MindCET Snapshot #2

# Big Data & Education

February 2014

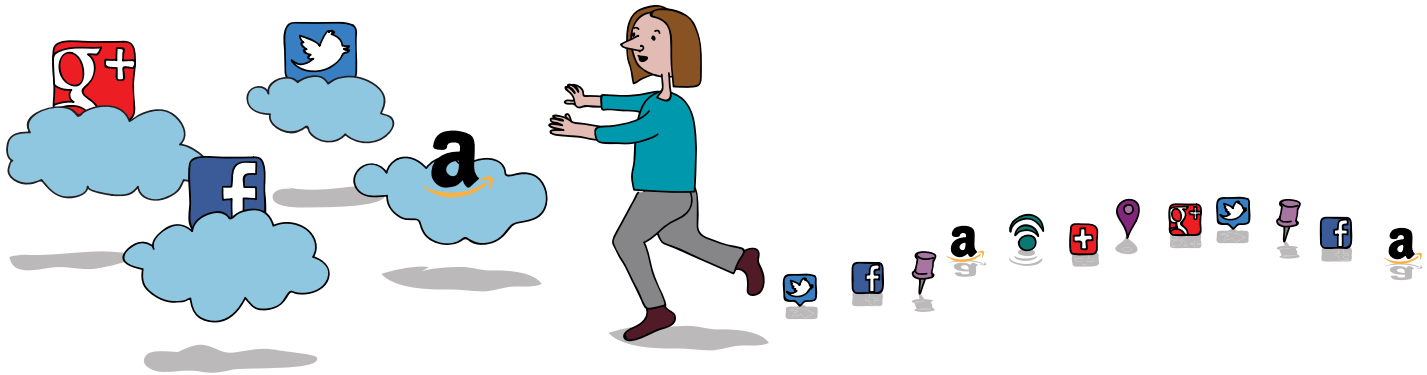
MindCET Snapshots highlight current trends within the developing field of EdTech, providing different perspectives: pedagogic, technological, business, and so on. This issue deals with Big Data's impact on education.

# Table of Content

# Overview

# Overview

**Whether we are aware it or not, our lives are increasingly being affected by data-driven decisions.**



Every time we type, click, touch, look – in other words, relate to digital devices – data is being gathered, and saved; we are continuously leaving a trail behind which is tracked, analyzed and even traded. Somewhere, a meaningful image of us is being formed which in turn allows a device to get to know each one of us, apart from the crowd. This familiar relationship with digital devices, through social networks, search engines, shopping sites and mobile apps, provide us with a sense of efficiency that we cannot live without. GPS helps me get home from anywhere, Facebook newsfeed selects the posts of my closest friends, Google facilitates my searches selecting out irrelevant stuff; devices that "know" me facilitate my life. In order for this to happen, we constantly feed, mostly passively, seas of data that are turned into meaningful information to create our individual and common digital identities. **These ever-growing, intelligent systems of understanding data are being developed to help us make sense of who we are and about the world we live in. This phenomenon is what surrounds the media phrase - Big Data.**

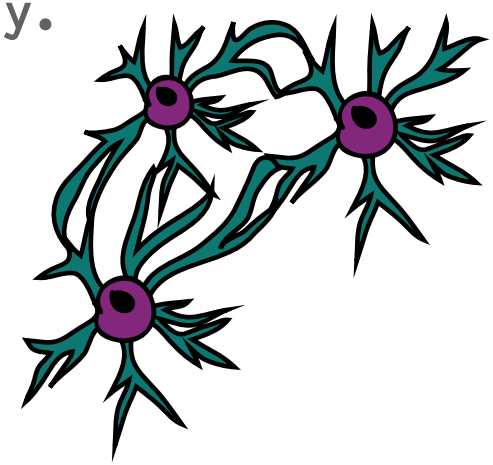
# An "elusive" concept

The overwhelming rise of allusions to Big Data throughout the media affirms its undeniable impact on all areas of society, from business to the academy. However, it is still hard to find a clear and unified definition. According to the MIT Technology Review (October 2013), **Big Data is revolutionizing 21st-century business without anybody knowing what it actually means.** To make it even gloomier, volume is losing value: "Big Data is one of the worst industry terms ever invented. Not only does it poorly describe the increasingly role data plays in our lives, but it has created an obsession with the exact wrong parameter: volume of data."[1]

