Building Global Interest in Data Literacy: A Dialogue

Workshop Report

Data drive discovery, decision making, and innovation in all aspects of human endeavor.

Data allow us to track the progress of diseases and identify cures; to predict and prepare for natural disasters and national security threats; and to develop new products that will have broad usefulness and spur new avenues of economic development. Data support wise decision making about everyday challenges—from choosing the best car to buy to diagnosing how to fix it. Data about our actions, our movements, and our personal decisions are being collected via our tablets, our computers, and our cell phones, and will potentially be used in ways no one has currently imagined. These changes bring with them an imperative to prepare students with new kinds of knowledge and skills—not only so they can be productive members of the workforce, but also so they can participate as educated citizens, making evidence-based decisions in a complex world.

What does it mean to be data literate in the world of "big data"? What should we be teaching students to better prepare them to participate in today's workforce and society? What steps need to be taken to develop critical data literacy skills in schools? To seek answers to these questions, EDC's Oceans of Data Institute (ODI) convened expert panels for two workshops. In late 2014, a panel of big data analysts from leading business, government agencies, and universities produced a profile of the "big-data-enabled specialist" (BDES),

which described what they do in their jobs and the essential knowledge, skills, and behaviors they must have. This profile was reviewed and validated by more than 150 big data professionals from nearly every industry sector. In October 2015, ODI and IBM convened a second workshop, Building Global Interest in Data Literacy: A Dialogue, which brought education experts into the conversation.

Over three days, the panel of big data analysts and education experts engaged in a structured dialogue that produced new insights into the nature of data literacy in today's big data society, envisioning what the data-literate person does, and delving into the meaning of analytical thinking and how we might teach this in K–16 classrooms.

Key Workshop Outcomes

The panel developed and endorsed the following definition of data literacy:

> The data-literate individual understands, explains, and documents the utility and limitations of data by becoming a critical consumer of data, controlling his/her personal data trail, finding meaning in data, and taking action based on data. The data-literate individual can identify, collect, evaluate, analyze, interpret, present, and protect data.

The panel also acknowledged the importance of teaching a range of skills that are important foundations of data literacy and, in particular, emphasized the critical importance of encouraging curiosity, skepticism, and persistence in the classroom.

The Call for Action to Promote Data Literacy, produced and signed by all panel members at the end of the workshop, calls for "a revolution in education, placing data literacy at its core, integrated throughout K–16 education nationwide and around the world." The experts on the panel "see a growing urgency for the promotion of global data literacy for the following reasons:

- Our world economy and our jobs are increasingly shaped by data and by the knowledge and skills required to use it effectively.
- We are all perpetually, and often unknowingly, producing streams of data, which we need to be more aware of and shape and manage to ensure our privacy and personal security.
- Effective use of data empowers us to make objective, evidence-based inferences and fundamental decisions affecting our lives, both as individuals and among societies."

The Global Data Initiative proposed at the end of this workshop set the ambitious goal of bringing data literacy to 100 million students by 2021. To achieve this goal, the panel articulated the importance of substantive efforts in the following three areas:

- Engaging stakeholders and building a global data literacy community
- Developing and disseminating educational resources
- Incorporating data literacy into the international standards and assessments that are used to define well-educated students

The Full Report

Subsequent sections of this report provide a more complete presentation of the key outcomes of the workshop **Building Global Interest in Data Literacy: A Dialogue**

Section 1
What Is Data Literacy?

What do we mean when we speak about data literacy? Who is the person both comfortable with, and adept at, employing data prudently in the workplace as well as in their personal life? How do we know if someone is data literate? What exactly are data-literate individuals able to do that demonstrates their capacity to work with data effectively? What resemblance do the activities of data-literate individuals bear to the activities performed by data-enabled professionals? Questions like these propelled energetic dialogue between members of the expert panel and resulted in two key products—a definition of data literacy and a description of activities the data-literate person is able to do. These two products are presented below, along with brief descriptions of how they were developed.

Definition of Data Literacy

To initiate the dialogue, ODI offered the panel a draft definition of the data-literate individual for their consideration. The panel was then challenged to create its own definition, one that was succinct and comprehensive, and most importantly, one that described what a data-literate individual is able to do. The panel worked through several iterations before coming to consensus on the following definition, which anchored all subsequent conversation:

> The data-literate individual understands, explains, and documents the utility and limitations of data by becoming a critical consumer of data, controlling his/her personal data trail, finding meaning in data, and taking action based on data. The data-literate individual can identify, collect, evaluate, analyze, interpret, present, and protect data.

Activities the Data-Literate Person Is Able to Do

To develop a more comprehensive description of what the data-literate person is able to do, the panel worked in small groups comprising experts in education and data analytics. They analyzed ODI's profile of the big-data-enabled specialist (BDES), produced in 2014, to determine the extent to which the activities in the profile corresponded to the range of activities in which a data-literate individual might engage. The panel discussed which tasks should be eliminated and whether any needed to be added. The following summarizes the key activities and their components. (Note: These descriptions are intended to fully capture the range of tasks associated with data literacy; it is not expected that any one individual would be able to perform all of these activities.)

