Abstract
The Oceans of Data project is an NSF-funded, 2-year interdisciplinary review of literature and expert opinion pertinent to making interfaces to large scientific databases accessible to and usable by novice learners and their instructors. The Visualizing Oceans of Data Knowledge Status Report (KSR) presents cross-cutting and visualization-specific guidelines that highlight how interface design features can address and ameliorate challenges novice users face as they navigate complex databases to find data, and construct and look for patterns in maps, graphs, animations and other data visualizations. Integrating information from hundreds of diverse sources, the resulting report is a set of over seventy guidelines that represents a synthesis of knowledge derived from key bodies of literature in the cognitive sciences, neurosciences, learning sciences, computer sciences, and many other disciplines. This paper presents the results of this work in the form of a small sampling of example guidelines from the full Visualizing Oceans of Data report.

Key words – Education, visualizations, interface, animations, guidelines

I. INTRODUCTION
Ocean science and engineering is being revolutionized by the development of cyberinfrastructure for accessing near real-time and archived observatory data (e.g. Ocean Observatories Initiative (OOI)). Large cyberinfrastructure projects have the potential to transform not only how oceanographic research is conducted, but also the way ocean science is taught in high school and undergraduate level classrooms. Yet while ocean science research is data-intensive, 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 scientific cyberinfrastructure 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 that they have not seen.

Access to sophisticated data and visualization tools can provide students with experiences using state-of-the-art techniques for scientific investigations, and afford them opportunities to analyze and draw conclusions from many kinds of complex oceanographic data. However, expert data access interfaces pose a significant barrier to novice (non-expert) users. Data portals, analysis tools and visualizations designed for professional scientists typically include specialized terminology and data structures, and assume knowledge that only experts within a specific discipline possess. Moreover, the process of finding, downloading, and visualizing/analyzing data requires a significant investment of time on the parts of instructors and their students, and only a small proportion of this time is spent extracting meaning from the data.

Knowledge that can inform the design of non-expert interfaces to scientific databases is broadly dispersed across dozens of disparate disciplines, in thousands of books and journals, with no collation or synthesis to guide best practice, making this knowledge difficult for educational interface designers to access and use in their design. To inform efforts at bridging scientific cyberinfrastructure to high school and undergraduate classrooms, Education Development Center, Inc. and the Scripps Institution of Oceanography conducted an NSF-funded, 2-year interdisciplinary review of literature and expert opinion pertinent to making
interfaces to large scientific databases accessible to and usable by novice learners and their instructors. The Visualizing Oceans of Data Knowledge Status Report (KSR) presents cross-cutting and visualization-specific guidelines that highlight how interface design features can address and ameliorate challenges novice users face as they navigate complex databases to find data, and construct and look for patterns in maps, graphs, animations and other data visualizations. The KSR has two key components:

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

The guidelines are underpinned by three broad, cross-cutting design principles (cross-cutting guidelines) that help make scientific databases more broadly accessible, specifically: 1) designing the user interface and visualizations so that they don’t contain more information than the learner can actively process; 2) drawing attention to important features and patterns; and 3) enabling customization of visualizations and tools to meet the needs of diverse learners.

This paper provides examples of the visualization specific guidelines for interface/tool developers to use to facilitate novice access to scientific databases. It is organized into four sections: Accessing Data, Georeferenced Data Representations, Graphs and Animations. Each section contains a brief synopsis of key considerations that inform the guidelines and highlights some of the key guidelines from the full Visualizing Oceans of Data report that should inform the design of educational data interfaces. Note the guidelines presented here are a small sampling of the full set of seventy-three discussed in the full KSR, which also includes an introduction to the research and theory underlying the report’s findings as well as cross-cutting guidelines that apply across the board to the design of interfaces and data visualizations.

II. ACCESSING DATA

This section focuses on the first thing that students will encounter when using an online data portal: the user interface that allows them to select and view data. At the time of this writing, very little research has been directly targeted at how to design Web interfaces that facilitate novice access to professional scientific databases. However, if this first step in the process of working with data isn’t handled well, students (novices) will go no further.

