DATA DETECTIVES

BIG DATA CLASSROOM PROGRAM FOR TEACHERS AND THEIR STUDENTS

SUBJECT An introductory session on Big Data

AUDIENCE Students (ages 13-18)

TIME REQUIREMENT 45-55 minutes

ACTIVITY COMPONENTS Interactive survey, discussion, reading, and Q&A

REQUIREMENTS Computers with online access

GOAL Introduce students in a lively, fun way to the topic of Big Data

About Data Detectives and The Human Face of Big Data Project

Imagine students from all over the world sharing and comparing their lives in real-time via the Internet.

Data Detectives, a student component to the parent project The Human Face of Big Data, is for teenagers around the globe (ages 13-18), designed to bring the topic of Big Data to life through a customized, interactive, online experience.

The Human Face of Big Data is designed to demystify and humanize the topic of Big Data. Earlier this the year, 100 of the world's leading photographers were dispatched to over 30 countries to capture stories and images that are presented in The Human Face of Big Data project. The stories, photographs, and illustrations capture the effect that Big Data is beginning to have on civilization in every corner of the globe.

About the Classroom Teaching Program

The world's data is now doubling every two years, so to help students understand the impact of this data deluge this curriculum is aimed at demystifying and answering the question: "What is Big Data?" It's an easy-to-follow, 45-55 minute classroom activity for educators consisting of discussion, reading, Q&A, and an online, interactive survey with real-time results. Students are invited to share and compare aspects of their lives in real-time, to explore stories and images from the book, and learn about how students around the globe are already utilizing Big Data.

The Classroom Program Components

The Data Detectives classroom program consists of the following components:

- Program flow and guide
- Discussion and Q&A outline
- An online survey (students will need computers with Internet access)
- Data visualizations and comparisons
- Big Data essays, infographics, and additional resources

The Program Inspires

The Data Detectives curriculum session inspires the following:

a. Gives the ability for students to measure, compare, and learn about themselves

b. Offers a fun way for students to engage with data and experience how data affects their lives

c. Demonstrates the ability to apply data filters and understand behavioral comparisons

d. Shows students how data can give them new ways to look at and understand the world

The Data Detectives Big Data Classroom Program and Guide

PART 1 ICEBREAKER/TEACHER STARTS DISCUSSION (5-10 MINUTES)

Designed as a quick icebreaker to show what students currently believe Big Data to be.

a. Without advance discussion and/or prompting, teachers ask students, "What is Big Data?"

PART 2 STUDENTS READ A SUMMARY ABOUT BIG DATA (5-10 MINUTES)

a. Teacher hands out a summary article on Big Data ("Reflections in a Digital Mirror" by Juan Enriquez) for students to read. The essay can be found at the bottom of PDF file and is available at HumanFaceOfBigData.com/teacher-resources.

PART 3 STUDENTS PARTICIPATE IN DATA DETECTIVES WEB EXPERIENCE (10 MINUTES)

Students are prompted to take the Data Detective survey. By participating in the survey, students will experience first-hand data gathering, instant analysis, and filtering of their answers. They'll be able to see themselves compared to their classmates, and to other students from around the world, in real-time.

a. Students take online survey (each student will need their own computer with Internet access)

b. Survey results are instantly transformed into compelling, rich visuals

c. Results are viewed, compared, and discussed within the classroom

d. Students and teachers can print survey results and post in their classrooms

Here are a few things to consider regarding taking the survey. (Keep in mind, there are several different ways the teacher and students can interact with the survey activity.)

a. Ideally, each student has their own computer, otherwise students will need to share (this will lengthen the time required for taking the survey)

b. Each student will have their individual results from the survey to review

c. For purposes of a group discussion, we recommend the teacher project one of the student's results onto a large screen or board to instigate a general discussion. If this isn't possible, we recommend students form small discussion groups to discuss the findings

Part 4, below, is a Q&A discussion guide

PART 4 DISCUSSION AND Q&A (15-20 MINUTES)

Now that the students have explored the Data Detectives website, teacher can engage students in a discussion and Q&A on their experience. To facilitate a discussion, teachers can ask students the following questions:

a. What are the key take-aways from the essay?

b. What are the key take-aways from your web experience?

c. Based on reading the essay and the web experience, how is Big Data different from what you thought? (Refer to the Big Data Facts, below, to guide Q&A)

d. Do you think in the future Big Data will be more or less important than the Internet? Why?

e. How much data do you think is collected about you every day? (Refer to infographics and ideally project them onto a screen to make the conversation visually interesting and engaging)

f. How is Big Data changing and/or impacting your daily life?

g. Are there ways you could interact with data in your daily activities that would be helpful?

h. How do you envision Big Data and technology being used to help solve problems facing the world and humanity?

i. How do you think Big Data could be utilized to solve a problem in your school or community? Refer to:

   i. "Data Making a Difference" and the "Kids Making a Difference" for story examples

   ii. **The Human Face of Big Data** website

   iii. Stories, essays, and infographics from the book (identified below)

Reference Materials & Description of Big Data

What is Big Data?

