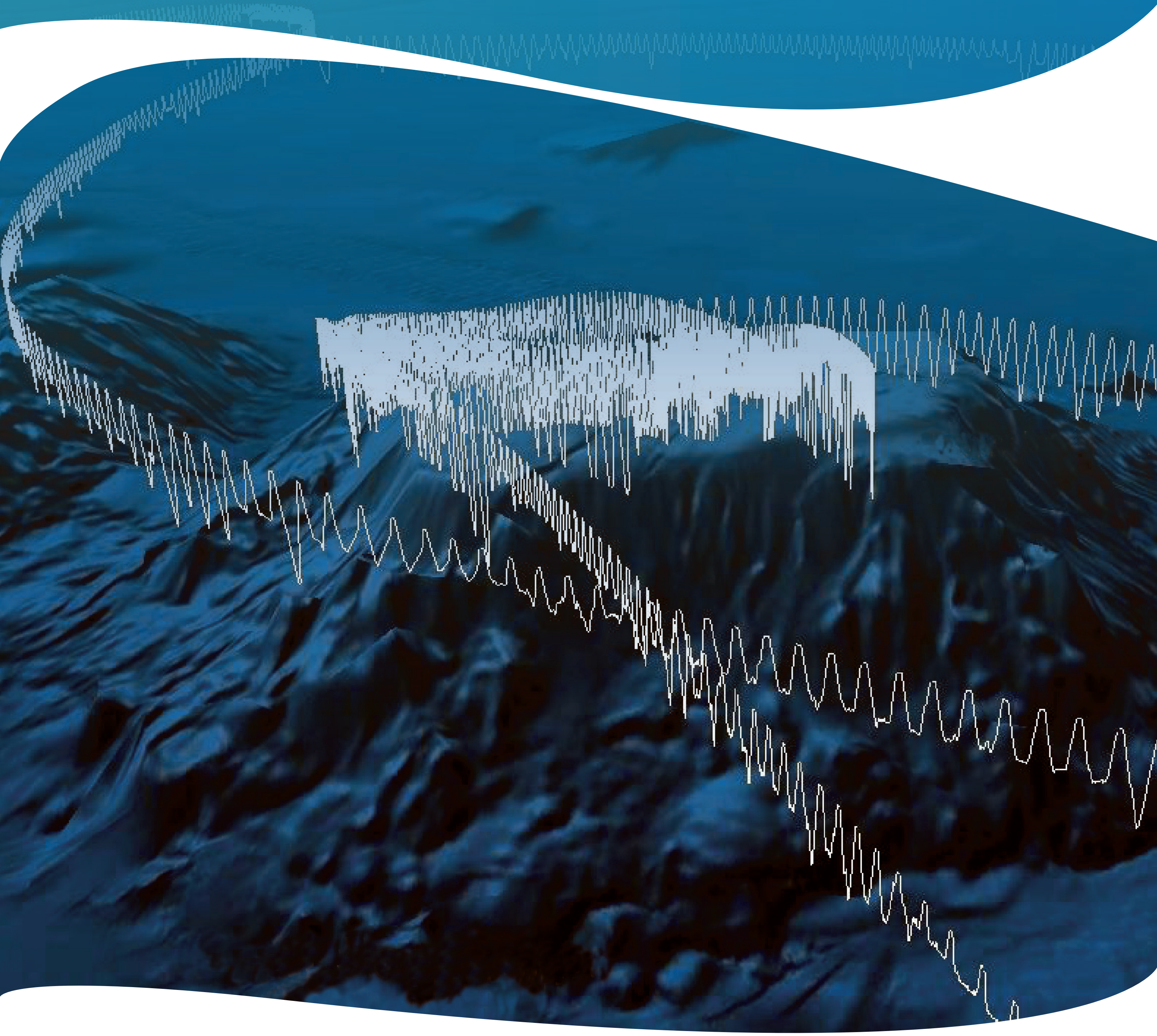
VISUALIZING OCEANS OF DATA

Educational Interface Design

2013 KNOWLEDGE STATUS REPORT



EDUCATION DEVELOPMENT CENTER, INC. • SCRIPPS INSTITUTION OF OCEANOGRAPHY

COVER IMAGE CREATED BY PATRICK ROBINSON

Maxwell, S. M., J.J. Frank, G.A. Breed, P.W. Robinson, S.E. Simmons, D. Crocker, J. Gallo-Reynoso, and D.P. Costa (2012) Benthic foraging on seamounts as a specialized foraging behavior by a deep diving marine mammal. *Marine Mammal Science 28*(3): E333-E344.