The considerations and guidelines in this section are based on the most relevant research we were able to find, as well as the experience of the project team and our advisors. It should be emphasized that beyond the suggestions in this section that specifically apply to data portals, it is especially important that developers adhere to good general design principles for user interfaces and universal accessibility (Johnson, 2010; National Center on Accessible Instructional Materials, n.d.; Shneiderman & Plaisant, 2010).

A. Considerations

The goal of an interface should be to maximize the degree to which students are able quickly find and view data, focusing their attention on important learning tasks, rather than the process of accessing the data. Novices’ lack of familiarity with domain vocabulary, and with the technologies associated with remotely collected data, can frustrate their ability to find appropriate data. If a digital information database does not support easy and intuitive browsing and discovery, this can pose a significant barrier to the use of the data.

B. Example Guidelines

There should be low or no barriers to downloading and visualizing a data set. Students should be able to get very quickly to the point where they are able to explore and generate questions about the data. To the maximum extent possible, processes that are not important to learning goals should be automated. One possible approach is to provide interesting default data sets that are immediately available, which are in turn connected to a larger variety of data that students can access on demand. Downloading data should connect seamlessly with generating data visualizations, and students should be able to modify visualizations or change data sets without going back to step one. The default settings for visualizations should be customized and optimized to the data so that the visualizations are immediately viewable, minimizing the need to go through steps to change settings (see Figure 1).
Make important controls stand out. Harrower and Brewer (2011) point out that learning how to use an interface involves “at least two critical steps: knowing what the buttons do and knowing the order in which to use them” (p. 266). They lament the fact that software engineers often hide key controls or options deep within the interface, rather than organizing them so that the most important controls are highest in the visual hierarchy. Making the most important controls stand out means keeping the webpages simple and clear of distracting information, such as advertisements, extraneous links, images, and animations (Nivala et al., 2011). Harrower and Brewer (2011) also suggest having controls appear only when they are needed to guide the user through the interface. Medycki-Scott and Blades (1992) suggest that presenting a map of the system and/or signposts to show where to go next can reduce the possibility that users will become lost. Gordin et al., (1994) suggest using a standard and familiar English-style text-informed arrangement for information on a webpage—arranged from general to specific, from controls to results, from left to right, and from top to bottom. Clearly marking steps that have been completed (by, for example, changing the color of a button) also helps users stay on track (Johnson, 2010).

Minimize expert terminology. Avoid using expert terminology or jargon (including that associated with the underlying programming architecture) and use familiar, task-focused descriptors on the user interface (Edelson et al., 1997; Johnson, 2010). Research has found that skilled (fast) reading is an automatic process of recognition and uses a different part of the brain from unskilled reading, which involves sounding out words and deciphering their meaning (Johnson, 2010). Thus, unfamiliar vocabulary can significantly disrupt students’ thinking processes and distract their attention from more important learning tasks.

Simplify and structure information. When navigating through a website, it is easier for people to scan quickly and automatically for relevant information when that information is presented in a concise, structured way (Johnson, 2010). For example:

- **Visually group related information** (such as labels and values) or controls by proximity, without adding bordered boxes or other labels that add to the visual clutter.
- **Have new information pop up over related content** to shift students’ focus temporarily without losing their connection with larger goals (Johnson, 2010).
- **Provide options that students can choose**, rather than requiring them to type in search terms—recognition is much easier for humans than recall.
- **Use pictures** as a clear and concise way to convey function, particularly if their meaning is consistent across the interface and, ideally, similar interfaces (Johnson, 2010).
- **Minimize text on webpages** and use restricted and highly consistent vocabulary that is understandable to a broad audience
- **Use text fonts that are easy to read**—sans serif for shorter bits of text in a screen display and serif for longer narrative reading (Evergreen, 2012). Avoid busy backgrounds and colors that interfere with each other.
- **Arrange text using a visual hierarchy**, with headings, bullets, and tables that make the text easy to scan and read automatically (Johnson, 2010).