"It is the accumulation and analysis of information. Lots of information. Oceans of information. Every time someone clicks on something at Amazon, it's recorded and another drop of is added to the ocean. Every time a scanner beeps at the supermarket checkout. Every time a home electricity meter reports a reading. Every time a parcel passes a FedEx checkpoint. Every time a customs officer checks a passport, every time someone posts to Facebook, every time someone does a Google search — the ocean swells."

—Dan Gardner, "An Ocean of Data"
(excerpted from The Human Face of Big Data)

Big Data is shaping businesses, shifting markets, and transforming our world. During the first day of a baby's life, the amount of data generated by humanity is equivalent to 70 times the information contained in the Library of Congress.

Big Data is a simple term used to describe the emergence of incredibly powerful ways to gather and analyze digital information to gain new insights about nearly every aspect of our world and lives.

It is the ability to extract meaning: to sort through huge masses of numbers and find the hidden patterns, the unexpected correlations, and the surprising connections.

You can think about Big Data as the process of helping the planet grow a nervous system, one in which we are just another human-type of sensor.

What's Creating the Deluge of Information?

Each of us now leaves a trail of digital exhaust, an infinite stream of phone records, texts, browser histories, GPS data, and other information that will live on forever. This data comes from:

- Sensors on mobile phones
- Online shopping
- Medical imaging
- GPS-enabled cameras and smartphones
- Satellites
- Smart electrical grids
- Video surveillance
- Social media: Facebook, Twitter, Google
- Digital music
- Digital photographs
- Large databases in the banking sector
- Gene sequencing

How to Measure Big Data and What's a Byte?

(A byte reminder – each one is 1,000 times larger than the one before it!)

- A byte = one character, or a grain of sand
- A kilobyte = a sentence, or a couple of pinches of sand
- A megabyte = a 20-slide PowerPoint show, a small book, or a tablespoon of sand
- A gigabyte = 10 yards of books on a shelf, or a shoebox full of sand
- A terabyte = 300 hours of good-quality video, a tenth of the Library of Congress, or a playground sandbox
- A petabyte = 350,000 digital pictures, or a mile-long stretch of beach
- An exabyte = half the information generated worldwide in 1999, or a beach stretching from Maine to North Carolina
- A zettabyte = unimaginable, or a beach as big as all the coastlines in the world

Big Data Facts and Stats

- The worlds data doubles every two years
- In 2011, humans created 1.8 zettabytes of data. This is equivalent to 200 billion high-definition movies that are at least 120-minutes long. It would take one person 47 million years to watch all those movies
- YouTube users upload 48 hours of new video every minute of the day
- 20 petabytes of data are processed daily (that's 20 billion megabytes)
- There are one billion Tweets created every 72 hours
- 30 billion pieces of content are shared on Facebook monthly
- A personal computer holds about 500 gigs, so it would require about 20 billion PCs to store all of the world's data
- $600 buys a disk drive that can store all of the world's music
- Lady Gaga has 28 million Twitter followers. Barack Obama has 18 million
- Wal-Mart handles more than one million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes of data

- More than five billion people are calling, texting, tweeting, and browsing on mobile phones worldwide
- Decoding the human genome originally took 10 years to process; now it can be achieved in one week
- 571 new websites are created every minute of the day

Brands and organizations on Facebook receive 34,722 "Likes" every minute of the day

Today, a street fruit vendor in Mumbai can access more information, maps, statistics, academic papers, price trends, futures markets, and data than a U.S. president could only a few decades ago

Examples of Big Data

The following are summary examples of how Big Data is being utilized:

"Instead of 'find my iPhone,' some auto insurance companies are offering a service that may enable parents to 'find my teenager.' Progressive Insurance, for example, offers the Snapshot, a tracking device that reports on a car's location, acceleration, braking, and distance traveled. Owners who install the device can get a 10 to 15 percent discount on their policy. Privacy activists, however, fear the technology is ripe for abuse."