REPORT DESIGN

Designed by Jennifer Matthews *(jbmatthews@ucsd.edu)*, Scripps Institution of Oceanography, UC San Diego

Visualizing Oceans Of Data:
Educational Interface Design

Citation:
Krumhansl, R., Peach, C., Foster, J., Busey, A., and Baker, I. (2012). *Visualizing Oceans of Data: Educational Interface Design.* Waltham, MA: Education Development Center, Inc.

This material is based upon work supported by the National Science Foundation under Grant Nos. 1020002 and 1019644. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

CONTENTS

PREAMBLE

Science is data-intensive, but today's science education is not. In most classrooms, students' work with data is limited to reading graphs prepared by others, or at best collecting simple data sets themselves. While these student-collected data sets allow students to begin building their data proficiency, the conclusions that can be drawn and the lessons that can be learned from these data are limited in scope and can sometimes be compromised by data quality. The large, high-quality scientific data sets that are newly available online allow today's science students to incorporate working with authentic data into their learning experiences, giving them virtually unlimited opportunities to participate in real scientific work.

However, the fact remains that the educational promise of large scientific cyberinfrastructures will not be met without concerted effort. It is a huge leap to bridge from reading graphs or maps that have been carefully prepared to illustrate a particular concept to interpreting data visualizations that may not have ever been seen before, may have data problems, and may not show any obvious trend. It's also a huge leap to bridge from data that students have collected themselves to data that were collected remotely, by instruments students do not understand, in an environment they have not seen.

As one of our advisors, Jim Hammerman (August 22, 2012), noted:

> *It's a really hard and important problem. It shouldn't be so hard for people in schools to use [these professional data sets], but we all know it is. I'm interested in having these sorts of tools available for schools and citizen groups who want to make a difference in the world, making it possible for people to be curious, and making the case for what matters to them using data.*

The Oceans of Data project has made an attempt to define and confront what is "hard" for students and teachers who attempt to use large, online professional data sets. We feel passionately that it's important for us to do this to prepare today's students for tomorrow's world.

THE OCEANS OF DATA PROJECT TEAM

RUTH KRUMHANSL, principal investigator at Education Development Center, Inc.(EDC), provided technical leadership to the project team and coordinated project work with Scripps, the project's advisory board, and NSF. Her focus during implementation of the project was on reviewing and analyzing literature and developing guidelines relevant to Accessing Data and Geo-referenced Data Representations, and she led the synthesis and development of Visualizing Oceans of Data.

CHERYL PEACH, principal investigator at Scripps Institution of Oceanography, played a key role in ensuring the relevance of the Oceans of Data work to scientific cyberinfrastructure projects such as the Oceans Observatories Initiative. In addition, she arranged, co-planned, and hosted the advisory committee meetings, helped to visualize the structure of Visualizing Oceans of Data, and assumed primary responsibility for the dissemination of project findings,

JUNE FOSTER, co-principal investigator at EDC , was instrumental to the conceptualization of the Oceans of Data project, helping to shape the project goals and the research methodologies, and in particular contributing her expertise in Universal Design for Learning to the project. She was primary reviewer of all sections of Visualizing Oceans of Data and lead writer of the Cross-cutting Guideline section Enabling Customization.

AMY BUSEY of EDC was a primary author of Visualizing Oceans of Data. Her particular focus during the literature review and writing was on visual perception and cognitive load theory, and she was lead writer of the related Cross-cutting Guidelines sections. She also researched and wrote the specific guidelines for Animations.

IRENE BAKER of EDC completed the review and coding of literature related to graphs, and was lead author of the Graphs section of Visualizing Oceans of Data. She also conducted interviews with existing Web-based data providers as part of an initial needs assessment.

JACQUELINE DELISI of EDC acted as an internal methodological advisor, advising the project team as they refined the research methodologies, developed coding protocols, and analyzed findings.

KIRA KRUMHANSL of EDC played a critical role by searching for and obtaining literature relevant to the Oceans of Data project work.