Defines the Problem

Includes articulating questions, identifying the audience and context for the problem, and determining what will be observed and the tools to be used for measurement and for assessing risk and bias involved in conducting the study.

Wrangles/Handles Data

Includes exploring the data that are available and/or collecting data, mapping data across heterogeneous sources, cleaning/transforming/synthesizing and visualizing data, writing software to automate data analysis tasks, documenting the data handling process, handling data with the understanding that the data set is an imperfect construction that involves decisions and assumptions.

Self-Manages Data Resources

Involves planning with an awareness of a data lifecycle, fulfilling ethical and legal obligations, demonstrating an awareness of risks to privacy and confidentiality of data for oneself and others, and protecting and documenting data to maintain reproducibility.

Chooses Appropriate Methods and Tools

Includes selecting methods and tools that align to purposes, demonstrating an understanding of validation, and documenting methods and tools.

Analyzes Data

Involves developing an analysis plan and conducting exploratory analyses (to identify anomalies, outliers, and bias in sampling), drawing insights from results, evaluating confidence in results and comparing the results with other findings, and documenting findings.

Communicates Findings

Includes telling the "data story"—describing the problem, method and analysis, and limitations of the study, and preparing visualizations to communicate results and recommendations.

Engages in Lifelong Learning

Involves seeking out mentors and mentoring others, and staying current on new data sources and analytical tools.

Section 2
Key Skills Associated with Data Literacy

To identify the skills possessed by a data-literate individual, the expert panel turned again to the BDES profile. The BDES profile includes a list of skills considered to be essential to enabling a BDES to perform effectively. Each panel member was asked to review this list of skills and to identify the six skills they believed were most essential to data literacy.

The skills and habits identified as most essential by the panel include (in alphabetical order):

- Adheres to ethical and legal rules/implications/ consequences
- Creates/enters into a data structure
- Formulates productive questions
- Is cognizant of data security
- Is competent in data visualization design
- Is familiar with software
- Is knowledgeable about research methods
- Is knowledgeable about statistical thinking/methods
- Makes inferences (from non-random sampled data)
- Makes privacy choices
- Manipulates data
- Solves problems
- Thinks analytically
- Thinks computationally
- Thinks critically
- Thinks synthetically (big picture); creates contextual data implications
- Understands design thinking/data life cycle
- Works with spreadsheets

Section 3
Analytical Thinking

The BDES expert panel in 2014 developed a list of the skills essential to their profession. Nearly 100 big data analysts who completed the validation survey affirmed this list, giving their strongest endorsement to the skill of analytical thinking, as shown in the top survey results below.

Knowledge:
An effective big-data-enabled specialist has knowledge in:

Skills:
An effective big-data-enabled specialist possesses skills in:
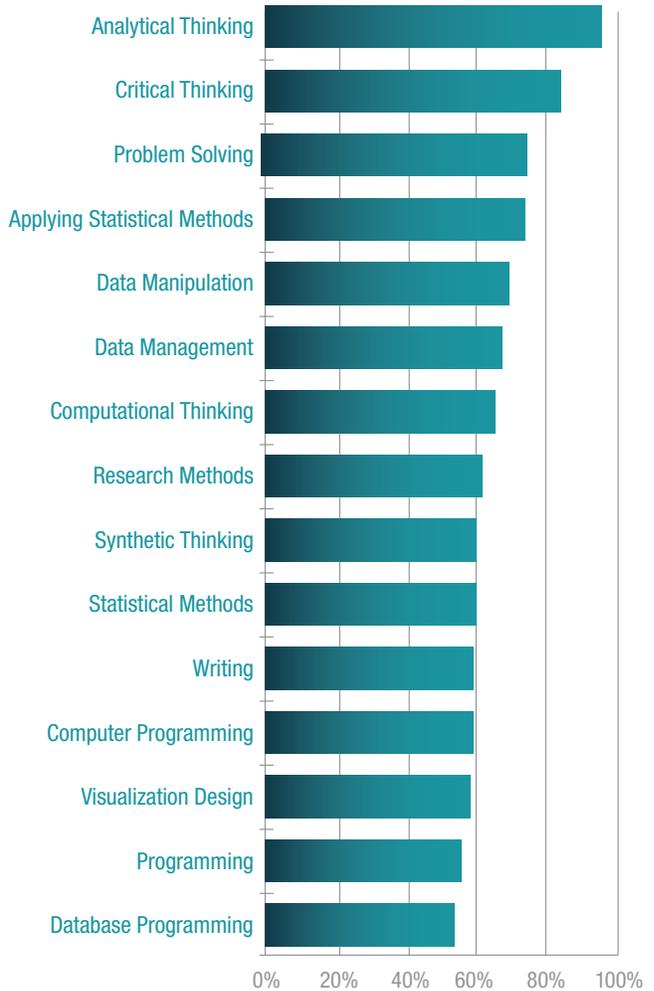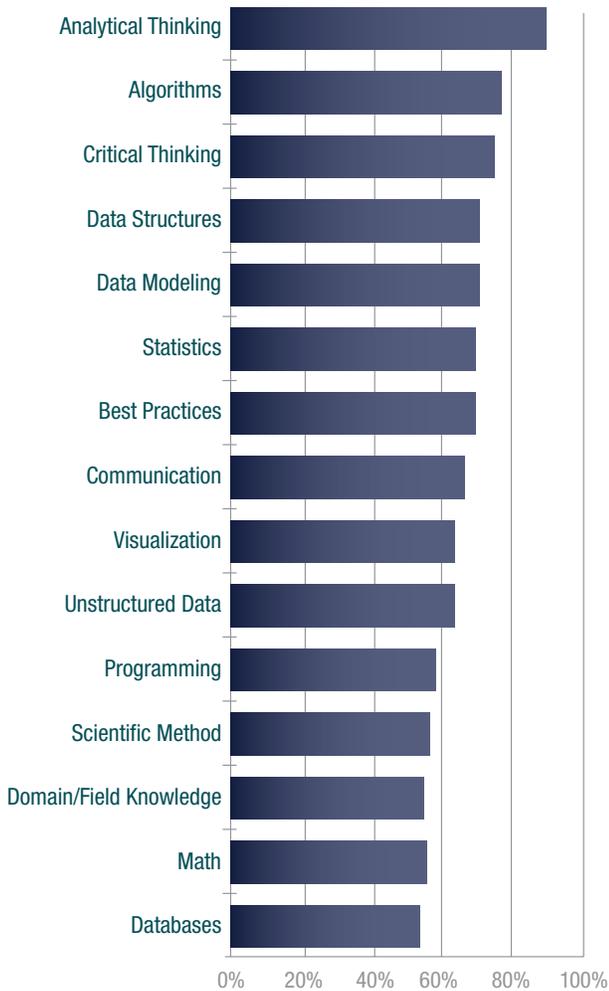
Percent of respondents who checked this item (out of N=93)

Given this strong endorsement, educators at ODI were left wondering what was meant by "analytical thinking"? How does the analytical thinker approach working with data, and how might analytical thinking skills be engendered in K–16 students?

The convening of our expert panel in 2015 provided a unique opportunity for education leaders to observe five panel members who work in data analytics as they were asked to describe how they applied analytical thinking in their work. The "fishbowl" conversation was prompted by key questions asked by the facilitator:

- What is analytical thinking within the context of your job?
- Can you provide an example of how you have used analytical thinking in your work?

After listening to the conversation among data analysts, educators were invited to ask clarifying questions and reflect on the implications for education. The entire conversation was recorded, transcribed, and analyzed, providing the three key elements of analytical thinking:

Analytical thinking is…

Clearly defining your goals and how they can be achieved with data

Taking a (sometimes) difficult problem, breaking it into pieces, and building it back up again to gain interesting insights

Exploring patterns in data with a skeptical but open mind
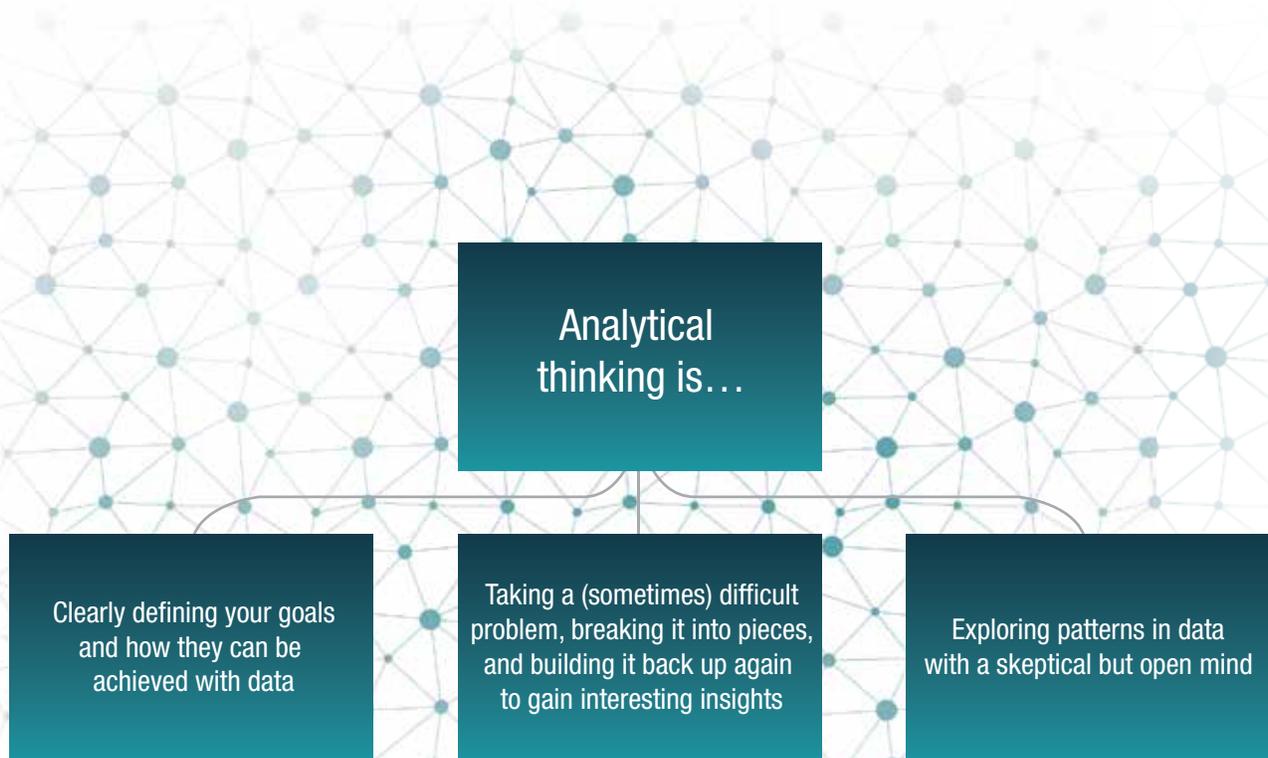
The discussion provided many different yet related perspectives, some of which are reflected below.