—Michael Malone
(excerpted from The Human Face of Big Data)

"Relying on social networks and analytics, companies are gathering volumes of data from the web to help musicians and music companies better understand their audiences. By tracking hundreds of thousands of artists' new fans, plays, views, and comments across social media, and then correlating them with events such as song releases or shows, a company called Next Big Sound can identify a band's popularity on the web before record sales rise. Next Big Sound's technology singled out Alabama Shakes, headed by lead singer Brittany Howard, as one of the 15 fastest-accelerating artists across the Internet, even before the band's shows sold out and before they appeared on the Late Show with David Letterman."

—Michael Malone
(excerpted from The Human Face of Big Data)

Essays

Learn more about Big Data through these insightful and illuminating essays. You won't think about the world in the same way again.

a. "Reflections in a Digital Mirror" by Juan Enriquez

b. "An Ocean of Data" by Dan Gardner

c. "Our Data, Ourselves" by Kate Greene

d. "How Crowdsourcing Is Changing Science" by Gareth Cook

Infographics

View the world of Big Data through the colorful interpretation of powerful infographics. Google, Twitter, Facebook, and more come to life in the following infographics. Click the links (below):

a. "Googling Google – What Happens When You Google?" by Nigel Holmes

b. "Facebook: Taking Our Emotional Temperature" by Nigel Holmes

c. "The World Speaks Up! – Twitter" by Nigel Holmes

Additional Resources

Enjoy more resources and learn about cool projects by going to the following links:

a. Nick Feltron - **http://feltron.com/**

b. Aaron Koblin - **http://www.ted.com/talks/lang/en/aaron_koblin.html**

c. **MIT SENSEable City Lab -** http://senseable.mit.edu/trashtrack/trashtrack_release.mov

d. Jer Thorp - **http://blog.blprnt.com/about**

e. Spatial Information Design Lab - **http://www.spatialinformationdesignlab.org/projects.php?id=16**

About the Sponsors

The Data Detectives component of The Human Face of Big Data project is editorially independent and is made possible through the generous support of EMC Corporation, which serves as the primary sponsor. Supporting sponsorship comes from Cisco Systems, FedEx, VMWare, Tableau, and Originate.

EMC's primary interest and expertise lies in advancing science, technology, engineering, and math (STEM) education to build a highly-trained and innovative workforce for the future. EMC's efforts are focused on strengthening educational systems and raising standards locally and globally. "The demands of life, work, and citizenship cannot be met in through a public education system that has remained virtually unchanged for more than century. The time for fundamental, systemic change is now."
—Joe Tucci, President,
CEO, and Chairman of the Board of EMC Corporation

Each year Cisco Networking Academy, Cisco's largest corporate social responsibility program, teaches over a million students worldwide the skills needed to build, design, manage, and secure computer networks. Creating economic opportunities by improving student career prospects while filling the global demand for networking professionals. With 10,000 academies in 165 countries, Networking Academy helps individuals prepare for industry-recognized certifications and entry-level information and communication technology (ICT) careers in virtually every type of industry. Students develop foundational skills in ICT while acquiring vital 21st-century career skills in problem solving, collaboration, and critical thinking.
cisco.com/go/netacad/us.

FedEx is proud to be the Official Shipping Sponsor of The Human Face of Big Data. In 1978, FedEx chairman and CEO Fred Smith said, "The information about the package is just as important as the package itself." Today, FedEx provides customers with access to near real-time information that has enabled new supply chain models and efficiencies. This unprecedented access to information is connecting customers around the world to economic markets and communities. FedEx, the company that revolutionized global business practices by providing next- business-day delivery service, leveraged the reach, strength, and reliability of its network to ship 10,000 individual copies of The Human Face of Big Data photo book around the globe. From the earliest commercial handhelds in 1980, to equipping vehicles with radio computers in 1984, to making package tracking available on the world's first transactional website in 1994, FedEx has always sought the edge of innovation.

For More Information

To learn more about The Human Face of Big Data and its Student Data Detectives initiative, teachers, students and parents should visit the program Web site at StudentFaceOfBigData.com. To receive more information regarding he EMC Learning Solutions commitment, interested school administrators should go to emc.com/education.

About The Human Face of Big Data

The Human Face of Big Data is a globally crowdsourced media project focusing on the new ability to collect, analyze, triangulate, and visualize vast amounts of data in real-time. The project includes a free mobile app, Mission Control briefing events, a student Data Detectives initiative, a large-format book (with more than 200 photographs, as well as essays and infographics), an iPad app, and a documentary.