III. GEO-REFERENCED DATA REPRESENTATIONS

Geographically referenced data sets —such as those associated with oceanographic, atmospheric, and geophysical databases— capture information about the geospatial distribution of measurements. Maps (plan views), cross-sections (depth views), and 3D visualizations (which present multiple spatial perspectives in one image) are powerful tools used by scientists to identify spatial patterns in data.
Most of the relevant literature available at the time of this writing pertains to how novices extract information from maps. More research is needed to understand how to help students understand cross-sectional views, which are commonly used in oceanography and the geosciences but much more rarely encountered elsewhere. Three-dimensional depictions and interactive visualizations show promise for helping students relate different data views. However, many of the tools that currently exist (such as the tool that allows users of Google Maps to draw a profile line on a map and generate a profile view) focus on conveying the shape of surfaces, not the three-dimensional distribution of subsurface measurements.

A. Considerations

Working with and thinking about geo-referenced data requires spatial and visualization abilities that vary significantly from one individual to another. Moreover, the visual features and patterns that a novice student perceives in a geo-referenced data representation will not be the same as those perceived by an expert. Geo-referenced representations of oceanographic and other Earth science data are complex and therefore likely to be inherently difficult for novices to understand. Moreover, creating new geo-referenced data representations using a Web interface is likely to be particularly difficult for those new to working with geo-referenced data. Appropriate design choices for geo-referenced data visualizations will vary according to the nature of the data and task, as well as the intended user.

B. Example Guidelines

Make visualizations simple and unambiguous without extraneous information. Reduce the number of data classes in a visualization (thus reducing the number of colors or the number of line symbols in a contour map, for example), which research has indicated can be a straightforward way to reduce demands on working memory (Miller, 1955; Phillips & Noyes, 1982; Slocum, 2008, as cited in Goldsberry & Battersby, 2009). Minimize unnecessary clutter in maps by changing the amount of detail displayed as the scale changes; include less detail in smaller-scale maps and place more emphasis on general form (Imhof, 2011).

Include information to minimize confusion and help students establish and maintain orientation. Missing information can make a visualization harder to read and understand, imposing unnecessary demands on working memory. To add clarity to a visualization consider the following:

- Clearly explain the meaning of colors and symbols.
- Display the units of measurement employed in the visualization.
- Give each visualization a clear title that allows students to quickly identify the type of data they are viewing.

Include features that orient the student (Edelson et al., 1997). For example, physical and political boundaries and labels can allow users to quickly register the location and scale of a map. Sometimes the addition of a simple locus map can be beneficial and help students relate detailed views to more global views.

**Identify areas of missing data.** Often, areas where data are missing are not adequately explained. For example, white areas designating “no data” on a map may be erroneously confused with ice-covered areas, and, depending on the color palette used, the areas with no data may be more eye-catching than those with data. Because novices are not likely to be familiar with data collection instrumentation and measurement techniques, they will not anticipate or understand these areas where data are lacking, which may cause confusion and/or push students beyond their capacity to easily grasp all of the information contained on the map. Therefore, areas where data are missing should be distinct from other areas through the colors or patterns used to designate them, and should be clearly labeled either directly or in an explanatory legend.

User color to make important features and patterns stand out. While experts have robust schemata related to the phenomena they are investigating and the patterns they’re looking for, novices often lack these schemata and will have difficulty recognizing the scientifically significant patterns buried in the “noise” of complex data visualizations. It is critical for designers to apply what is known about how to use color, textures, shading, and graphical interplay to advantage, ensuring that important features stand out to novice users.

- Use color palettes that take advantage of our ability to automatically process certain color differences.
- Do not use color in a way that overemphasizes certain features so that they mislead the student into seeing more than is actually in the data that are displayed (Imhof, 2011).
- Use color hue to represent categorical (nominal data).
- For ordered spatial data, use color palettes that vary primarily in luminance and saturation, or use a sequence of hues that are perceptually ordered.
- To show divergence in two directions from an average or threshold value, use two very different hues to represent the extremes, and decrease
saturation to a neutral color at the central threshold value

- Highlight important thresholds with distinct color changes.
- Provide alternative color palettes so that students can customize data representations to the data and task, with clear steps and guidance about how to do so effectively.

