THINKING BIG

Transitioning your students from working with small, student-collected data sets toward "big data."

Kim Kastens, Ruth Krumhansl, and Irene Baker

Data are the foundation of science. Every insight and fact in science textbooks is grounded in evidence based on data. The *Next Generation Science Standards* (NGSS Lead States 2013) positions "analyzing and interpreting data" as one of eight science and engineering practices, and the *Common Core State Standards, Mathematics* (NGAC and CCSSO 2010) recognize "Measurement and Data" as one of the domains of mathematics to be fostered at all K–12 levels.

In the past, scientists—such as Galileo, Faraday, or Curie—generally worked with data collected personally or within a small team. Nowadays, scientific breakthroughs are more likely to come from complex data sets larger than one scientist could collect. These data, such as genomic DNA data stored at the National Center for Biotechnology Information (see "On the web"), may come from many individual labs. Or they may be from centrally coordinated instruments, such as the geoscientific EarthScope array (see "On the web"). Either way, the data are usually centrally archived and made available online free of charge.

Such data, now accessible to science teachers and students, are often of a quality and quantity previously seen only at research institutions. Large data sets are also increasingly important in other fields, from real estate to marketing to education to criminal justice to sports.

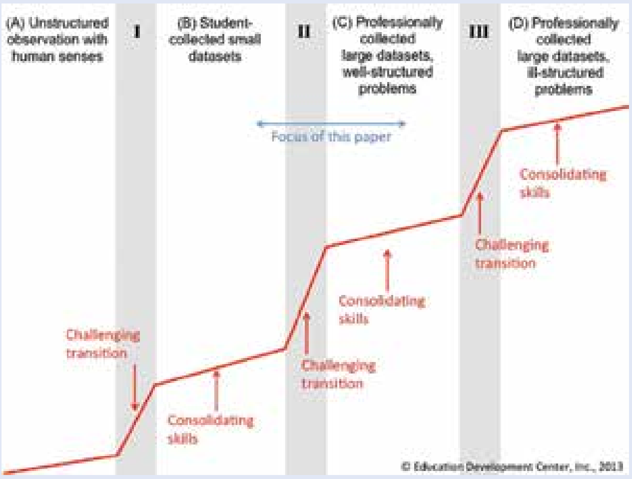
No matter what careers your students may pursue, analyzing and interpreting big data will be an important skill set.

This article is aimed at teachers already experienced with activities involving small, student-collected data sets and who are now ready to begin working with large, online data sets collected by scientists and engineers. We discuss challenges, instructional strategies, and sources of appropriate lesson plans.

FIGURE 1

Pathway to becoming a skilled data user.

The authors hypothesize that the pathway from child to skilled data user involves several challenging transitions; this article focuses on the transition from small, student-collected data sets to large, professionally collected data sets.



Challenges and changes

Figure 1 shows our framework for thinking about how children mature into skilled, data-using adults. At first, children observe the world in an unstructured way with their senses (domain A). They learn to make predictions and generalizations about phenomena experienced directly, such as: "Water will flow downhill." Next, as students in school, they work with small, self-collected data sets (domain B). Most traditional school-based investigations, including those described in the *National Science Education Standards* (NRC 1996), fall in domain B. An example is growing plants on a classroom windowsill under differing conditions of hydration.

Later, students work with larger data sets that they did not collect themselves. Most often, these are obtained online. At first, they work on fairly well-defined problems, such as locating tectonic plate boundaries from a database of earthquakes and volcanoes (domain C). Finally, they learn to work with large, complex data sets around ill-structured questions such as whether we should allow hydrofracking in our community or which medical treatment is best for an elderly relative (domain D). Figure 1 shows a trajectory with intervals of gradually increasing proficiency as the learner hones skills within one domain, interrupted by transitions involving big steps in learning.

FIGURE 2

Changes and challenges in transitioning toward big data.

Before	→	After
Student-collected data	→	Data professionally collected by scientists, engineers, technologists
Relatively small data sets	→	Relatively large data sets
Simple empirical observations (e.g., rainfall measured by rain gauge)	→	Complex derived parameters (e.g., precipitation derived from satellite remote sensing data)
Direct knowledge of context and setting in which the data were collected based on one's own senses	→	Sparse and abstract understanding of setting from metadata (e.g., date/time, lat./long., location map)
Participation in design of data plan yields direct knowledge of trade-offs in depth versus breadth	→	Secondhand knowledge of how data were collected
Aware of some potential data problems from personal experience (e.g., data loss, measurement error, instrument failure, operator error)	→	Must recognize possible data problems from attributes of the data themselves and/or the metadata
Simple, comprehensible analysis tools: pencil and paper, graph paper, calculator, spreadsheet	→	Powerful but opaque data display and/or statistical software

This article is about the transition from small, student-collected data sets to large, professionally collected data sets, labeled as Transition II in Figure 1. Many challenges and changes happen across this transition (Figure 2). While student-collected data typically have at most a few hundred data points, professionally collected data sets are often measured in gigabytes. Scientists' data sets are also likely to cover a longer time interval and use a higher sampling rate, more sampling stations, larger population, more conditions, and more data types, or have other forms of complexity. Students typically display their own data in simple tables or graphs but may need to use more complex statistical or visualization tools when using large, scientist-collected data sets (Krumhansl et al. 2012).

Furthermore, when students collect data themselves, they gain a deeper understanding of the process by which the data were generated and possible limits on data quality (Hug and McNeill 2008). They can develop a personal sense of the circumstances or environment from which the data were extracted and then draw on this understanding in making their interpretations (Roth 1996). All of this is lacking when working with data collected by others.

Instructional strategies

The transition from small, student-collected data sets to large, professionally collected data sets is challenging but well within the reach of high school students. Below we offer four, classroom-tested instructional strategies that support students in their first explorations of large, professionally collected data sets (Figures 3–6).

Strategy # 1: Data puzzles: This approach can be infused throughout your lessons and doesn't require technology. Seek out activities and test questions where students are asked to interpret data visualizations (for example, graphs or maps) of scientists' data (Figure 3). In well-designed data puzzles, the data snippets have been chosen because they contain a clear-cut manifestation of an important structure or process and thus offer the student a high insight-to-effort ratio (Kastens and Turrin 2010). In Figure 3, students can see evolution manifested in the beak depths of finches in the Galápagos Islands who did and did not survive a severe drought (Education Development Center, Inc., 2014).