The Human Face of Big Data is produced by Rick Smolan and Jennifer Erwitt, co-founders of Against All Odds Productions, which specializes in the design and execution of large-scale, crowdsourced global projects that combine compelling storytelling with state-of-the-art technology. To produce these projects, Smolan and Erwitt invite teams of leading journalists and photographers (plus members of the general public) to come together to focus on emerging topics such as the global water crisis, the effect of the Internet on civilization, and how the human race is learning to heal itself. Their projects result in best-selling illustrated books, TV specials, apps, exhibits, and hundreds of millions of media impressions. Fortune magazine described Against All Odds Productions as "one of the coolest companies in America."

To learn more, go to HumanFaceOfBigData.com and follow the project on Twitter and Facebook.

REFLECTIONS IN A DIGITAL MIRROR

Juan Enriquez

MOST MODERN HUMANS are now attempting to cram more data into their heads in a single day than most of our ancestors did during entire lifetimes. In the 15 minutes it takes you to read this essay, the amount of information being generated by the human race will have expanded by about 20 petabytes, equivalent to about three times the amount of information currently in the Library of Congress, or about one-half of all written works from the beginning of recorded history in all languages. The world's total data is doubling every two years.

It's not just what humanity is collectively generating that's overwhelming us; it's what we, as individuals, attempt to digest daily. Every year we try to cram in, read, understand, and remember at least 5 percent more words than the year before. That means that instead of coping with a mere 100,000 words per day five years ago, we are now coping with more than 130,000—plus billions of compounding bits. Even what used to be the calming act of looking at the stars has been transformed: within weeks of its launch, the Sloan Digital Sky Survey satellite collected more astronomical data than had all of mankind in its entire history. But even this amount of data seems a trifle when compared with the daily output of the new Large Hadron Collider (LHC) in its successful hunt for the "God particle" and the origins of the universe. And yet, as unbelievable as these achievements seem, the reality is that we are actually just at the beginning stages of the Big Data era.

We are shifting from a world in which we "know," because we sampled a little and extrapolated a lot, into a world in which we *know*—where all data is collected, analyzed, and stored. We are all becoming citizens of the new realm of truly Big Data. And it will soon touch not only every *corner* of our lives but every *time* in our lives: for three years, MIT's Deb Roy recorded over 250,000 hours of video and audio of his baby son's life—categorizing and analyzing nearly every word heard by and, eventually, spoken by his son, providing an accurate map of how one learns to speak and reason.

As we transition from sampling and polling to having a complete census of enormous data sets, we also transition from "I think I'm sure" to "I know it, and I can prove it." For instance, Google has created a census of every word published since 1500. We now know exactly who used more than 500 billion words, in more than five million books. And we can trace the specific use, frequency, and context of every one of those words and phrases. Thus we know the words *love* and *war* battled for frequency of usage from 1800 through 1914. (Love usually triumphed.) Since 1914, war has been the overwhelming victor. And the word sex? For better or worse, it's been steadily gaining on the ever-declining use of love.

All of this is taking place within a massive and explosive evolution in how we use, store, and transmit information. In 1986, only 6 percent of the world's data was digital and "www" was still three years away. There was no Google. Today, more than 99 percent of the world's written words, images, music, and data are transmitted in the two-letter Boolean alphabet of 1s and 0s. Other than perhaps the agricultural revolution of 10,000 years ago, no event in human history has ever generated as much wealth and changed as many lives as this transition into a digital world.

According to the global market intelligence firm IDC, in 2011 we played, swam, wallowed, and drowned in 1.8 zettabytes of data. (A zettabyte is a trillion gigabytes; that's a 1 with 21 zeros trailing behind it.) IDC Digital notes that if you were inclined to store this data on 32-gigabyte iPads, you would need only 86 billion devices—just enough to erect a 90-foot-high wall 4,000 miles long from the bottom of your shoes to the center of the Earth. Today, a street fruit stall in Mumbai can access more information, maps, statistics, academic papers, price trends, futures markets, and data than a U.S. president could only a few decades ago.

Lest all this seem a little abstract and distant from your daily life … then imagine Facebook, Google, Twitter, and other social media as electronic tattoos. They are very different from yesterday's ink tattoos. At once trivially easy to apply as well as seemingly painless, these new tattoos can also be far longer lasting and potentially more damning. Every time we blog, tweet, Facebook, or Google, or visit Amazon, LinkedIn, Meetup, or foursquare, or upload a video to YouTube, we leave little marks, some more visible than others, of who we are, whom we are with, and what we like. We electronically tattoo ourselves, our preferences, our lives, in a far more comprehensive and nuanced way than any inked skin.