IV. GRAPHS

Graphs are one of the most important ways that scientists communicate and interpret relationships among data. Making graphs is a ubiquitous practice among scientists (Latour, 1990). But why are graphs so central to scientific practice?

A graph both transforms and reduces large amounts of data into a visual representation, which can then communicate patterns in data that typically cannot be directly visualized, due to the nature of the phenomena being studied or the number of data points that are being sampled. Graphs use spatial and visual features, such as length, angle, area, and color, to represent quantitative and categorical data and to show the relationships among data. The references for the data are usually, but not always, represented on axes. Graphs are used to analyze, explain, and predict phenomena.

An educational data interface might feature many types of graphs including scatter plots, line graphs, bar graphs, pie charts, box plots, and anomaly charts.

A. Considerations

The interpretation and creation of graphs by novice students will not be the same as an expert’s. Most students do not understand the nature of data. Graph comprehension consists of many kinds of cognitive tasks, and the complexity of the process is challenging for students relatively new to working with data. Moreover, accuracy of visual perception varies according to the visual features found in a graph, an important consideration in designing for non-experts.

B. Example Guidelines

Describe important values and trends associated with each data set (e.g., quantitative and/or qualitative information about trends or data features). A way to make a data set less abstract is to include supplemental information that gives further details about the data. These supplements can be quantitative (e.g., historical average and benchmark values) or qualitative (e.g., verbal descriptions of historical trends and shifts in the data).

Define and reduce visual chunks. When related parts of visual features are grouped by perceptual proximity, it is easier for students to identify visual chunks and recognize visual patterns. Carpenter and Shah (1998) and Shah, Mayer, and Hegarty (1999) found that grouping of visual features positively influences viewers’ spontaneous interpretations of graphs. It is important to understand that it is visual grouping rather than the type of graph, per se, that helps students recognize visual chunks. Color hues, boundaries, and placing related data in spatial proximity are ways to group visual chunks.

Multiple visual chunks of data make a graph difficult to interpret (Ratwani et al., 2008). Providing tools for drawing one or more statistical trend lines can help students interpret a graph by reducing the complexity of the visual features and by showing trends that are hard to distinguish by eye (see Figure 2). Variability in large data sets can make visual chunks and trends difficult to see. The data sets can be more easily visualized by providing a tool to reduce the number of points by averaging. Automatic scaling can obscure data details and affect the density of data points in scatter plots, making trends difficult to see and variability difficult to understand (Cleveland, 1993; Cleveland et al., 1982; Lauer & Post, 1989). Providing a tool to change the scale and aspect ratio (the ratio of width to height of a graph) can support students in identifying visual chunks and trends and more clearly identifying patterns in variability. Some data points are the result of error, and removing these “outliers” can clarify the trend of the data. However, a removal should always be justified. Make sure that students are able to explain their reasoning when removing what they perceive to be erroneous data.

Make it easy to examine and extract quantitative information. Students often need to review individual data points before looking for visual chunks and trends in graphed data (Ben-Zvi, 2002). Tools that show point values include those that enable presentation of a data set table and graph simultaneously, those that allow a
student to “mouse over” a point and see its value, and those that add grid lines to line graphs and bar graphs. Tables and graphs can be dynamically linked so that if a student clicks on a data point in one representation, that same data point is highlighted in the other representation.

Data from a scientific database are often labeled with abbreviations, quantities, and units that students have never seen, and the graphs created from those data sets will incorporate those labels. As much as possible, ensure that quantities and units in data sets and graphs are written out and are familiar to students. If unfamiliar terms or abbreviations must be used, include an explanatory caption (Kosslyn, 2006).

V. ANIMATIONS

Animated visualizations include one or more features that exhibit change over time. This change can take on a variety of forms, including transformation (changes to an object’s attributes, such as size or color), translation (changes to location in 2D or 3D space), and transitions (appearance or disappearance from a visualization) (Lowe, 2003).

There are multiple ways that animated visualizations might be integrated into an educational interface for scientific data, including animated geo-referenced or graph visualizations, animations that provide contextual information to support data exploration, or animated coordination between multiple data representations.