It's recognizing problems that can be solved with data and analytics, then thinking about both the data and the analytical approaches that would be applied.

It's persisting and recognizing that data are messy. It's about calibrations—there's a baseline you have to establish, and then you can look for the signal beyond that.

It's detective work—recognizing something is going on/an anomaly that warrants further investigation.

It's a combination of looking for patterns, and using reasoning to make sense of those patterns.

What is analytical thinking?
(The Discussion)

It's running lots of controlled experiments to test your ideas (most of which will fail), and avoiding confirmation bias. Experts have to unlearn some things and maintain a "beginner's mind."

It's curiosity—seeking answers, questioning the data. Do I have good data? Bad data? Taking information and being able to tell a story in a really specific, concise, truthful, ethical way.

It's trying different approaches.

It's recognizing that each type of data is appropriate for some uses and not for others. No data set is perfect—all have a bias, limitations.

Not analytical thinking = doing a Google search, getting a result and posting it. The analytical thinker gives it deeper thought.

The experts in data analytics on the panel shared with us a number of stories about how they've applied analytical thinking in their work to achieve important insights.

Ryan Kapaun, crime analyst for the Eden Prairie, Minnesota, Police Department, shared this example…

When I was first made an analyst for my police department, I had no training in data science or statistics. I was the first analyst, and there was no mentor to guide me. They just gave me access to 10 different databases with millions of pieces of data and asked me to tell them where the squad cars should be, and why. I didn't accomplish much at first, but my curiosity drove me—I had all of this data in my hands and I wanted to make sense of it and to tell a story. Can I forecast when break-ins into cars are going to happen and in what areas of the city? Can I also use the data to try to explain who is responsible? Eventually I was able to do that by looking at indicators of, for example, who's pawning specific items, and probing why they might be pawning them—have they been arrested five times for drugs and need to fuel their drug habit? Are they out of treatment and living in the vicinity of the break-in? But then this also brings up ethical questions—how do you balance the potential of data to do good versus invade someone's privacy?

There was a string of residential burglaries—of people who were burglarized while they were at the theater, generally on Friday nights. Residential burglaries like that are typically rare, because people are usually home on Friday night, or you might have a babysitter. How did this individual know people were gone and their houses were empty? We went through all sorts of data to try to figure out who this individual was—was it someone going through theater employee lists? Or lists of people buying tickets? Who had access to the ticketing data? That wasn't it. Who has a history of similar crimes in the metro area? Are those people out of prison? Are they in? Do they have warrants? Then we have a specific car—a green Chrysler Concord between 2000 and 2003. Who in the Twin Cities owns Chryslers that are green in that model year, and who has a drug problem or history of burglary? It's trying to narrow down the list of possible suspects and trying to tell that story of why we believe we should be targeting this individual or these three individuals. And then ruling them out.