ACKNOWLEDGMENTS

We'd like to acknowledge the contributions of our advisors, who shared their considerable experience and insights at two lively and stimulating meetings, as well as in telephone interviews, written comments, and e-mail communications. Their comments on the draft Knowledge Status Report greatly improved its content, particularly in areas where directly relevant literature is sparse. They brought diverse experience in education research, science research, teaching, educational software development, and cyberinfrastructure development to our work, which led to particularly interesting exchanges where we struggled to understand each other's language and perspectives. These productive struggles convinced us that more of these types of conversations are essential if we want to bring expert databases to students.

The Oceans of Data Advisory Board comprised the following members:

YI CHAO, Principal Scientist, Jet Propulsion Laboratory

DANIEL EDELSON, Vice President of Education, National Geographic

ALLISON FUNDIS, Research Scientist and Education and Public Outreach Liaison, Oceans Observatories Initiative RSN, University of Washington

BORIS GOLDOWSKY, Director of Technology, Center for Applied Special Technology

JAMES HAMMERMAN, Senior Researcher and Evaluator, TERC

KIM KASTENS, Doherty Senior Research Scientist, Lamont-Doherty Earth Observatory, Columbia University

JULIANNE MUELLER-NORTHCOTT, Biology and Earth Science Teacher, Souhegan High School

JOHN ORCUTT, Professor of Geophysics, Scripps Institution of Oceanography, UCSD

WILLIAM SANDOVAL, Associate Professor of Psychological Studies in Education, Graduate School of Education and Information Studies, UCLA

We'd also like to thank cognitive scientists Jess Gropen of EDC and Thomas Shipley of the Spatial Intelligence and Learning Center at Temple University for their thoughtful review and insightful feedback that enhanced the quality of this product, the National Science Foundation for funding this work, and our program officer Elizabeth Van der Putten for her support and encouragement along the way.

Introduction

I. INTRODUCTION

I. INTRODUCTION

About the Oceans of Data Project

The practice of science and engineering is being revolutionized by the development of cyberinfrastructures for accessing near real-time and archived observatory data. The NSF-funded project Oceans of Data aims to make it possible for students and their teachers to join that revolution.

The potential exists for classrooms to use state-of-the-art resources and techniques for scientific investigations and to analyze and draw conclusions from many kinds of complex data. But realizing that potential requires breaking new ground. As they stand now, the interfaces and data visualization tools for large science cyberinfrastructure databases are industrial-strength—designed by experts for use by experts—which significantly impedes broad use by novice learners.

What is needed are more "egalitarian" interfaces and data representations that make large scientific databases accessible to, and usable by, nonscientists (some of whom, hopefully, are budding scientists). But doing so is no easy matter for the software developer. Efforts to create interfaces and tools that bridge to the science classroom must be informed by state-of-the-art knowledge. The problem has been that such knowledge is dispersed across dozens of disparate disciplines, in thousands of books and journals, with no collation or synthesis to guide best practice. It is no wonder that developers sometimes have to rely on best hunches, rather than best practices, in their design efforts.