Tim Smith animates this "elusive" concept and talks about the uncomfortable ever-growing digital information that has been challenging us, throughout the last five decades, to create new means to store, connect and analyze data. Smith reminds us of the major impact on the Big Data development of CERN Accelerating Science Lab, where the World Wide Web was developed by British scientist Tim Berners-Lee.

At Cornell University, Ward & Baker have recently published a paper trying to arrive at a common definition, based on a survey of how Big Data is perceived by the leading companies, and concluded that "Big data is a term describing the storage and analysis of large and complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning."[2]

To help us get a clearer picture, we looked at what a few of the leading players have to say.



"Big data - the lifeblood of our new global nervous system - is the resource for addressing the big global challenges of today. Leveraging the ever-increasing power of networked computers, big data provides the clearest lens for examining how society functions in fine-grain detail"

(Alex 'Sandy' Pentland, Toshiba Professor of Media, Arts and Sciences, Massachusetts Institute of Technology).[3]


"From the dawn of civilization until 2003, humankind generated five exabytes of data. Now we produce five exabytes every two days...and the pace is accelerating"

(Google's executive chairman Eric Schmidt).[4]

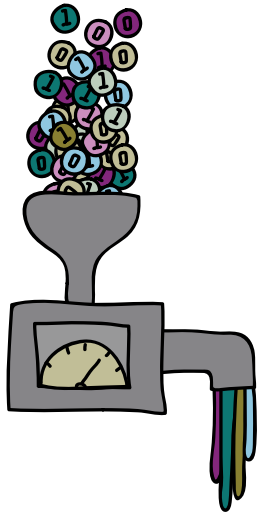
"Big data refers to our ability to collect and analyze the vast amounts of data we are now generating in the world, 'Big Data Analytics' and not 'Big Data' as such are the real game changer"

(Bernard Marr).[5]

The development of Big Data is like "watching the planet develop a nervous system" (Yahoo chief Marissa Mayer).[6]

"It is set to become one of the greatest sources of power in the 21st Century"

(BBC Horizon 2013).[7]



"This new world of business in the era of Big Data requires radically different thinking, new organizational structures and processes, and new leadership skill sets to interpret and connect data in more creative and meaningful ways"

(Sir Terry Leahy, Tesco's Former CEO, 2013). [8]

"It has become a new natural resource. An amazing natural resource"

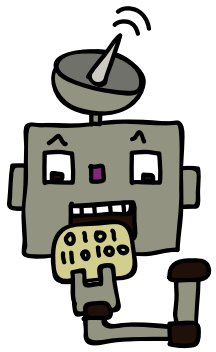
(Jim Spohrer, Director of IBM Global University Projects).[9]

Gartner analyst, Doug Laney, in 2001[10] proposed a three-fold definition encompassing the "three Vs": Volume, Velocity and Variety. This idea has since become popular in defining Big Data, including a fourth V: veracity, to cover questions of trust and uncertainty.

"Big data is the derivation of value from traditional relational database-driven business decision making, augmented with new sources of unstructured data" (Oracle).

Big Data "allows us to find a needle in a haystack"
(Michael Eisenberg,VC,Alpha).