Kirk Borne, principal data scientist for management consulting firm Booz Allen Hamilton told this story…

A large financial services firm had customers for their products, and some of those customers were taking their business elsewhere after a certain amount of time. This business was losing money, and so I said, "Think about your customers. What might be a behavioral signal of someone about to defect?" So instead of building some sort of really complex model, they just looked at Web clicks—something as trivial as how many times the person logged into their account per week or per month or whatever. And they discovered that over a period of many years there was a sort of nominal level of inquiry into their accounts and checking their balances. But all of a sudden there was a big uptick in interaction with their account the month before they pulled their money out. They were probably doing some comparison shopping and were checking their balances, etc. And then, boom, their money was taken out. So they said "Huh, maybe this is a signal. This

is a precursor signal." And so they built a model that looked for this signal in the data. They had a control sample, and for the treated sample they applied a soft customer response to this uptick—they would send an email, say, "Hey, have you seen the new products we have?" and, "Here are answers to frequently asked questions." After the first business quarter, they estimated (based on comparison with the control sample) that they saved their business $1 billion in that first fiscal quarter alone.

Hunter Whitney, UX and data visualization consultant, author, and instructor, said this…

We look at network traffic to detect anomalies. But the bad guys are aware of this and are trying to figure out ways to create signatures that are able to communicate out, but also don't look anomalous, or look like the sort of garden-variety anomaly. So they're operating in a data environment and trying to blend into it. And so the analysts who I'm designing visualizations for are trying to find that extra little something, that extra insight and analytic ability, to see if an anomaly is really interesting and different. Does it make that cut so that we should follow up with it and go deeper, dive into the server logs and so forth? But you have to see the pattern and then have a gut feeling potentially, and reasoning, to go in there and get it.

Ben Davison, data scientist at Google, told this story…

You might notice that a number of Web searches goes up in January and say, "Ok, well, maybe we're successful with recent launches." However, this actually tends to correlate with people getting new devices over the holidays. So your basic population has changed as well, and that doesn't come

through in the basic graphs. You have to be able to think about the context in which the data was collected and recognize all of the factors that might be influencing patterns and trends that seem to emerge.

Analytical Thinking: Implications for Education

When education experts on the panel joined the discussion, they articulated the following insights regarding the implications of what they had heard for education.

Encourage curiosity, skepticism, and persistence

- Encourage curiosity—*"I have all the data in my hands. I want to make sense of it, and to tell a story."*
- Encourage questioning when something doesn't "make sense," particularly regarding the data sets and what they do/do not represent
- Push students to identify multiple possible explanations for patterns in data
- Allow students to fail; an idea that fails is part of the process of innovation
- Allow students to struggle
- Teach persistence

Teach about the process of working with data

- Teach how to formulate productive questions
- Show students how to break down the problem and show work
- Model for students how to design and think critically about controlled experiments
- Scaffold students' analytical thinking to get them going (e.g., get them thinking about the boundaries of the problem and variables)

Other implications for education

- Because of the scale of big data sets, the data have to be handled differently. Students need to learn about how to frame their questions, and to be able to evolve,

adapt, and iterate them as their analysis progresses.

- Data visualizations provide powerful ways for people to engage with data, and new kinds of data visualizations, which go beyond the graphs and tables students have traditionally worked with in school, are being developed to highlight patterns in very large data sets. Students need to learn how to read and interpret these visualizations.

- It's important for students to do things, practice, and play around with data.

- Students need to represent data in a visual format.

- Students can grapple with the potential of data to do good versus concerns about privacy.

- Learning how to communicate the data story is important, including the associated risks and limitations.

Following the fishbowl conversation, panelists worked in breakout groups that included both data analysts and experts at different levels of education—elementary, middle school, high school, and post-secondary—to describe examples of classroom activities that they felt would develop students' skills in analytical thinking. These examples reflect very preliminary ideas about the types of activities that would, over the course of students' schooling, develop their analytical thinking skills. Some of these examples are described below.

- Categorize simple objects and group them based on different properties/characteristics

- Seek out an answer with different data collection strategies and compare the results

- Tell a story using data (e.g., use data from a school-based survey to describe their school to a visitor)

- Tell a coherent and compelling story with multiple tables and charts

- Use existing models or descriptions of trends to make predictions about possible outcomes not observed in the data

- Explain and use a data distribution—look for correlations, associations, clusters

- Collect evidence to solve a mystery

- Carry out a survey of classmates and explain limitations and possible inaccuracies of the data

- Design and run a randomized experiment

- Evaluate and compare multiple data sources to answer questions (e.g., compare political poll results conducted by different organizations)
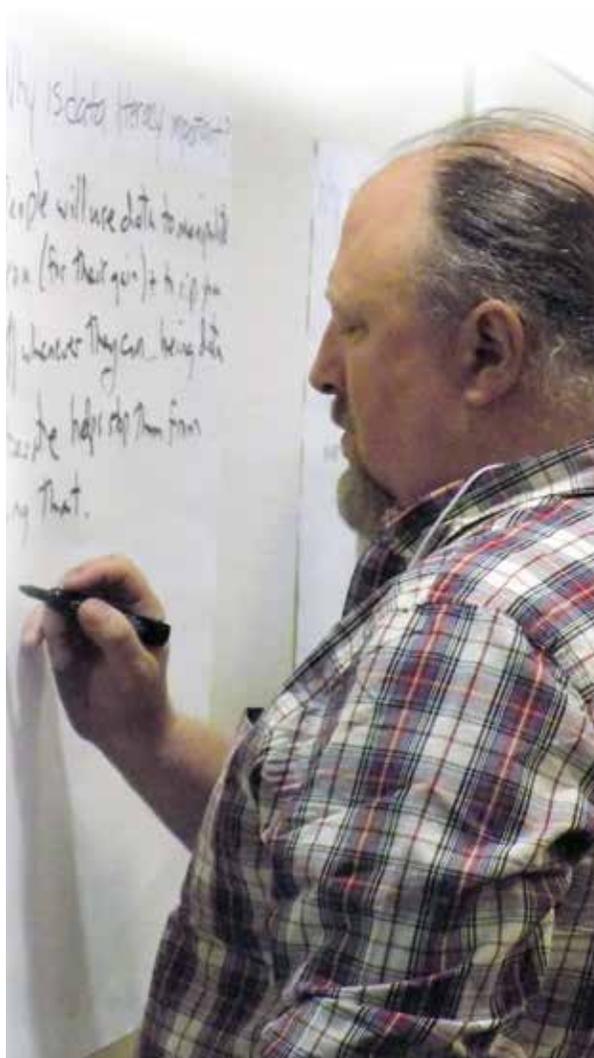
Section 4
Global Data Literacy Initiative—Next Steps

An ambitious mission, as well as the outline of a path that would get us there, was set forth by ODI, IBM, and the panel.

> ## Mission:
> Bringing data literacy to 100 million students by 2021 with the ultimate goal of everyone becoming data literate over time.

What would it take to get us there?

Engage stakeholders and build a global data literacy community

- Identify and engage key stakeholder groups whose support is needed to achieve goal (policymakers, practitioners)
  - Identify what they care about and how they can contribute to the initiative
  - Identify which stakeholders are potential funders, and what they are most likely to fund
  - Develop action plans for what we want from them and how to engage them in the initiative (e.g., endorsements, financial support, in-kind contributions)
  - Raise stakeholder awareness through events and activities (including online social engagement sites)
  - Develop memoranda of understanding with stakeholder groups
- Cultivate and endorse data literacy ambassadors/mentors

- Partner with existing relevant communities of practice
- Engage with communities of researchers in building a body of work around effective methods of developing students' data literacy

Develop/collect and disseminate educational resources

- Develop and/or translate curricula into all major languages across disciplines, STEM and non-STEM
- Develop or collect appropriate interfaces/digital analytical tools that provide student access to data sets
- Curate technology tools and platforms providing broad access to curricula, data, and teacher training
- Create and implement teacher professional development
- Develop assessment guidelines/tools
- Compile "beyond the classroom" experiences that promote data literacy
  - Research best practices and create a framework and model for data literacy competitions and hack-a-thons
  - Create/endorse competitions that engage students at various levels and scale up globally
  - Create online forums for sharing and solving data problems
  - Work with museums and science centers to develop programs that engage students in data and data visualizations/representations

Incorporate data literacy into the international standards and assessments used to define well-educated students

- Identify relevant existing national/international standards/benchmarks that can be leveraged
- Promote the inclusion of data literacy items in international assessments (e.g., PISA)
- Develop teacher certification programs
- Develop global data literacy certifications

Section 5
Call for Action to Promote Data Literacy

Recognizing the critical need to develop data literacy skills in students worldwide, the panel worked together to craft the following call for action, which was signed by all panel members and the conveners from ODI and IBM.

Data drives discovery, decision making, and innovation. The quantity of data created globally is growing exponentially, and data is everywhere. Data touches all aspects of our lives, including our health, our environment, and our role as citizens. However, our current education systems have not been equipped to produce either the workforce or the citizenry with the skills, knowledge, and judgment to make wise use of the data streams that our technologies are delivering. Specifically, we, the undersigned, see a growing urgency for the promotion of global data literacy for the following reasons:

1. Our world economy and our jobs are increasingly defined by data and by the knowledge and skills required to use them effectively.
2. We are all perpetually producing streams of data, which we need to shape and manage to ensure our privacy and personal security.
3. Effective use of data empowers us to make objective, evidence-based inferences and fundamental decisions affecting our lives, both as individuals and among societies.

In light of this, we are calling for a revolution in education, placing data literacy at its core, integrated throughout K–16 education nationwide and around the world. By enabling learners to use data more effectively, we prepare them to make better decisions and to lead more secure, better-informed, and productive lives.