A. Considerations

According to principles of multimedia design, the structure and content of an external representation should correspond to the structure and content of the material it represents (Betrancourt, 2005). In theory, then, animated visualizations are a natural choice for illustrating changes that occur over time. In the context of exploring data, animations show the potential to enhance user understanding, for example, by facilitating comparisons (Nakakoji et al., 2001) or drawing users’ attention to specific aspects of a data set (Gordin et al., 1994). However, research also indicates that in order for students to reap these kinds of benefits, animations must be designed to address the challenges imposed by fleeting sources of information. It is important for interface designers to understand the potential difficulties that users might face and to design strategies that will support the effective use of animated visualizations, in appropriate contexts.

Novices have difficulty recognizing and remembering important changes, and animated maps can be particularly challenging. Novice users who have trouble reading static maps, which already require visual searching and scanning, can find it even more difficult when these representations are animated—and conversely many of the challenges that novices face with animations in general are exacerbated when they must grapple with an animated map.

B. Example Guidelines

Maximize users’ available working memory resources by presenting verbal explanations as audio narrations. Coordinating visual and verbal representations is thought to help novices build more coherent and complete schemata, based on the premise that each moves through unique processing channels and results in different forms of mental representations that can both reference and enhance each other (Mayer & Anderson, 1991; Paivio, 1986; Rieber, 1990).

However, trying to integrate written verbal explanations into an animation will most likely lead to a situation that overburdens users’ working memory. Users must either (1) take their eyes off important changes as they’re happening in order to read the text, or (2) try to hold the visual or verbal information in their working memory if the two are presented one after the other.

The literature offers a solution. Research on learning from animated multimedia supports the theory that visual and auditory stores of working memory are at least partially independent, and that one way to avoid overburdening users’ visual working memory resources and dividing their attention is to present verbal explanations and cues as narration rather than as accompanying text (Mayer & Moreno, 2003; Vekiri, 2002). This strategy might also be applied to legends in animated maps in the form of audible legends or audible supports to visual legends (Kraak et al., 1997).

Keep users’ attention focused by visually and temporally integrating related information. Particularly for learners with lower prior knowledge, it is important to provide both graphic and verbal information and to ensure that this information is integrated temporally as well as spatially (Plass et al., 2009; Vekiri, 2002). For example, the designer can add labels or explanations next to the animated object, show two animations with related information at the same time or back to back, or include an auditory narration over the animation (Mayer, 1997).

Give users control over the progression and pace of animations. Allowing users some form of control over an animation’s pace shows promise as a strategy for making animations feasible as educational tools. Whether users control the speed at which they progress through pre-identified segments or whether they have total control over the pace of an animation (using start/stop/pause buttons), interaction around an animation’s progression seems to provide a significant advantage by helping to optimize cognitive load (i.e., reducing extraneous cognitive load and increasing
VI. SUMMARY

Students’ use of large scientific databases holds tremendous promise — but this territory is largely uncharted. Student interfaces need to be developed and classroom tested through multiple cycles of research and development. More importantly, effective integration of ocean data into the classroom requires more than a student-friendly interface. New types of curricula are required to facilitate students’ learning with large, complex data sets. New types of instructor supports and professional development (particularly for pre-college teachers) are necessary to enable teachers to effectively guide students’ work. It will take the work of many researchers, curriculum developers and software developers along with iterative phases of development, design research, user testing, and pilot testing to realize the promise of professional scientific databases to transform science education.

ACKNOWLEDGMENTS

This work was funded by the National Science Foundation. We would like to acknowledge the contributions of our advisors, who shared their considerable experience and insights. Their comments on the draft report greatly improved its content, particularly in areas where directly relevant literature is sparse. We would also like to thank the Ocean Observatories Initiative Cyberinfrastructure Implementing Organization (OOI CI) for their role in inspiring the project, and for informing our thinking on the kind of guidance and information large cyberinfrastructure projects need to create interfaces and visualizations accessible to novice users.

REFERENCES


Mayer, R., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding