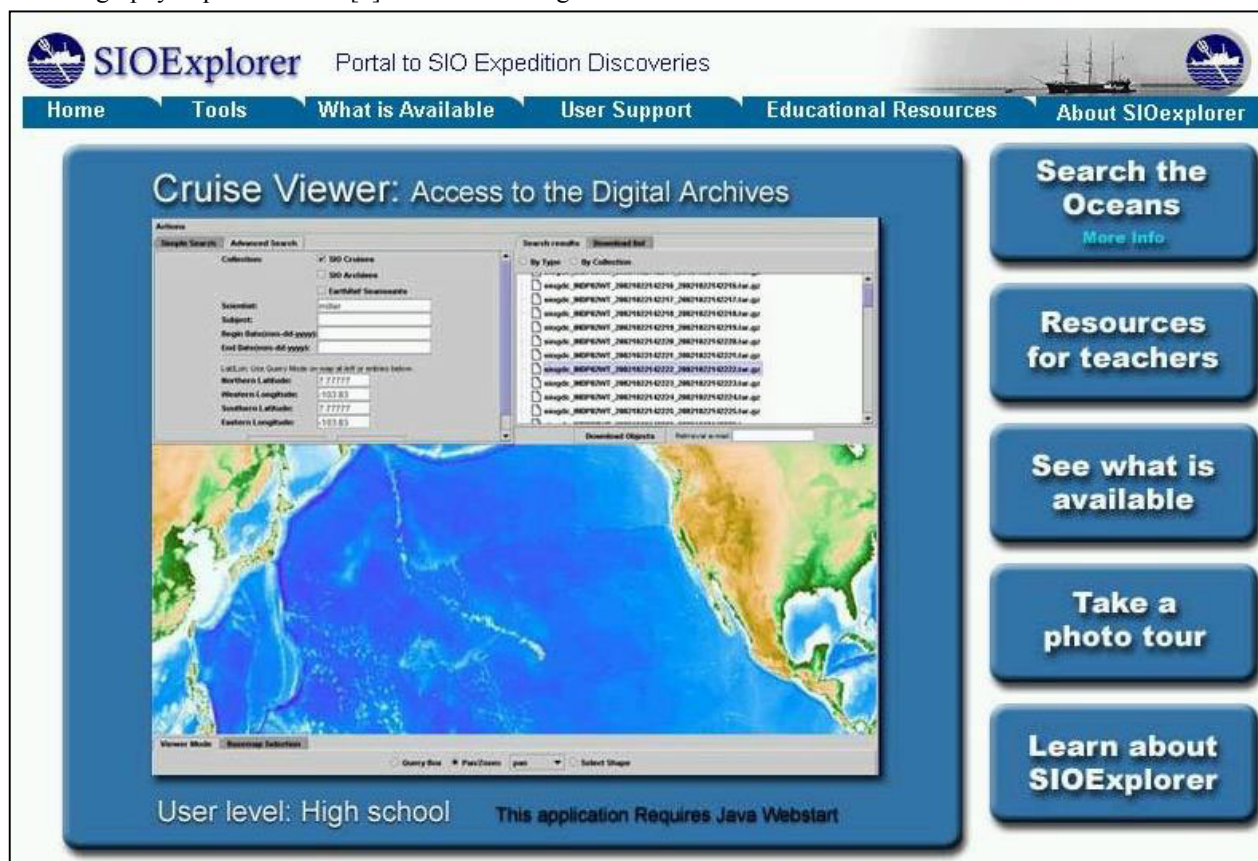


Figure 1. SIOExplorer web site, opening archives for research as well as for education.

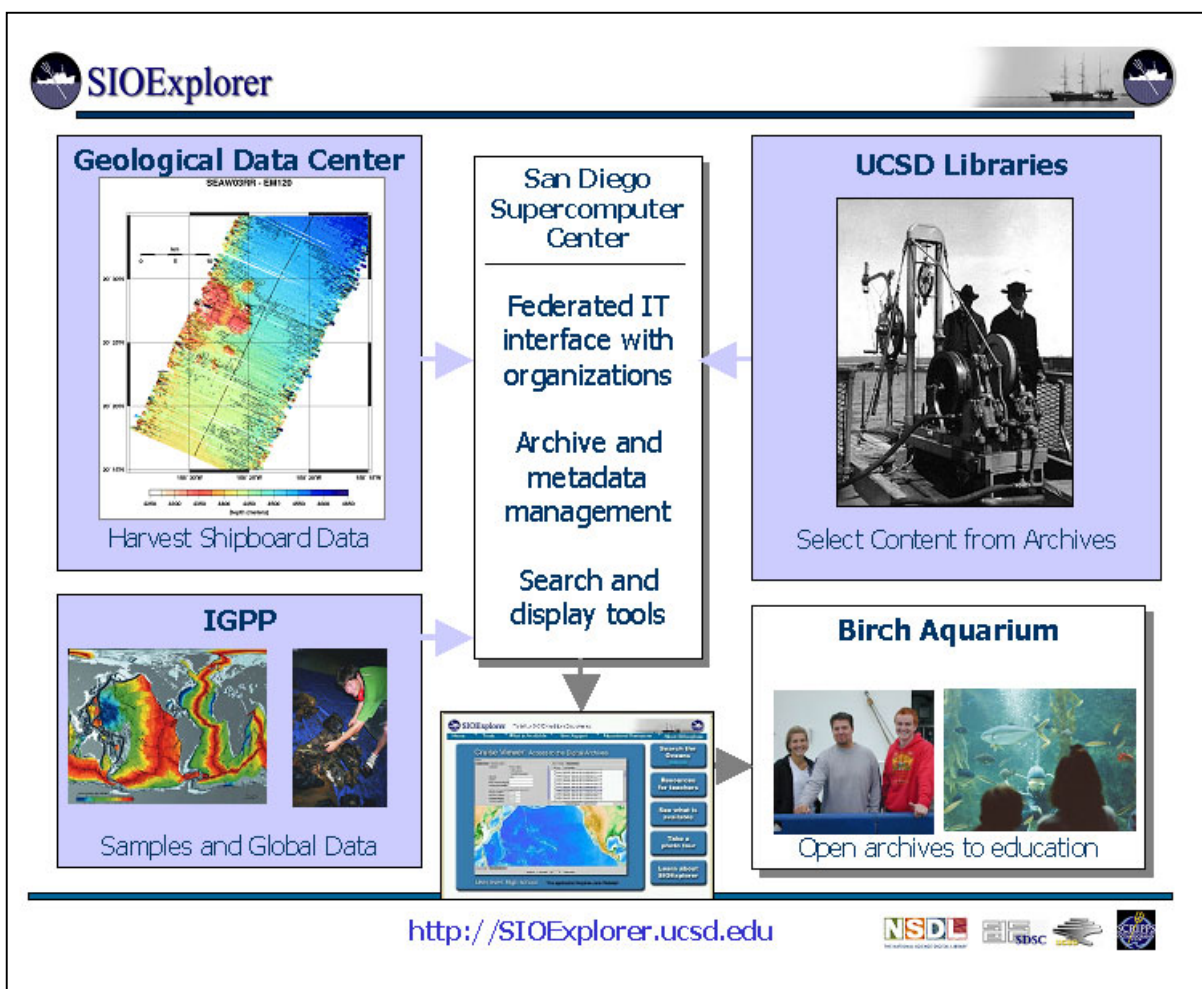


Figure 2. The SIOExplorer IT architecture federates the operations of three distinct data publishing authorities, integrates them into a searchable digital library, and makes them available for learning

digital library, and to include related historical images and documents from the SIO Archives in the SIO Library. Access to several global databases was added, along with the seamount catalog and geochemical resources of www.earthref.org [2]. An Advisory Board was formed, with representatives from academia, government and industry. SIOExplorer has now become a part of the National Science Digital Library (www.nsdl.org) and can be accessed directly at (<http://SIOExplorer.ucsd.edu>; Fig. 1).

It became clear that a scalable Information Technology architecture would be needed [3,4]. Data and documents from three separate organizations would need to be integrated initially, with more to follow in subsequent years (Fig. 2). Each organization had its own data standards and formats. Almost no metadata existed. With millions of files and approximately 1 terabyte of data, we realized that

a team approach would be required, combining the expertise of SIO, the UCSD Libraries and the San Diego Supercomputer Center. A "Collection Builders Toolkit" [5,6,7,8] has been developed to automate collection development, create and manage metadata, and geographically search the library (Fig. 3). Each digital object in the library has an associated metadata structure, which includes a Dublin Core block along with domain-specific blocks, as needed. Objects can be searched geospatially, temporally, by keyword, and by expert-level. For example, expert-level classification makes it possible to screen out research-grade contents, revealing material appropriate for the selected grade.

Now that the library has begun operation with an initial collection of cruises and supporting information, the next challenge will be to streamline the creation of educational material from the contents.

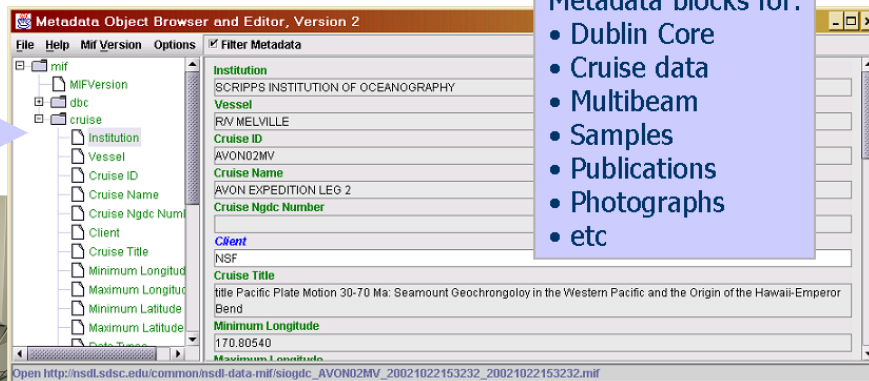
MOBE for Earth Science Metadata

Comprehensive approach for entire project

- Embed domain-specific metadata
- Avoid wading through details of other fields
 - Take advantage of recent developments in "human usability engineering"



Expand for your field as needed



Metadata blocks for:

- Dublin Core
- Cruise data
- Multibeam
- Samples
- Publications
- Photographs
- etc

"Making metadata as painless as possible for geologists"

Figure 3. A Collection Developers Toolkit includes modules for metadata harvesting and management. Metadata blocks for various data types or scientific disciplines are added, as needed.