Signed,

Appendix
Data Literacy Expert Panel: Biographies

Deborah Boisvert

Principal Investigator, BATEC

Deborah is the principal investigator for BATEC, an NSF-funded Advanced Technological Education National Center of Excellence aimed at creating a coordinated IT education system, spanning area secondary schools, community colleges, and four-year universities, and co-principal investigator for the Massachusetts Exploring Computer Science Partnership. Through these projects, Deborah has proven her extensive experience in developing and implementing educational programs for secondary school, community college, and university faculty while creating enduring and effective connections at the academic and administrative levels among the network of community colleges and secondary schools. Her partnerships with industry and community provide her with a broad-level perspective on regional planning for workforce training. Deborah serves in leadership positions on the Massachusetts Computing Attainment Network (MassCAN), the MA K–12 Computer Science and Digital Literacy Standards Team, the Board of Higher Education's Transfer Task Force, the National Advisory Board of IWITTS, and the ACM SIGITE Education Board.

Kirk Borne

Principal Data Scientist, Booz Allen Hamilton

Kirk is a transdisciplinary data scientist and an astrophysicist, with specific expertise in applications of data mining, machine learning, and advanced analytics within diverse domains, from marketing to education to scientific research. He is the principal data scientist in the NextGen Analytics and Data Science account within Booz Allen Hamilton's Strategic Innovation Group, where he provides advisory, consulting, mentoring, training, client solutions, and business development expertise in collaboration with the 1000+ data scientists within the organization across numerous industries. He has advised several federal agencies on data analytics and big data applications. He is a founding member of the American Astronomical Society's Astroinformatics and Astrostatistics Working Group and the International Astrostatistics and Astroinformatics professional society. Prior to moving to his current position in May 2015, Kirk was professor and data scientist at George Mason University for 12 years. Before 2003, he spent nearly two decades in positions supporting NASA projects. He received his BS degree in Physics from LSU and his PhD in Astronomy from Caltech.

Michael Bowen

Associate Professor, Science Education, Mount Saint Vincent University, Halifax, Nova Scotia, Canada

Mike has worked with developing and researching competency with data and graph literacy for over 20 years. His ethnographic PhD work examined the development of data literacy (i.e., analysis and representation of data) at various educational levels ranging from grade 6 to post-doctoral fellows and university professors to try to understand how these competencies develop. He has published numerous academic articles and chapters on these issues in education, sociology, and psychology journals. His recent focus has been on the development of these competencies in teachers, and his recent publications include two books to help develop inquiry and data analysis competencies in middle and high school teachers and their students. Mike has presented workshops on these issues at teacher professional development conferences over the past 15 years and works actively with the National Science Teachers Association. He has bachelor's degrees in Biology, Education, and Journalism (HonBSc, BEd, BJ); master's degrees in Toxicology and Sociology (MSc, MA), and a PhD in Education.

Ben Davison

Quantitative User Experience Researcher, Google Search

Ben is a data scientist in Google Search. His goal is to extract meaning from user activities in Google Search. The largest-n work involves A/B studies with billions of users on everything other than the main list of links or ads. This informs product teams and leadership on what users really want to see on the page. Ben started research with psychometric studies for accessibility in academia, where getting enough participants was a key problem. He has since discovered that big data has the reverse problem: spurious findings and results that are difficult to put into action. He found a distance in meaning not present in small-n studies. His current work in logs analysis, surveys, and user needs is triangulating on meaningful results and improving methods for product-driven research. Ben holds a PhD in Human-centered Computing from Georgia Tech.

Rob Gould

Faculty, UCLA Department of Statistics

Rob has been responsible for developing and directing the UCLA Statistics undergraduate program since 1998 and his primary interests are in statistics education. Currently, he is principal investigator of the Mobilize Project, an NSF-funded program to develop data science curriculum materials in high school STEM courses centered on participatory sensing, a data-gathering paradigm. He is currently on the American Statistical Association's joint committee with the National Council of Teachers of Mathematics and is chair of the ASA's joint committee with the Association of Mathematicians at Two-Year Colleges. He is also founder of the American Statistical Association's DataFest competition, a national data hack-a-thon for undergraduate students. In addition, he is founding editor of the e-journal *Technology Innovations in Statistics Education* and a co-author on the introductory statistics textbook *Exploring the World Through Data*. He was a collaborating author on the ASA's *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report*, published in 2007. He has a PhD in Mathematics from University of California, San Diego, and a BS in Applied Mathematics from Harvey Mudd College.

Ryan Kapaun

Crime Analyst, Eden Prairie, Minnesota, Police Department

Ryan is a crime analyst for the Eden Prairie Police Department (MN). He uses data to forecast crimes in and around Eden Prairie. After crimes occur, Ryan uses data to develop leads and locate suspects. He has 12 years of experience as an analyst, including five years as an intelligence analyst with the State of Minnesota, assigned to the Division of Homeland Security and Emergency Management, Bureau of Criminal Apprehension and the Minnesota Joint Analysis Center. He is a past president of the Minnesota Association of Criminal Intelligence Analysts and, in 2010, he was named the Minnesota Crime Analyst of the Year for his role in collecting and analyzing intelligence data for the Eden Prairie Police Department. Ryan has a BA in Communication from the University of Minnesota, Duluth.