In assessing students' understanding, begin with basic data-skills questions and be sure that at least some questions require students to reason about processes in the system the data represent (Baker-Lawrence 2013) and not merely to read values off the graph or map. For Figure 3, a question requiring students to decode and describe data would be: "Compare the mean beak size of the two finch populations." A reasoning question would be: "Based on your knowledge of life sciences and the data provided, suggest a possible explanation for the difference in mean beak size."

FIGURE 3

Data puzzle strategy.

The teacher devises questions to lead students to recognize connections between snippets of data and an important phenomenon or process.

Oceans of Data Institute Instructional Strategy #1

Data Puzzle

Procedure:

1) Curriculum developer identifies a small snippet of authentic data that embodies an important scientific concept, and develops data visualization(s) that foreground the patterns emerging from that concept.
2) Students view static data visualizations on screen or paper, and answer guiding questions about the system represented by the data.
3) Students experience a rewarding "Aha! moment" of recognition when they see the concept manifest in real world data.

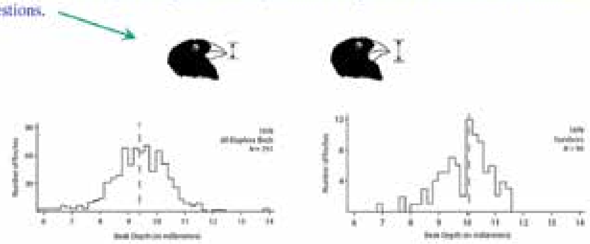
Theory of Action:

The dots, bars, squiggles, and blotches of color of a data visualization look nothing like the conceptual sketches and verbal descriptions by which students are typically introduced to scientific concepts. This type of activity allows students to see the connection between data and concept for clear-cut, unambiguous cases.

Example:

1) Students carry out a hands-on activity in which they use different implements—chopsticks, pin, sticky tape, fork straw, spoon, and so on—to pick up different foods as quickly as possible. Students make a table of the results and discuss the relationship between structural variations, adaptation, and survival.

2) Students compare data graphs on beak depth from Galapagos ground finches who did and did not survive a major drought. First, they answer straight forward decode and describe questions.



3) Finally, students discuss whether the process of natural selection can explain the change in beak depth and use the data to consider how new species could eventually evolve from the single species *G. fortis*.

From Learning Experience 4: So Many Species, So Much Time: *The Origins of Biodiversity*, in *Biology: Concepts & Practices*. Author: Education Development Center, Inc. 2014: Carolina Biological Supply Company, Burlington, North Carolina.

"This was challenging in a good way," said one teacher who used the finch activity in her classroom. "Some students got it right away and were able to answer the reasoning questions using the data. Others struggled and needed to go back to the data. I asked a few guiding questions to help them understand the connections. Next time I'll step back more and let students reason longer between the questions and the data."

Strategy #2: Nested data sets. Students first collect and interpret a small, student-collected data set and then interpret a larger data set (Figure 4). In this example, school groups first gain personal experience with the Hudson River estuary by collecting data near their school. Then they extend their inquiry upstream and downstream by bringing in data from other school groups and across time by bringing in professionally collected data from permanently installed sensors in the river. (For more on this activity, see "A Day in the Field," pp. 35–42.)

For assessment, combine questions addressing the student-collected data set with questions requiring use of the larger data set. A small data question for the Figure 4 example would be: "How did the salinity at our sampling site change as the tide rose?" A big data question would be: "Suggest three factors that seem to influence the salinity in the river and support your suggestions with data."

Strategy #3: Predict, observe, explain: This instructional sequence, often used for hands-on activities (Haysom and Bowen 2010), can help students explore large data sets (Figure 5). Based on their understanding of a natural process, students predict what a certain type of data will look like under a certain set of circumstances. In the example of Figure 5, students predict that *if* a distant star has an exoplanet orbiting it, *then* an observer on Earth should see the light intensity from the star diminish as the exoplanet passes between Earth and the star (Gould, Sunbury, and Krumhansl 2012). [For a complete description of this activity, see an article in the November 2014 issue of this journal (Gould, Sunbury, and Dussault 2014)]. Making a prediction gives the students a stake in the outcome, encourages them to observe more closely, and guides their inquiry through the database.

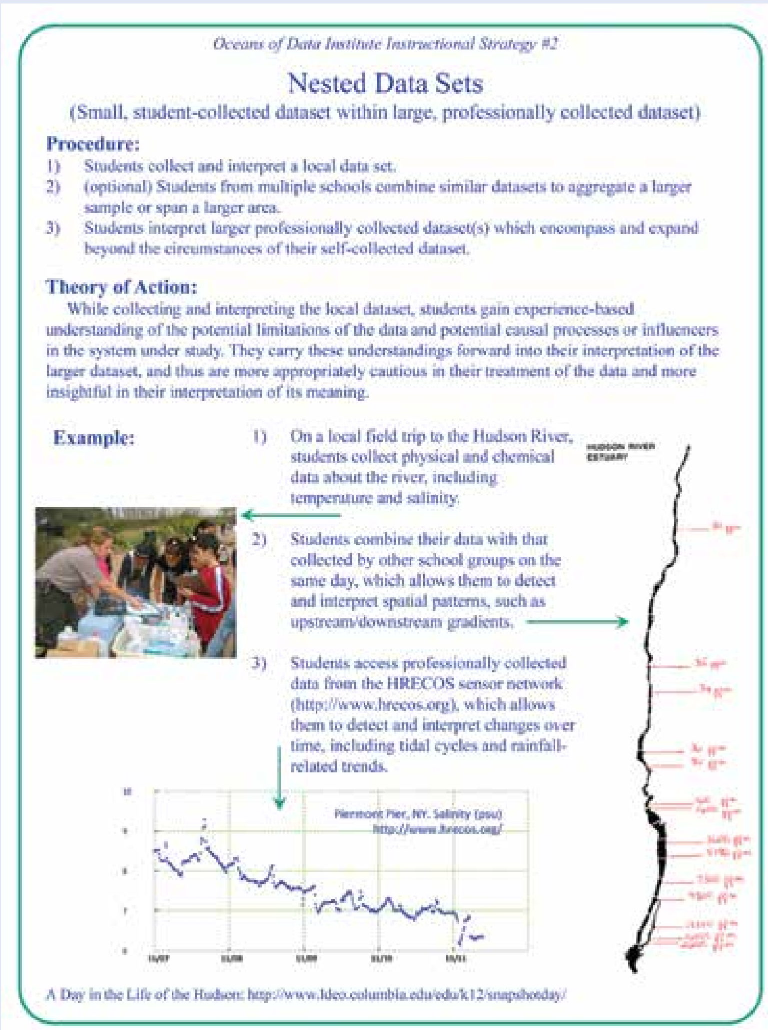
For assessment, probe students' understanding of the similarities and differences between the prediction and the observed data. When scientists use models, these similarities and differences reveal which aspects of the model have been successful and which need improvement (Kastens et al. 2013). For the example of Figure 5, you could ask: "In what ways did the data look like you expected?" "In what ways was it different?" "What are some possible causes for the similarities and differences?"