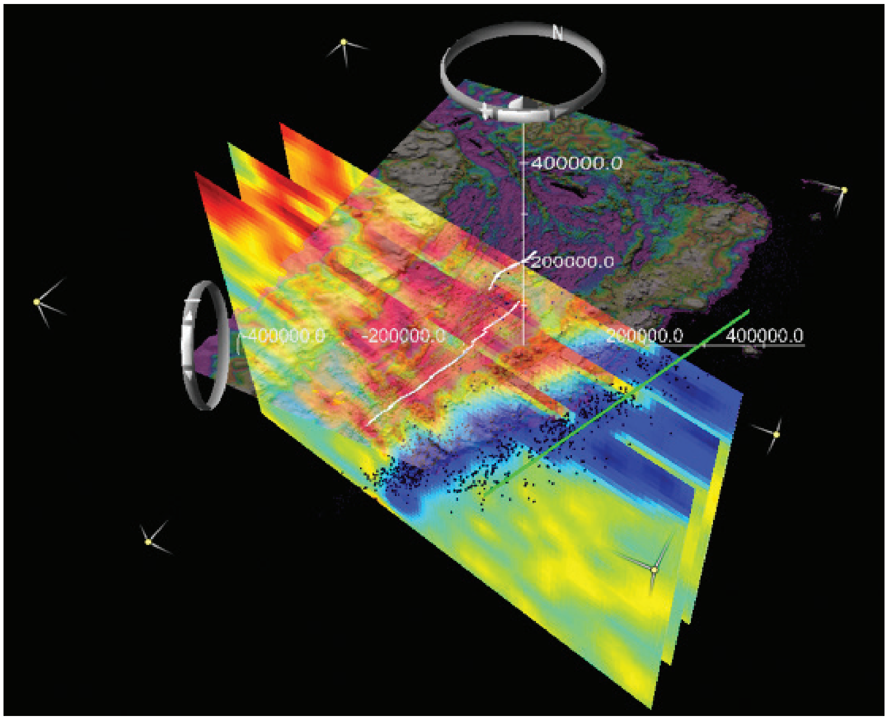


Figure 1. Experts use sophisticated data visualization techniques that may be very difficult for novices to understand. The displayed image is a snapshot from an interactive 3D visualization of the Lau Basin and Tonga Trench curtesy of Allison Jacobs. (Source: SIO Visualization Center, Scripps Institute for Oceanography Institute for Geophysics and Planetary Physics. Retrieved from siovizcenter.ucsd.edu/library/objects/detail.php?ID=138.)

To support interface and tool designers in their efforts to bridge cyberinfrastructure to the classroom, NSF funded Education Development Center, Inc. (EDC), and Scripps Institution of Oceanography to conduct the Oceans of Data project. Our goal has been to identify pertinent literature and expert opinion from the wide-ranging disciplines, to organize that knowledge into an initial integrated framework, to develop considerations and guidelines for educational interface design, and to present them in Visualizing Oceans of Data: Educational Interface Design, a knowledge status report (KSR).

We developed this KSR as a handbook with two key components:

- Guidelines for interface and data visualization tool development
- The considerations (principles, research, and theory) that inform these guidelines

Who Is the Audience for Visualizing Oceans of Data?

Our primary audience for the KSR is developers of interfaces for novice users. These developers will design and create interfaces that are easily navigable. They will define the capabilities that should be built into tools for visual representations of data, be they maps, graphs, or animations. They will construct important functionalities, such as varied color palettes suited to particular purposes, layering of information, alternative formats for representing particular data, and modes for scaffolding to support learning.

A caveat is in order: While the project goal was to array options for interface developers to consider, we recognize that, optimally, design decisions should be made in context—that is, taking into consideration the particular curriculum, the precise learning and teaching goals, and the needs and abilities of particular groups of students. Making appropriate design decisions therefore involves a cast of characters beyond interface developers (see Figure 2). This includes curriculum writers who understand how to guide students in their use of data to meet learning goals, and teachers who play perhaps the most critical role in facilitating students' use of data in the classroom.

Realizing the potential of large databases for student learning also requires the participation of an even wider set of actors. The scientists and database architects who develop the science cyberinfrastructure databases are pivotal. Professional development experts are necessary to help pre-college teachers gain confidence using scientific data and to help them develop strategies for engaging students with this new type of learning activity. Researchers are likewise central in continuing to fill knowledge gaps and build new understandings about learning in this new context. We hope that the KSR will be of interest and assistance to all of these key players as well.

This collaborative project considered in particular the complex observational data that are collected to support scientific research about the earth's oceans, atmosphere, and geosphere. However, the Key Underpinnings and guidelines in this document also have broader application to other scientific domains that hope to support students' access to and visualization of professional scientific databases.