Cliff Konold

*Director, Scientific Reasoning Research Institute,
University of Massachusetts, Amherst*

Cliff is director of the Scientific Reasoning Research Institute at the University of Massachusetts, Amherst, and senior research scientist at PRISM. A psychologist by training, he studies how people reason and learn about chance and data, and applies this research to the design of educational materials and software. He was the lead designer of TinkerPlots, the data visualization and modeling desktop application published by Learn Troop, and is currently part of the team developing the data analysis Web application, CODAP. As part of that project, he is continuing his research of students' understanding and use of data structures.

Juan Miguel Lavista Ferres

Principal Data Scientist, Bing/Microsoft

Juan Miguel Lavista Ferres is currently the principal data scientist for Microsoft Data Science team (DnA), where he works with a team of data scientists searching for insights in petabytes of data. Juan joined Microsoft in 2009 to work for the Microsoft Experimentation Platform (EXP), where he designed and ran randomized control experiments across different Microsoft properties. At Microsoft, Juan also worked as part of the Bing Data Mining team. Before joining Microsoft, Juan was the CTO and co-founder of alerts.com. Juan has two Computer Science degrees from the Catholic University in Uruguay, and a graduate degree in Data Mining from Johns Hopkins University.

Odette Merchant

Project Manager, Nova Scotia Community College (NSCC), Halifax, Nova Scotia, Canada

Odette is an experienced educator who has worked for more than 20 years in the adult education sector in both Canada and the United Kingdom. Odette has held various teaching and leadership roles within the field and is now project manager at NSCC supporting educational programs and partnerships with industry. Her work in post-secondary education has always focused on the development of pathways to learning, skills development, and applied learning. A true advocate of lifelong learning, Odette strives to support strategic collaboration efforts that bring together educational institutions and the community for the enhancement of both. Currently, Odette is working on a multi-institutional project in Nova Scotia that is bringing together high schools, community colleges, universities, government agencies, and industry partners in developing learning opportunities, pathways, curricula, and applied research around data and analytics. Odette holds a BA from Memorial University of Newfoundland, an MBA from the University of Glamorgan (University of South Wales), a Diploma in Community College Education from NSCC, and is currently studying for her MEd in Lifelong Learning at Mount Saint Vincent University.

Andrew Schaffner

Professor of Statistics, California Polytechnic State University, San Luis Obispo

Andrew is very interested in curriculum development as well as increasing statistical literacy and computing across the undergraduate curriculum. He introduced statistical computing using R to the Cal Poly BS Statistics program in 1997, which is now required in the curriculum. He also served as the university-wide curriculum chair from 2010–15. This year he will serve as the internal reviewer of Cal Poly's General Education program, where he will stress the importance of coursework to improve data literacy. He is co-principal investigator and principal statistician for a number of NIH-funded clinical trials investigating obesity and pregnancy as well as consulting statistician for multiple environmental projects funded through the National Estuary Program in Morro Bay and Tenera Environmental. He holds a BS in Mathematics from Cal Poly and an MS and PhD in Statistics from the University of Washington.

Hunter Whitney

Consultant, Author, and Instructor; UX and Data Visualization

Hunter is a consultant, author, and instructor who brings a user experience (UX) design perspective to data visualization. He has advised corporations, start-ups, government agencies, and NGOs to achieve their goals through a strategic design approach to digital products and services. His experience includes leading the designs of data analysis interfaces for uses ranging from biomedical research to cybersecurity. He has also consulted for the Monterey Bay Aquarium on various UX and visualization projects. Hunter is the author of *Data Insights: New Ways to Visualize and Make Sense of Data* and contributed a chapter to *Designing for Emerging Technologies: UX for Genomics, Robotics, and the Internet of Things.* He received dual bachelor's degrees in Biology (UCSC) and English Literature (UCLA)—and has completed post-graduate psychoneuroimmunology research at UCLA.

Michelle Williams

CEO, Williams Learning Solutions

Michelle is CEO of Williams Learning Solutions, a start-up company. Her focus is developing educational tools for elementary science. Michelle worked as an associate professor of science education at Michigan State University for almost 10 years. During that period she developed projects designed to determine effective approaches to successfully deliver science education to K–12 students. Her work combined topics in the biological sciences using a technology-enhanced learning environment to take advantage of research in the learning sciences that improves scientific literacy among precollege students. She received several million dollars in National Science Foundation funding to address the question, How can all students be assured the opportunity to learn significant STEM content? Through this initiative, Michelle investigated best practices in technology-supported instructional design and formative assessment that provide rich data on students' conceptual understanding of genetics. Michelle holds a BBA in Marketing from the University of Texas, Austin, and a PhD in Education in Development in Mathematics and Science from UC Berkeley.

EDC OCEANS *of* DATA
INSTITUTE