Strategy #4: Hypothesis array: In cases where students may not know enough about a system to make a well-formulated prediction, you can provide them with an array of working hypotheses (Figure 6, p. 30). The students explore and marshal the available data to support one of the

FIGURE 4

Nested data set strategy.

Students gain personal knowledge of the system while collecting their own data and then expand their scope of inquiry with professionally collected data.



WWW.LDEO.COLUMBIA.EDU/EDU/K12/SNAPSHOTDAY/; USED WITH PERMISSION.

hypotheses. In the example of Figure 6, the hypotheses take the form of sketches of landforms that might be represented in a topography/bathymetry database. Research shows that providing such an array of possibilities helps students explore data in a methodical, productive way (Mayer, Mautone, and Prothero 2002).

To assess well, students must not only choose the correct hypothesis but should also back up their selection with evidence and reasoning grounded in data. In the example of Figure 6, one good question would be: "For each of the hypotheses that you rejected, use data to show why that landform cannot be present in the area researched."

Sources of data and lessons

Armed with these strategies, you still need lesson plans and data for the specific topics you teach. Fortunately, as data of all types have become more abundant online, and data access tools more user-friendly, data-using lesson plans have proliferated. Figure 7 (p. 31) offers some reliable sources for classroom-tested lessons in which students work with scientist-collected data on a wide range of topics from weather to plate tectonics to evolution to genetics.

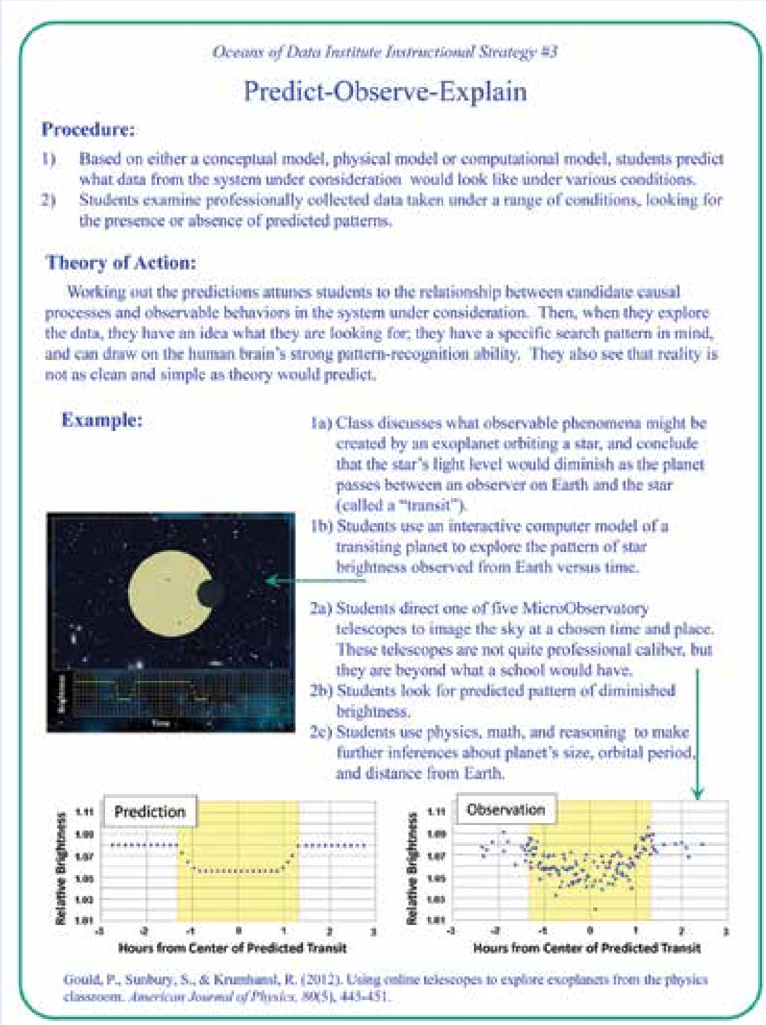
Conclusion

With your guidance, plus online data and their newly acquired big data skills, students can find evidence for some of the big ideas of science, rather than having to accept those ideas on authority from a teacher or textbook. They can conduct investigations about places and circumstances where they may never go, such as the bottom of the ocean. They can explore phenomena that are global in scope and span years in time, becoming a part of the community of users of that same data, and perhaps even make important discoveries of their own. Students' facility with complex data will be useful in a growing array of professions, from auto mechanics to health care, plus a necessary prerequisite for careers in science and engineering. ■

FIGURE 5

Predict-observe-explain strategy.

Students first predict what data will look like under certain circumstances and then make a data visualization to test their prediction.



Oceans of Data Institute Instructional Strategy #3

Predict-Observe-Explain

Procedure:

1) Based on either a conceptual model, physical model or computational model, students predict what data from the system under consideration would look like under various conditions.
2) Students examine professionally collected data taken under a range of conditions, looking for the presence or absence of predicted patterns.

Theory of Action:

Working out the predictions attunes students to the relationship between candidate causal processes and observable behaviors in the system under consideration. Then, when they explore the data, they have an idea what they are looking for; they have a specific search pattern in mind, and can draw on the human brain's strong pattern-recognition ability. They also see that reality is not as clean and simple as theory would predict.

Example:

1a) Class discusses what observable phenomena might be created by an exoplanet orbiting a star, and conclude that the star's light level would diminish as the planet passes between an observer on Earth and the star (called a "transit").
1b) Students use an interactive computer model of a transiting planet to explore the pattern of star brightness observed from Earth versus time.

2a) Students direct one of five MicroObservatory telescopes to image the sky at a chosen time and place. These telescopes are not quite professional caliber, but they are beyond what a school would have.
2b) Students look for predicted pattern of diminished brightness.
2c) Students use physics, math, and reasoning to make further inferences about planet's size, orbital period, and distance from Earth.

Gould, P., Sunbury, S., & Krumhansl, R. (2012). Using online telescopes to explore exoplanets from the physics classroom. *American Journal of Physics, 80*(5), 445-451.

Kim Kastens (kkastens@edc.org) *is a senior advisor, Ruth Krumhansl* (rkrumhansl@edc.org) *is a principal research scientist, and Irene Baker* (ibaker@edc.org) *is a senior curriculum/instructional designer at the Oceans of Data Institute of the Education Development Center, Inc., in Waltham, Massachusetts. In addition, Kastens is a special research scientist at the Lamont-Doherty Earth Observatory of Columbia University in Palisades, New York.*

Acknowledgments

We thank our colleagues in EDC's Oceans of Data Institute for sharing ideas and strategies, as well as the classroom teachers who collaborated with us on developing and testing activities for students to learn with data. This thinking has been funded by the National Science Foundation through grants EAR99-07689, GEO06-08057, and DRL11-38616 to Columbia

University, grants DRL04-39443, DRL10-20002, DRL12-22413, DRL13-13866, and DRL13-31505 to EDC, and by EDC's Oceans of Data Institute.

On the web

EarthScope array: *www.earthscope.org*

EDC's Oceans of Data Institute: *http://oceansofdata.org*

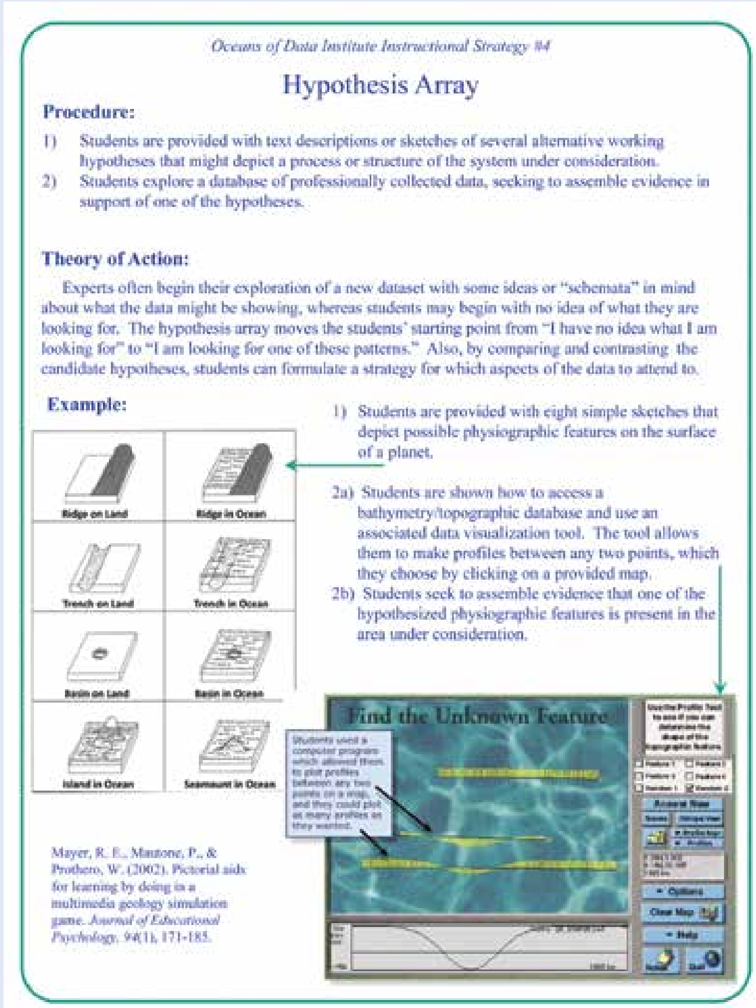
Larger versions of Figures 3–6: *www.nsta.org/highschool/connections.aspx*

National Center for Biotechnology Information: *www.ncbi.nlm.nih.gov*

FIGURE 6

Hypothesis array strategy.

Students exploring data from an unfamiliar system seek to support or refute several candidate hypotheses provided to them.



References

Baker-Lawrence, A.R. 2013. An investigation into instructional support for data analysis in high school science inquiry. *Dissertations and Theses.* Paper 1534.

Education Development Center, Inc. 2014. *Biology: Concepts and practices.* Burlington, NC: Carolina Biological Supply Company.

Gould, R., S. Sunbury, and M. Dussault 2014. In praise of messy data. *The Science Teacher* 81 (8): 31–36.

Gould, P., S. Sunbury, and R. Krumhansl. 2012. Using online telescopes to explore exoplanets from the physics classroom. *American Journal of Physics* 80 (5): 445–451.

Haysom, J., and M. Bowen. 2010. *Predict, observe, explain: Activities enhancing scientific understanding.* Arlington, VA: NSTA Press.

Hug, B., and K.L. McNeill. 2008. Use of first-hand and second-hand data in science: Does data type influence classroom conversations? *International Journal of Science Education* 30 (13): 1725–1751.

Kastens, K.A., and M. Turrin. 2010. *Earth science puzzles: Making meaning from data*. Arlington, VA: NSTA Press.

Kastens, K.A., A. Rivet, C.A. Lyons, and A.R. Miller. 2013. *Students' use of physical models to experience key aspects of scientists' knowledge-creation process*. Paper presented at the National Association for Research on Science Teaching, Rio Mar, Puerto Rico.

Krumhansl, R. 2014. *EDC Earth science.* Ronkonkoma, NY: Lab-Aids.

Krumhansl, R., J. Foster, A. Busey, I. Baker, and J. DeLisi. 2012. *Visualizing oceans of data: Designing educational interfaces.* Waltham, MA: Education Development Center, Inc.