"Big data is the term increasingly used to describe the process of applying serious computing power - the latest in machine learning and artificial intelligence—to seriously massive and often highly complex sets of information" (Microsoft).[11]



The US National Institute of Technology (NIST) argues that Big Data is data which "exceed(s) the capacity or capability of current or conventional methods and systems." In other words, the notion of "big" is relative to the current standard of computation.[12]

# A data Tsunami

Neologisms appear every day to try to define new data-driven experiences, as for example, "dataself"[13] to define our growing incapacity to separate our own subjective sense of who we are from data-driven personal experiences; or "webdata" to express this unstoppable growth of data being collected from our daily digital interactions, through active and passive ways, when we engage in social activities, domestic or professional functions and physical exercise. Today, gadgets record how much electricity each appliance in our house eats up, consumer genomics generate personalized medicine and Nike fuel bracelet tracks personal data while we exercise. "These data-hungry gadgets also harness the power of connecting people with their own data and getting them to see how that could change their lives."[14] Our digitized data and how it is represented back to us becomes "a new dimension of what makes our experiences 'real.'"[15]

We are under a data tsunami, as Chris de la Torre puts it: **"Now we live and love inside our devices - consumers every minute of every day, browsing our laptops, phones, tablets and soon Google Glass; and this data is churning up a bottomless well of ideas - we're consuming and creating. What's the world thinking? Swipe, click and see."**[16]


## Real-time data feedback

Data can now be contrasted with what is happening in real time. Assumptions and predictions are validated and corrected as information is being gathered. **This instant feedback revolutionizes our capacity to understand and act upon real events.** "The numbers and the analysis allow us to have data-driven connections that experienced people who use their hunches would never have guessed," said Professor William DeLone, expert in information systems' effectiveness. "That's the power of information, it doesn't lie."[17] Big Data analytics makes it possible to work through massive amounts of real-time and previously gathered information, in order to find unseen patterns and discover incongruities that can lead to new knowledge, and indicate opportunities for new services and products. Moreover, it allows for the development of ways of operating more efficiently, improving the transparency and accountability of institutions (McKinsey Report 2013, Open data).

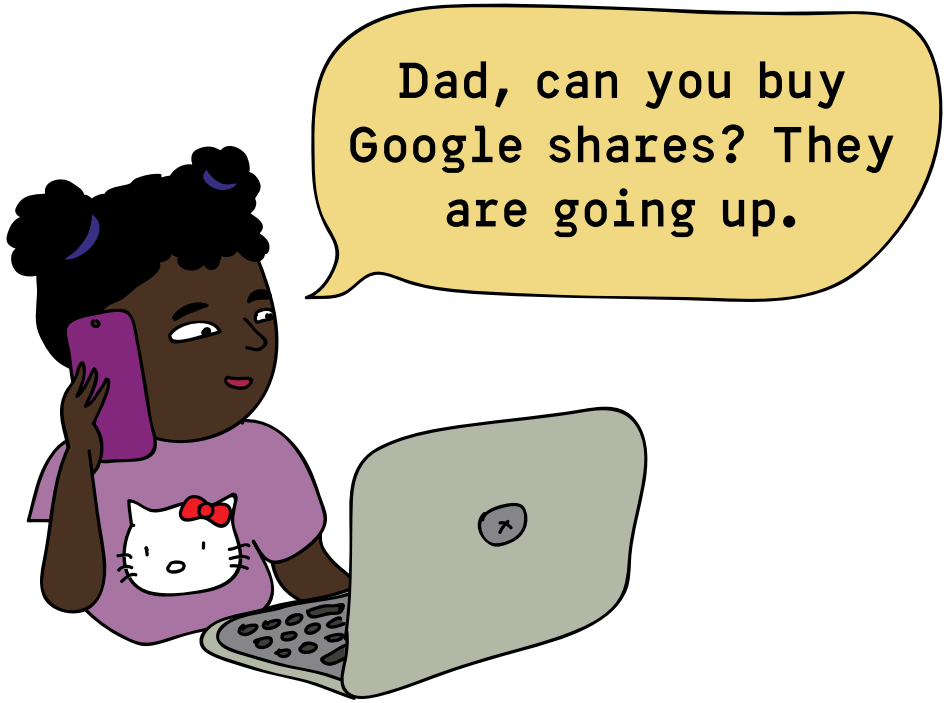
**Citizens are provided with opportunities to access more information than ever.** The impact on public systems, such as health services, is very significant, allowing each individual to take control of his/her own healthcare, by providing access to databases on personal information, relevant and tailored information about preventive measures, information on epidemics, health trends and different possibilities of available treatments and professionals.