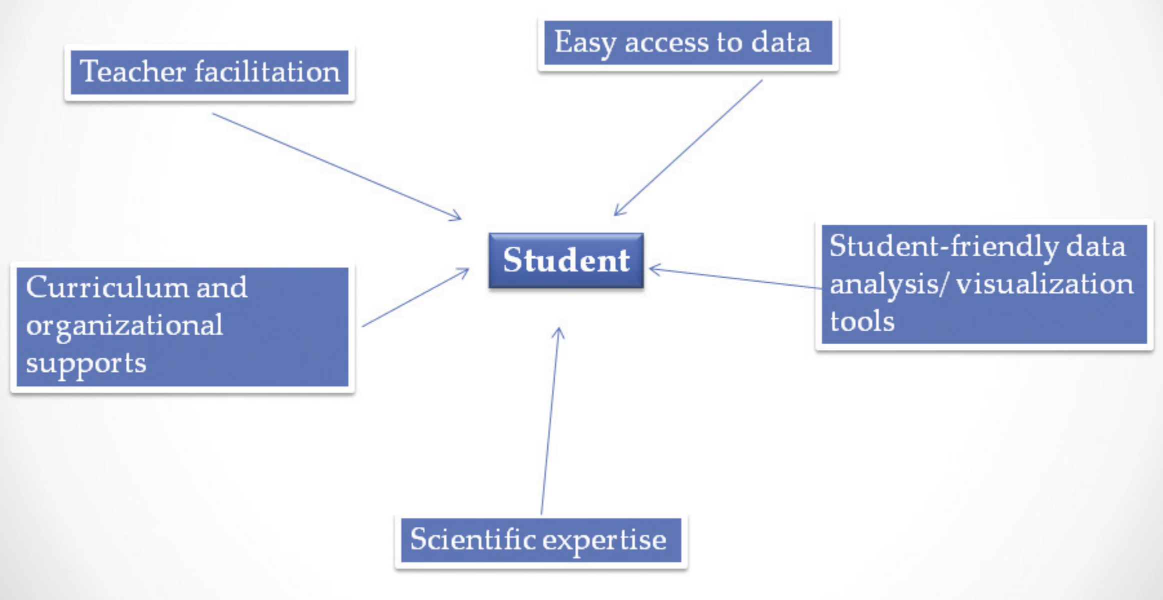


Figure 2. Careful design and testing of each of these elements is required to engage students in scientific practices using data in an online interface

The KSR at a Glance

By summarizing and organizing literature and expert opinion on the tenets underlying design recommendations, as well as the pros, cons, unknowns, and contradictions that sometimes emerge, we created this KSR to inform the process of developing interfaces and tools for data visualizations in the form of georeferenced data representations, graphs, and animations. The KSR is organized as follows:

II. KEY UNDERPINNINGS

Research and theory regarding three topics that are foundational to learning:

- Cognitive Load Theory: The mechanisms of working memory and long-term memory and how they relate to knowledge acquisition
- Visual Perception and Processing: How humans take in and make sense of visual information
- Schemata: How knowledge is stored, organized, and applied

III. CROSS-CUTTING GUIDELINES

Principles and corresponding recommendations that apply across the board to the design of interfaces and data visualizations:

- Adjust Cognitive Load: Designing the presentation of material so that it doesn't exceed the amount of information the learner can actively process
- Draw Attention to Important Features and Patterns: Promoting learning by using methods to highlight key information
- Enable Customization: Building in the capacity to meet different learner needs

IV. SPECIFIC CONSIDERATIONS AND GUIDELINES

The functions and tools particularly relevant to providing access to large scientific databases and facilitating students' work with these data. Design features to be used—or avoided—are addressed for the following:

- Accessing Data: Facilitating the selection and viewing of data parameters
- Geo-Referenced Data Representations (Plan Views, Cross-Sectional Views, and 3D Views): Promoting comprehension and analysis of geographically referenced data visualizations
- Graphs: Supporting interpretation of relationships among data using graphs
- Animations: Using dynamic presentations to represent change over time

V. FUTURE RESEARCH AND DEVELOPMENT: MAPPING THE TERRAIN

Questions relating to the following are presented to map the terrain of research and development that is needed and to focus on certain areas that we believe will be particularly fruitful:

- Authentic Data and Student Learning
- Interfaces and Data Visualization Tools
- Curriculum and Teacher Supports

How to Use the KSR

The KSR serves as both a reference and a tool. It is by no means a step-by-step blueprint for constructing interfaces and tools, for as yet there is no definitive state-of-the-art process for making large scientific databases usable by novice learners. What we offer, rather, is a resource to consult during the software planning and development processes. We know that the considerations and guidelines herein are many and complex. You may choose to pick the low-hanging fruit or to tackle a wide range of approaches. Whatever your *modus operandi,* we do have one recommendation for using the KSR: Please pay heed first to the Key Underpinnings and Cross-Cutting Guidelines chapters, for they offer an abridged orientation to the research, principles, and theories that too often remain under the radar. They also provide a basis for contemplating the considerations and guidelines in the subsequent chapter regarding data access, georeferenced data representations, graphs, and animations.

Students: The Ultimate Beneficiaries

Design decisions must of course be rooted in an understanding of the ultimate user group—students with limited prior experience working with professionally-collected scientific data. Throughout the KSR, we consistently discuss the characteristics and needs of the learners to be served.

The students for whom interfaces and visualization tools will be designed constitute a homogeneous yet diverse user group. Most will be in science classes that stress inquiry and will be called on to engage in key scientific practices, including, for example:

- Asking questions
- Developing and using models
- Planning and carrying out investigations
- Analyzing and interpreting data
- Using mathematics and computational thinking
- Constructing explanations
- Engaging in argument from evidence
- Obtaining, evaluating, and communicating information (National Research Council, 2012)

Virtually all K-16 students will begin their science studies as novices—that is, they will not have the expertise of scientists. As novices they will lack the kinds of knowledge and skill that shape what scientists "attend to and notice, how they organize new information and how they solve problems" (National Research Council, 2006, p. 95). Novices' reasoning and problem-solving will not be fluent. As a whole, they will probably have difficulty drawing inferences from data and making transitions from concrete to abstract thinking. And, of course, all novices will most likely lack any experience whatsoever in working with large science databases.

At the same time, these student users will differ markedly from one another. They will, for example, be divergent in the ways that they most effectively perceive and comprehend information that is presented in a data interface. While some will have more highly developed organizational abilities, some will be less well honed. They will bring different prior knowledge to the class, in terms of science content, mathematical and statistical reasoning, and experience with data visualizations. Their interests and motivation will likewise vary.

Suffice it to say that there is no perfect way to serve all students. But appropriately designed interfaces—in concert with the digital medium's capacity to provide for customization—can go far in igniting students' interest in working with large databases and in supporting their learning.

How We Developed the KSR

How did the notion of the Oceans of Data project arise? How did we go about constructing this resource? Here we describe in broad strokes the path taken . . .

THE INCEPTION

Our collective experience—as science teachers, curriculum developers, designers of student interfaces and curricula keyed to scientific databases, and scientists charged with making a new cyberinfrastructure database accessible to the public—made one thing quite clear: Developers of interfaces that enable nonscientists to work with large databases could use some help in the design process.

The idea of developing a resource to aid developers was exciting, ambitious, and a bit daunting. We marveled at the potential of putting scientists' databases and related tools (in modified forms) into the hands and minds of novice students. We knew that there are few studies of novice use of scientific databases, yet we were familiar with certain bodies of theory and research, as well as observations (our own and others'),that seemed quite germane. And we knew that potentially relevant knowledge was spread across a vast array of fields. Developing the KSR would not be straightforward.

THE PROCESS

From the beginning, we knew that we could not perform the typical literature review/synthesis, where only methodologically rigorous research studies are addressed, because there was so little research regarding access to and use of large scientific databases. We decided on an alternative, though pragmatic, route—addressing theory, expert opinion, and our own experiences, in addition to whatever research existed.

To establish the parameters for our search, we first identified key bodies of knowledge, reviewed some literature, tracked and reviewed some prominent citations in that literature, and conferred with the Oceans of Data Advisory Board and other experts. Thus emerged the focus on two key parameters: the different types of *data representations* that students might encounter (such as georeferenced representations, graphs, and animations), and the processes of *working with data* in which students would likely engage (for example, pattern recognition, finding or selecting data, and reading data representations). Through applying the preliminary coding protocol to several seminal works, we identified a third parameter, dubbed *cross-cutting* issues. This parameter refers to cognitive processes and other factors that relate across the board to various types of representations and actions involved in working with data. The cross-cutting parameters comprise such elements as cognitive load, spatial perception and visualization, prior knowledge, scaffolds and supports, navigation, and schemata. We then established the final coding protocol, while continuing to search for new literature related to our parameters. Testing for inter-rater reliability, we found that the protocol was appropriate to the task at hand and that coders were in agreement.

Our hunt for literature was wide-reaching. We searched a panoply of disciplines, including geosciences education, mathematics education, cognitive psychology, informatics, visual perception, cartography, neuroscience, computer science, learning science, and Universal Design for Learning. We followed up on citations from seminal works in order to ensure that our search was comprehensive and represented the current state of thinking across these fields. All in all, we reviewed over 300 documents (journal articles, books, and presentations), conferred with our ten project advisors, and consulted other experts from a variety of disciplines. We entered articles and other source information into NVIVO software, flagged relevant passages with codes so that we were later able to run queries on individual topics (e.g., animations) and cross-referenced topics (e.g., animations and Cognitive Load Theory) and obtain compilations of relevant quotes. We then summarized the considerations and guidelines that emerged from each query.

Given this burgeoning mass of information from disparate sources, how did we decide what literature to include, guidelines and considerations to report on and how to organize the findings? Following qualitative methods, we noted patterns and themes, identified "disconfirming evidence" (contradictory results), and clustered findings. As we made our judgments, we drew heavily on the collective expertise of the project team:

- 59 years of curriculum development work, including primary authorships of full-year high school Earth science, physics, and chemistry courses
- 20 years in applied science, focused largely on creating and looking for patterns in visualizations of georeferenced data
- 7 years of experience in cognitive science research
- 24 years of research on student learning and pedagogy in science
- 13 years of science teaching in public school classrooms
- 26 years of work in the development of educational software supports for science curricula and computer interfaces to authentic scientific data

Our combined efforts constitute a first step in harnessing knowledge to inform interface development. It is our hope that this KSR will serve as a catalyst for much-needed research, development, and testing so that the field gains a clearer understanding of what design features work (or don't), why, in what contexts, and for whom.

REFERENCES

National Research Council. (2006). *Learning to Think Spatially: GIS as a Support System in the K-12 Curriculum.* Washington, DC: The National Academies Press.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington, DC: The National Academies Press.

Key Underpinnings

II. KEY UNDERPINNINGS

II. KEY UNDERPINNINGS

This section briefly summarizes a large body of research that is fundamental to understanding how people take in information (such as a map or graph) and make sense of it. These discussions of Cognitive Load Theory, visual perception, and schemata form in large part the basis for the cross-cutting and specific guidelines in the sections that follow.

Cognitive Load Theory

The human brain offers two cognitive structures for storing information:

- *Long-term memory* provides subconscious and permanent storage for practically unlimited amounts of information (Atkinson & Shiffrin, 1968).
- *Working memory* is where information from the environment and/or long-term memory becomes the focus of active attention and processing. Unlike long-term memory, working memory can only hold a finite number of items simultaneously and for a quite limited period of time (Miller, 1956; Peterson & Peterson, 1959). During the learning process, new information is integrated with existing knowledge using working memory resources, and so the way these resources are allocated defines the limits of learning (Paas, Renkl, & Sweller, 2004; Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

Cognitive Load Theory describes three types of demands on working memory:

- *Intrinsic cognitive load* refers to mental effort due to the inherent difficulty of the content to be learned. As the complexity of the content (i.e., the number of interacting elements to be processed) increases, so does the intrinsic cognitive load (Sweller, 1994; Sweller & Chandler, 1994). Visualizations impose increased intrinsic cognitive load when the phenomena or data they represent are complex enough to be challenging to the user.
- *Extraneous cognitive load* refers to any effort required to understand material that's not directly related to the learning process (e.g., mental energy spent trying to find a poorly placed legend on a map). It is of particular concern to interface designers, as extraneous cognitive load often stems from a representation's design or format (Sweller, 1994; Sweller, Chandler, Tierney, & Cooper, 1990). Visualizations imposing extraneous cognitive load require the use of working memory for processing that is not pertinent to the task at hand, thereby reducing the cognitive resources available to engage with new, important, and challenging information.
- *Germane cognitive load* refers to any effort devoted to the construction of new knowledge (Sweller, van Merrienboer, & Pass, 1998). Visualizations that impose germane cognitive load support meaningful engagement with the content and the processing of new information in ways that lead to new or enhanced understandings.

Cognitive Load Theory is an important consideration for those providing students with access to large scientific data sets, such as oceanographic data, and it forms the basis for many of the guidelines in this KSR. Oceanographic and other Earth science data impose a high level of intrinsic cognitive load due to the number of interacting elements typically involved in Earth systems. As a result, it is critical that interface designers take steps to reduce extraneous load, alleviate intrinsic cognitive load, and maximize germane cognitive load.

A key point is that expert scientists already have well-formed domain knowledge in their long-term memory that they can apply automatically, freeing up the necessary working memory resources to read and interpret complex data representations (Kalyuga, Chandler, & Sweller, 1998; Pass & van Merrienboer, 1994; Sweller, 1994). However, novice learners must devote much more of their working memory to knowing how to approach the task, making sense of unfamiliar data sets and visualization formats, and constructing new understandings from what they see. Interface developers need to provide visualizations and other interface features