Mayer, R.E., P. Mautone, and W. Prothero. 2002. Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology* 94 (1): 171–185.

National Governors Association Center for Best Practices and Council of Chief State School Officers (NGAC and CCSSO). 2010. *Common core state standards.* Washington, DC: NGAC and CCSSO.

National Research Council (NRC). 1996. *National science education standards*. Washington, DC: National Academies Press.

National Research Council (NRC). 2012. *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: National Academies Press.

NGSS Lead States. 2013. *Next generation science standards: For states, by states.* Washington, DC: National Academies Press.

Roth, W-M. 1996. Where is the context in contextual word problems? Mathematical practices and products in grade 8 students' answers to story problems. *Cognition and Instruction* 14 (4): 487–527.

FIGURE 7

Sources of K–12 activities using professionally collected data sets.

Source	Sample activities
The Center for Innovation in Engineering and Science Education, Stevens Institute of Technology *(http://ciese.org/realtimeproj.html)*	• Air pollution: What's the solution? • The Gulf Stream voyage • Catch a wave • Tsunami surge
Cooperative Institute for Meteorological Satellite Studies, Space Science and Engineering Center, University of Wisconsin–Madison *(https://cimss.ssec.wisc.edu/satmet/index.html)*	• Satellite winds • Weather forecasting • Wild weather • Satellite images • Monitoring the global environment
NOAA *(http://dataintheclassroom.noaa.gov/DataInTheClassRoom/)*	• Investigating El Niño using real data • Understanding sea level using real data • Understanding ocean acidification • Drawing conclusions: Weather maps
EDC *(http://oceantracks.org)* *(http://oceansofdata.org/our-work/edc-earth-science)*	• Investigating the migration of marine animals in response to seasonal changes in their environment (See "Harvesting a Sea of Data," pp. 43–49). • Patterns in ocean currents • Comparing regional climates
Earth Exploration Toolbook *(http://serc.carleton.edu/eet/index.html)*	• Analyzing plate motion using EarthScope GPS data • Climate history from deep sea sediments • Cool cores capture climate change (ice cores) • Detecting El Niño in sea surface temperature data • Water availability
Evolution and the Nature of Science Institutes *(www.indiana.edu/~ensiweb/lessons/p.tut.db.html)*	• Investigating evolutionary questions using online molecular databases • Footsteps in time: Analysis of Laetoli footprints • Varve dating: Dating sedimentary strata • Cytochrome-C Lab (amino acid sequences in several different animals)
Northwest Association for Biomedical Research *(http://bit.ly/1xXoAWw)*	• How bioinformatics is applied to genetic testing • How bioinformatics is used in genetics research
Bioinformatics Activity Bank *(http://bit.ly/19Jv67E)*	• Researching genetic disorders using BLAST • Medical problem solving: What is the cause of the seizures? • Eye color: Is blue really blue? • DNA surveillance unit: Is that an endangered whale you're eating? • Bear evolution

Harvesting a Sea of Data

Using authentic data to investigate marine migrations

Amy Busey, Ruth Krumhansl, Julianne Mueller-Northcott, Josephine Louie, Randy Kochevar, Kira Krumhansl, and Virgil Zetterlind

A baby elephant seal on the Valdes Peninsula, Argentina.

"My students were hot on the trail of a giant elephant seal," says Julianne Mueller-Northcott, a science educator at Souhegan High School in Amherst, New Hampshire. "They saw it leave the shoreline south of San Francisco, dive 700 meters, and then head north toward Alaska." The students weren't on a research ship but thousands of miles away in a marine biology classroom, following elephant seals and other marine animals virtually, via laptop.

With data from the Tagging of Pelagic Predators research program (see "On the web"), the students navigated the terrain of the ocean floor using Google Earth, following a bright red line tracking the elephant seal (Figure 1) as it circled off the coast of the Aleutian Islands. Adding tracking data from other elephant seals, they looked for patterns across multiple animals—noticing, for example, that some seals traveled north while others went directly out to sea. "I prompted them with such questions as: 'Where is the animal going?' 'What would motivate the seal to migrate like this?' 'Why is it circling in the one area—mating? feeding? avoiding a predator?'" Mueller-Northcott says. "My students were hooked! They had questions and access to a wealth of scientifically collected data and were ready to try to find answers."

FIGURE 1

A tagged northern elephant seal.

Because of their size, elephant seals can wear large electronic instruments with little effect on their behavior. The tags they wear can record location, salinity, temperature, and depth information. With such data, students can develop multiple lines of evidence to inform their ideas and support their claims.

DAN COSTA/TOPP

Ocean Tracks

Ocean Tracks: Investigating Marine Migrations in a Changing Ocean (see "On the web") is an innovative program that provides students free access to authentic data collected from migrating elephant seals, white sharks, albatross, tuna, drifting buoys, and satellites, as well as customized analysis tools modeled after those used by scientists. Ocean Tracks allows teachers and students to use large, professionally collected data sets to investigate scientific questions of current, real-world importance: What do marine animals' movements tell us about areas of the ocean that are critical in supporting biodiversity? In what ways are human activities affecting these areas?

Big data in the high school classroom

Scientific research is undergoing a "big data" revolution, as probes deployed in oceans, the atmosphere, and outer space provide near real-time data streams. As more and more data sets such as Ocean Tracks become available online, opportunities to engage students in the *Next Generation Science Standards* (NGSS Lead States 2013) practice of "analyzing and interpreting data" are blossoming (Figure 2). Students and teachers have unprecedented access to weather and climate data, images of stars and galaxies, seismic recordings, and more—data that take them not just outside the classroom but to the edges of our planet and beyond. With such abundant new data, students can ask and answer their own questions, perhaps identifying patterns that have yet to be discovered by scientists.

While large scientific data sets can potentially transform teaching and learning (Barstow and Geary 2002; Borne et al. 2009; Ledley et al. 2008; Marlino, Sumner, and Wright 2004; NSF Cyberinfrastructure Council 2007; Rainey et al. 2013; Slater, Slater, and Olsen 2009), access to data often comes with a catch: Data portals meant for scientists can be unintelligible to students and teachers due to cryptic labeling, unintuitive navigation structures, unfamiliar data visualizations, and complicated analysis tools. There is a need for critical scaffolds, including customized interfaces, guiding curricula, and tools that allow teachers to assess students' progress (Edelson, Gordin, and Pea 1997; Krumhansl et al. 2012; Quintana et al. 2004; Sandoval 2001).

To tackle these challenges, Oceans of Data, a National Science Foundation–funded project, set out to find and summarize what is known about designing data interfaces and

FIGURE 2

Connecting to the *Next Generation Science Standards* (NGSS Lead States 2013).

The materials/lessons/activities outlined in this article are just one step toward reaching the standards listed below. Additional supporting materials/lessons/activities will be required.

Disciplinary Core Ideas	Example
ESS2.E: Biogeology • The many dynamic and delicate feedbacks between the biosphere and other Earth systems cause a continual co-evolution of Earth's surface and the life that exists on it.	The Ocean Tracks interface and curriculum engage students in thinking about how Earth and ocean processes influence life in the oceans. For example, students use tracking data to identify areas of the ocean heavily used by marine life and investigate the oceanographic processes that create these biologically productive areas. (Curriculum Module 2)
ESS3.C: Human impacts on Earth systems • The sustainability of human societies and the biodiversity that supports them requires responsible management of natural resources.	The Human Impacts overlay uses data on a variety of activities and processes (e.g., pollution, shipping) to show the intensity of human impacts in different regions of the Pacific Ocean. Students use this overlay to describe human impacts on areas of the ocean that students have determined to be of importance to the Ocean Tracks species. (Curriculum Module 4)
LS2.A: Interdependent relationships in ecosystems • Organisms would have the capacity to produce populations of great size were it not for the fact that environments and resources are finite. This fundamental tension affects the abundance (number of individuals) of species in any given ecosystem.	Students use tracking data to identify the coast of California as an area heavily used by the Ocean Tracks species. To understand why, students learn how the process of upwelling creates productive areas for the prey of the Ocean Tracks species. Students must then understand the link between the Ocean Tracks species and the prey. (Curriculum Module 3)
LS2.B: Cycles of matter and energy transfer in ecosystems • Photosynthesis and cellular respiration provide most of the energy for life processes.	Students investigate the behavior of elephant seals by taking measurements from their tracks and linking these to a chlorophyll overlay to generate a map of where the elephant seal prey are likely to be found. Students construct a food web to illustrate the levels of energy transfer between these two groups of organisms. (Curriculum Module 2)
LS2.C: Ecosystem dynamics, functioning, and resilience • A complex set of interactions within an ecosystem can keep its numbers and types of organisms relatively constant over long periods of time under stable conditions.	After generating support for hypotheses about how the Ocean Tracks species are influenced by environmental conditions, students make predictions about how human impacts may affect marine species. (Curriculum Module 4)
LS4.C: Adaptations • Natural selection leads to adaptation, that is, to a population dominated by organisms that are anatomically, behaviorally, and physiologically well suited to survive and reproduce in a specific environment.	As students investigate the tracks of the Ocean Tracks species, they make discoveries about the habits of these animals. They examine how deep elephant seals can dive, the trans-Pacific journeys of the bluefin tuna, and the Laysan albatross's ability to fly incredible distances over short periods of time. Resources in the Ocean Tracks library help students understand how adaptations enable Ocean Tracks animals to accomplish remarkable feats. (Curriculum Modules 1–4)
LS4.D: Biodiversity and humans • Humans depend on the living world for the resources and other benefits provided by biodiversity . . . sustaining biodiversity so that ecosystem functioning and productivity are maintained is essential to supporting and enhancing life on Earth.	The Ocean Tracks "Hot Spot" tool measures the density of track points in a particular area of the ocean. Using background information on biodiversity from the Ocean Tracks Library, students consider whether the hot spots they identify are species hot spots or biodiversity hot spots. Using the Human Impact overlay, students see how extensively human activity affects their hot spot. With this information, students construct a plan to mitigate the effects of overexploitation, pollution, and other factors on their hot spot. (Curriculum Modules 3 and 4)

Science and Engineering Practices	Example
Practice 4: Analyzing and Interpreting Data Analyzing data in 9–12 builds on K–8 experiences and progresses to introducing more detailed statistical analysis, the comparison of data sets for consistency, and the use of models to generate and analyze data.	Students take measurements from tracks of Elephant Seals (speed, depth, and track curviness), and interpret this data to generate support for hypotheses about where the animals are displaying feeding behavior. (Curriculum Module 2)
Practice 6: Constructing Explanations Constructing explanations and designing solutions in 9–12 builds on K–8 experiences and progresses to explanations and designs that are supported by multiple and independent student-generated sources of evidence consistent with scientific ideas, principles, and theories.	Students use animal tracking data and oceanographic data overlays to identify habitat hotspots in the Pacific Basin. They then construct explanations for these phenomena, which requires them to integrate their measurements and observations with their understanding of the underlying mechanisms that create productive ocean habitat. (Curriculum Module 3)
Practice 7: Engaging in Argument from Evidence Engaging in argument from evidence in 9–12 builds on K–8 experiences and progresses to using appropriate and sufficient evidence and scientific reasoning to defend and critique claims and explanations about the natural and designed world(s).	Students make a case for the design and location of a marine protected area. Engaging in this debate requires that students provide data supporting why some areas of the ocean are more important than others for marine species, what human activities may affect those areas, and whether the location of these areas changes over time. (Curriculum Module 5)
Crosscutting Concepts	**Example**
Patterns • Different patterns may be observed at each of the scales at which a system is studied and can provide evidence for causality in explanations of phenomena.	Students describe and quantify patterns in the tracks of elephant seals, white sharks, bluefin tuna, and Laysan albatross. They then make quantitative comparisons between the migration patterns of these four species to determine the species that travels the fastest and farthest. (Curriculum Module 1)
Cause and Effect • Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects.	Students observe patterns in habitat use by large marine animals and use data and background information to identify the underlying causes of these patterns. (Curriculum Modules 2 and 3)

Ocean Literacy Principles (Ocean Literacy Network 2013)

Literacy Principle	Example
Principle 5: The ocean supports a great diversity of life and ecosystems.	Students observe the ocean in four dimensions as they follow the individual animals across the surface, down to the seafloor, and examine environmental factors that vary across space and time. Through an in-depth investigation of the Ocean Tracks species, students gain an appreciation for the remarkable adaptations of these animals to their ocean environment. (Curriculum Modules 2 and 3)
Principle 7: The ocean is largely unexplored.	Ocean Tracks uses data collected by the Tagging of Pelagic Predators research program. Using satellite technology, electronic tags allow us to learn more about ocean inhabitants and their environment than ever before. Students investigate questions of interest to practicing scientists as they examine where these animals live, travel, feed, and breed. (Curriculum Modules 1–4)

FIGURE 3

Learning science with big data.

Students discuss measurements they've taken from an elephant seal track. Working collaboratively allows students to support each other's thinking and often encourages spontaneous discussion about how to analyze and interpret the data.

BURT GRANOFSKY/EDUCATION DEVELOPMENT CENTER, INC. (EDC)

visualizations for high school students. Guidelines emerged (Krumhansl et al. 2012) that are being implemented and tested in the Ocean Tracks project introduced above.

Piloting the data interface

To develop Ocean Tracks, a team of marine biologists, geoscientists, curriculum developers, web designers, teachers, and education researchers collaborated to generate a web interface and teaching resources and to conduct preliminary research on the program. One of the teachers, Mueller-Northcott, piloted Ocean Tracks in her high school marine biology classrooms in the spring and fall of 2013. Her efforts and experiences, as well as those of the other pilot teachers, have provided insights about the potential of such programs to facilitate learning with big data in high school classrooms.

RANDY WILDER/MONTEREY BAY AQUARIUM

Pacific bluefin tuna migrate thousands of miles.

"After my marine biology students developed ideas about the seal's behavior based on their observations and background research," Mueller-Northcott says, "I challenged them to gather and use quantitative evidence to paint a clearer picture of the factors that might be influencing migration patterns." Working in pairs, the students identified key portions of the track that might support their hypotheses and created plots of the animal's speed and deepest daily dive (Figures 3 and 4), recording their measurements in a data table. Based on these measurements, along with their observations and the research they conducted using the online Ocean Tracks library, they formed hypotheses that they tested by gathering additional evidence. They added sea surface temperature and chlorophyll concentration overlays and took measurements from these data in areas of interest along the track.

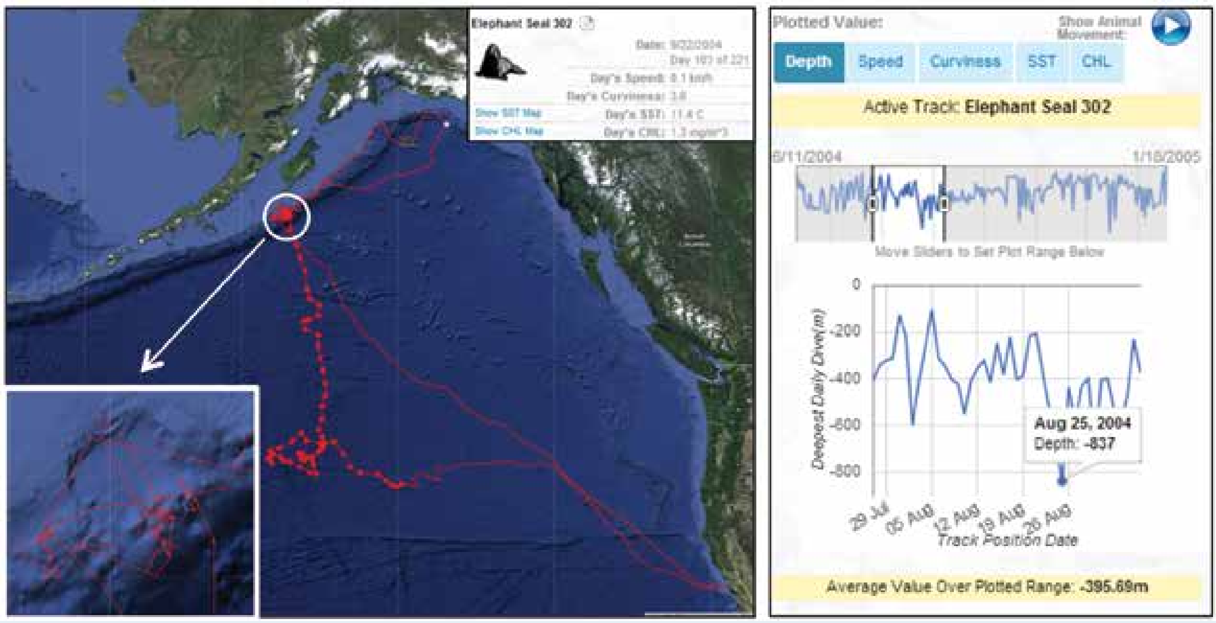
The students looked for patterns consistent with their hypotheses. Feeling confident in their claims, the students moved about the classroom, comparing other groups' seal track measurements displayed on posters with their own data. At first, some students were alarmed: "Our measurements aren't the same!" But then they surmised that the different groups may not have chosen exactly the same track intervals to measure and that, in fact, the patterns were similar in all the groups' data: "The elephant seal's average speed is lower, and the dives are deeper in the portions of the track where we think they're feeding," one student said.

"This was not the kind of discussion that we have very often in my marine biology course," Mueller-Northcott says. "When confronted with variation in their measurements and conflicting claims, my students helped each other grapple with issues such as what constitutes evidence, what the data (and differences in the data) actually represent, the significance of patterns and the meaning of data that don't fit those patterns, and how to assess confidence in your own and others' claims and evidence." As students delved into the data more deeply, the teacher pushed their thinking further by asking: "What do the data tell you?" "What other information can you use to help you?" "Who can gather the most evidence to support their hypothesis?"

FIGURE 4

Investigating elephant seal migrations in Ocean Tracks.

Students focus on sections of the elephant seal's track to measure the animal's speed, depth, and curviness (how much a track deviates from a straight line). Students determine the average, maximum, and minimum values for each variable and examine their data to look for patterns.