Mostly self-designed and self-inflicted, electronic tattoos are so easy to copy, reproduce, spread, store, and retrieve that they will likely long outlive our bodies. In a very new way, these brandings will make us immortal tomorrow … and inescapable today. Tattoos are serious. Every parent knows this; most kids do not. Once inked, a tattoo is a lifelong commitment to a culture, cause, person, passion, hatred, or love. Once inked, there is no hiding, and it is hard,

if not impossible, to change sides. Tattoos publicly advertise membership, fidelity, dedication, love, hate, and—often—stupidity. Beware, Andy Warhol's 15 minutes of fame may now turn into an unbearable eternity.

Immortality and Big Data are linked in other ways as well. One of the fastest-growing, and most interesting, aspects of Big Data is at the intersection of the digital and the biological. We've come very far, very quickly: a decade ago, reading the code of a single genome was a historic breakthrough. This year we will decode thousands of individual genomes. When Steve Jobs had the genes of his cancer tumor and of his normal DNA sequenced a few years ago, it was done in a specialized lab and cost more than $100,000. Today the cost of DNA analysis has plummeted to under $10,000, and with a simple spit test sent to a lab and $299, you can have your DNA analyzed for your health and ancestry. As we standardize, even trivialize, the coding of life—increasingly not just gene by gene, but entire genomes at a time—we will inevitably also change every carbon-based industry, including energy, medicine, chemicals, biotech, and agriculture. This may seem an outlandish claim, but imagine your reaction had someone come to you in the late 1970s or early '80s and told you that your work, play, education, entertainment, and citizenship would be forever transformed by a worldwide web of invisible, incredibly fast wireless digital networks connected to devices smaller than your wallet or perhaps implanted under your skin.

Having the data is only the beginning. As we know more and more about life, the universe, and all other subjects, as we double the amount of data generated by all humans within the next five years, we can begin to model, build, and scale to the point where we directly and deliberately guide the evolution of ourselves and many other species. In 2010, three scientists—Craig Venter, Hamilton Smith, and John Glass—programmed a computer with a basic genetic sequence. Then robotic arms assembled this specific code using the four building blocks of DNA (adenine, thymine, guanine, cytosine). It was a little like building a very complicated, very tiny Lego-like structure. By the end they had constructed the world's largest organic molecule. After discovering how to get this life program into a cell, the trio found they could turn one species into a different species. Some called it the world's first synthetic life form, but it is really the first fully programmable life form.

As programmable cell platforms like these begin to act, and be programmed, like computer chips, we can make fuels or chemicals, absorb excess carbon, develop miniature vaccine factories, change the growth cycles of plants, regrow organs, and extend the human life span substantially. As we bring together Big Data and life sciences, this will drive further discovery, as well as the world's economy, which in turn will drive unimaginable amounts of new data. MIT's Sebastian Seung, who creates 3D images of mouse brains, estimates that moving from current MRI machine resolution to light microscopy resolution will require increasing storage, per brain image, from one megabyte to a petabyte. But for us to truly understand the connections between neurons, we will need to use an electron microscope, meaning that each of our brains will generate a zettabyte-scale file.

So when we talk about Big Data and where it's headed, remember: all of humanity, thus far, has generated 1.8 zettabytes of data in history—roughly the equivalent of the image of an entire human brain.

We are barely at the cusp of the Big Data–life sciences revolution, and we are already running out—of bandwidth, of storage, of space. Every day FedEx trucks arrive at facilities around the globe to deliver hard drives by sneakernet, because it's faster and cheaper to ship a big hunk of iron from China or Bangalore to Silicon Valley than to stream that data through painfully clogged pipes. Even the cloud, as we know it, could be overwhelmed by life sciences data accumulating 50 percent faster than we can store it. These challenges will, in turn, breed enormous new companies and breakthroughs.

Big Data started as a series of small waves but is morphing into the greatest tsunami of information that humans have ever seen. What we choose to do with all of this new data may lead to one of the biggest adventures of all time.

Juan Enriquez is the managing director of Excel Venture Management and the author of the global bestseller As the Future Catches You: How Genomics and Other Forces Are Changing Your Life, Work, Health, and Wealth. He has helped found or guide more than a dozen technology start-ups, including Synthetic Genomics, Zipcar, Xcellerex, and Activate Networks. Prior to founding Excel, he was the founding director of the Harvard Business School's Life Sciences Project.