Metadata for multibeam

Use MB-system

Create metadata at acquisition time on ship

Insert metadata into swath file as comment

Works with all vendors

Simrad, SeaBeam, etc.

Browse metadata with MOBE for
Bounds
Quality
Corrections
Processing steps
Beampoint editing status
Proprietary hold

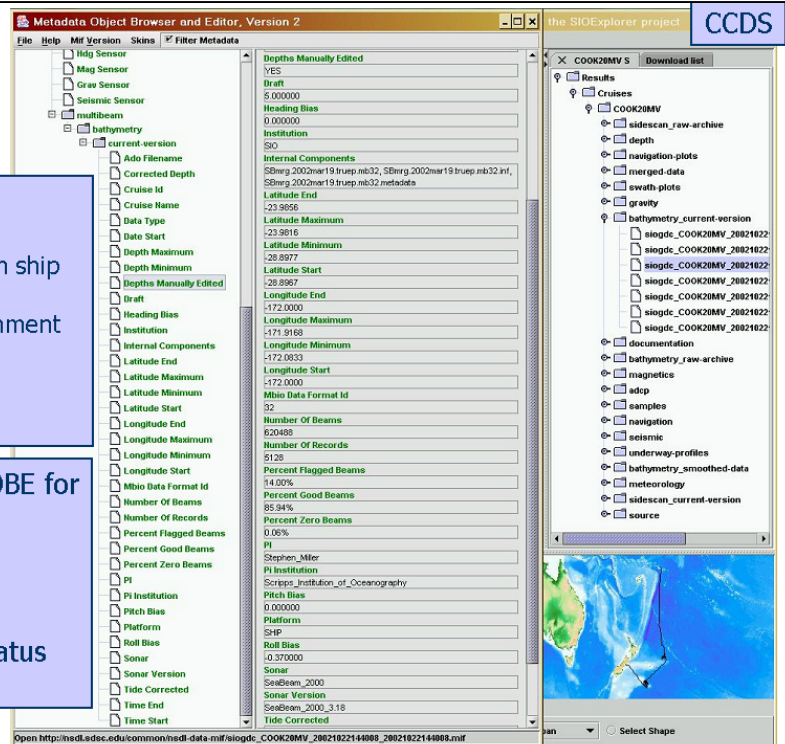


Figure 4. A sample CruiseViewer session shows a selected cruise (lower right, New Zealand to Samoa), the resulting data available (upper right), and an expansion of the metadata for a particular multibeam swath bathymetry file.

ACKNOWLEDGMENTS

These efforts are supported by the NSF/NSDL 0121684, "Bridging the Gap between Libraries and Data Archives," NSF/ITR 0113999 "SIOExplorer: Web Exploration of Seagoing Archives," as well as by SIO institutional funds.

REFERENCES

- [1] Miller, S.P., J. Helly, A. Koppers, P. Brueggeman, SIOExplorer: Digital Library Project, in *An Ocean Odyssey*, MTS/IEEE Oceans 2001 Proceedings, Volume 4, pp. 2288-2296
- [2] Keizer, P., A.A.P. Koppers, H. Staudigel, and J. Helly, 2001. The seamount catalog in EarthRef.org. EOS 82: F591.
- [3] Helly, J., T. T. Elvins, D. Sutton, D. Martinez, S. Miller, S. Pickett and A. M. Ellison, *Controlled Publication of Digital Scientific Data*, CACM, May 2002.
- [4] Helly, J., H. Staudigel, A. Koppers (2003). "Scalable Models of Data Sharing in the Earth Sciences." Geochemistry, Geophysics, Geoscience. DOI number 10.1029/2002GC000314.
- [5] Miller, S.P., J. Helly, M. Africa, U. Peckman, D. Day, and D. Clark, SIOExplorer: Overview, initial results and next steps, *Eos Trans. AGU*, 83 (47), Fall Meet. Suppl., Abstract OS61C-06, 2002.
- [6] Clark, D., S.P. Miller, U. Peckman, A. Chase, and J. Helly, SIOExplorer: Managing data flow into a digital library, *Eos Trans. AGU*, 83 (47), Fall Meet. Suppl., Abstract OS62B-0255, 2002.
- [7] Chase, A., and J. Helly, Bite-sized data management software: Developing specialized tools for manipulating metadata, *Eos Trans. AGU*, 83 (47), Fall Meet. Suppl., Abstract OS61C-12, 2002.
- [8] Sutton, D.W., J. Helly, S.P. Miller, A. Chase, and D. Clark, CruiseViewer: SIOExplorer graphical interface to metadata and archives, *Eos Trans. AGU*, 83 (47), Fall Meet. Suppl., Abstract OS62B-0254, 2002.