EDC AND STANFORD UNIVERSITY

Lessons learned

Teachers' and students' pilot work with Ocean Tracks yielded valuable lessons about how to engage students with professionally collected data sets accessible online. The following suggestions are based on pilot teachers' experiences and reflections (Sickler and Cherry 2011):

- Finding and preparing authentic data can be labor intensive for the teacher, and student experiences with these data are often limited to analyzing and interpreting a single set or type of data. In contrast, easy access to rich data sets can inspire questions, allow for explorations, and spur classroom discussions that are only possible when students explore multiple lines of evidence.
- Even when using a customized data interface, students still need support interpreting data and reading data visualizations, such as sea surface temperature maps or depth plots. Teachers who piloted Ocean Tracks found that students were particularly engaged when working in pairs or small groups and that it's important to have whole-class discussions and spontaneous debates about data in the classroom.
- Orientation experiences and teacher-support materials (e.g., curriculum guides, suggestions for implementation, content supports) are important, particularly for new teachers, as they help students learn to use scientific data. (Ocean Tracks pilot teachers used a teacher guide and attended a one-day training session on the interface and curriculum materials.) Ocean Tracks plans to offer expanded opportunities for virtual and in-person professional development in future phases of the program.
- To achieve the best results, supplement students' time on the computer with offline activities that let them practice their new skills.

DAN COSTA/TOPP

Male elephant seals can weigh up to 4,500 pounds.

- Encourage students to face the challenges of working with scientific data. Many will find it difficult to work without definitive "right" or "wrong" answers, to assess their work on how well data support their claims, and to make claims based on multiple types of data and repeated measurements. Remind students that the data are authentic and that their investigations are similar to the work of real scientists.
- Connecting students' experiences with data to their own lives can motivate them to explore questions related to that data.

Conclusion

Students and teachers today are poised alongside scientists at the frontier of the big data phenomenon. The opportunities for providing better access to big data sets are rife, and Ocean Tracks is serving as a model. However, much still needs to be learned about what teaching strategies may help students (and teachers) learn to work with and analyze big data. This article can spur further exploration of using big data in the high school classroom. Armed with the right tools and instructional strategies, the possibilities for learning about the world through data are boundless. ■

Amy Busey (abusey@edc.org) *is a research associate, Ruth Krumhansl* (rkrumhansl@edc.org) *is a principal research scientist and director of the Oceans of Data Institute, Josephine Louie* (jlouie@edc.org) *is a research scientist, and Kira Krumhansl* (kkrumhansl@edc.org) *is a research associate at Education Development Center, Inc., in Waltham, Massachusetts; Julianne Mueller-Northcott* (jmueller-northcott@sprise.com) *is a science educator at Souhegan High School in Amherst, New Hampshire; Randy Kochevar* (kochevar@stanford.edu) *is a marine biologist at Stanford University's Hopkins Marine Station, in Pacific Grove, California; and Virgil Zetterlind* (vzett@earthnc.com) *is the chief technology officer at EarthNC in Delray Beach, Florida.*

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1222413. The authors would also like to thank Christine Brown, Jacqueline DeLisi, Silvia LaVita, Jessica Sickler, Diane Stanitski, and Brad Winney for their invaluable contributions to this work.

On the web

Ocean Tracks learning modules (a series of investigations); teacher guide, including supplemental activities; and multimedia supports (virtual library and video resources for students and teachers): *http://oceantracks.org*

Programs, products, and research focused on unlocking the potential of big data in education: *http:oceansofdata.org*

Tagging of Pelagic Predators (TOPP): *www.topp.org*

References

Barstow, D., and E. Geary (eds). 2002. *Blueprint for change: Report from the national conference on the revolution in Earth and space science education*. Cambridge, MA: TERC. *http://bit.ly/18SpnvP*

Borne, K.D., S. Jacoby, K. Carney, A. Connolly, T. Eastman, M.J. Raddick, and J. Wallin. 2009. The revolution in astronomy education: Data science for the masses. State of the Profession Position Paper, submitted to the Astro2010 Decadal Survey. *http://bit.ly/1MFAg1s*

Edelson, D., D. Gordin, and R. Pea. 1997. Creating science learning tools from experts' investigation tools: A design framework. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Oak Brook, IL.

Krumhansl, R., C. Peach, J. Foster, A. Busey, and I. Baker. 2012. Visualizing oceans of data: Educational interface design. Waltham, MA: Education Development Center, Inc. *http://bit.ly/1Cn4dUn*

Ledley, T.S., A. Prakash, C. Manduca, and C. Fox. 2008. Recommendations for making geosciences data accessible and usable in education. *Eos* 89 (32): 291.

Marlino, M.R., T.R. Sumner, and M.J. Wright. 2004. Geoscience education and cyberinfrastructure. Boulder, CO: Digital Library for Earth System Education Program Center, University Corporation for Atmospheric Research. *http://bit.ly/1MEPMMO*

National Science Foundation (NSF). 2007. NSF's cyberinfrastructure vision for 21st century discovery. *http://1.usa.gov/1x87BR2*

NGSS Lead States. 2013. *Next Generation Science Standards: For states, by states.* Washington, DC: National Academies Press.

Ocean Literacy Network. 2013. Ocean literacy: The essential principles and fundamental concepts of ocean sciences for learners of all ages. *http://1.usa.gov/1DrWSCT* .

Quintana, C., B.J. Reiser, E.A. Davis, J. Krajcik, E. Fretz, R.G. Duncan, and E. Soloway. 2004. A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences* 13 (3): 337–386.

Rainey, D., L. Coyne, J. Gibson, S. Cammer, J. Schulman, B. Tretola, and O. Crasta. 2013. A project-centric course on cyberinfrastructure to support high school STEM education. *Systemics, Cybernetics and Informatics* 11 (3): 8–13. *http://bit.ly/1OL6DNd*

Sandoval, W.A. 2001. *Students' uses of data as evidence in scientific explanations.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Sickler, J., and T. Cherry. 2013. *Ocean Tracks interface development: Evaluation findings*. Columbus, OH: Lifelong Learning Group.

Slater, S.J., T.F. Slater, and J.K. Olsen. 2009. Survey of K–12 science teachers' educational product needs from planetary scientists. *Astronomy Education Review* 8 (1): *dx.doi.org/10.3847/AER2009005*

U.S. Global Change Research Program. 2009. Climate literacy: The essential principles of climate sciences. *http://1.usa.gov/1DrWSCT*