# From an elite few to every user

Professor Mayer-Schönberger, Oxford University, said at a recent talk that the beauty about data is that "its value is not exhausted once used; data should be shared, not guarded; and any entity that tries to do so is dangerously close to behaving dictatorially" (Online Educa Berlin, Dec 2013).

**Learning from data, once an exclusive experts' domain, is now being offered as normal practice in different settings, acting as a catalyst for changes on system practices.** "Personal tracking is doing to healthcare what the PC did to computing: It liberated it from the province of an elite few to a tool for the masses" (App developer, Rajiv Mehta).[18] The possibility for users to have access to information validated by real data brings transparency and enables much more collaborative decision-making activities. Google analytics, digital games' dashboards, Facebook or Twitter's statistics are only a few examples of using Big Data to provide meaningful, real-time information to regular users.

According to McKinsey Institute Report 2013, a new trend is emerging which will potentiate even more the general public accessibility of data: open and liquid data. Open data is "the release of information by governments and private institutions and the sharing of private data to enable insights across industries."[19] Liquid data is making open data widely available and in shareable format creating value for the regular user. Entrepreneurial initiatives are seizing this opportunity and creating value out of liquid data.



## Connectivity as the basis for Innovation

Professor Alex Pentland, of MIT, emphasizes the connectivity aspect which makes earlier centrally controlled networks, that solved problems separately, inappropriate to our contemporary challenges. "Instead of focusing only on access and distribution systems, we need dynamic, networked, self-regulating and resilient systems that take into account the complex socio-economic interdependencies of today's hyperconnected world."[20] He acknowledges that the flow and combination of information can lead us to see new patterns, basic to triggering innovation.
Richard Marciano, Digital Innovation Lab at UNC Chapel Hill, talks about the collaborative opportunities of Big Data as a potential resource for innovation. "It is not just the messiness of all this data, but the notion that **big data can create big collaborations,** which invites key questions: How can people get along and bring diverse points to the table? Big collaborations also lead to bigger ideas, so how can we guide research directions and develop innovative approaches that benefit from that kind of diversity?"[21]

## Allowing for common actions

There are several projects which try to use the power of Big Data for the common good, such as raising awareness about world problems, providing spaces for influence on public systems like health or education, or providing access to being active in the development of tools which are useful to society (OpenStreetmap in Haiti or Code for America in the US).[22]
Initiatives like Data without Borders (DataKind.org) support the impact of Big Data as a tool for the common good to be used by the people for the people. The project brings together data scientists and social organizations who believe that the improvement of the quality, access and understanding of data in the social sector can lead to better decision-making and greater social impact. "Data has the potential to make hidden relationships crystal clear, to be a common language between people who might never have spoken, to inspire collaboration, to offer metrics for decision making, and to turn seemingly unrelated ideas into powerful insights that can solve the most complex and intractable problems we face."

## Impact on Science

Science is being boosted and research is currently being transformed by the new possibilities Big Data brings as we realize the infinite complexity of things, from nano biology to the discovery of new universes. Macrosystems, or "big ecology," as David Schimel,[23] senior computer scientist at NASA, calls it, becomes possible only with broad-scale data. **Having large, rich data sets enables scientists to incorporate the complexity and variability of the real world into their models of large-scale phenomena.** CERN, the Swiss nuclear physics lab, uses the computing powers of thousands of computers distributed across 150 data centers worldwide to analyze the data and unlock the secrets of our universe. At Cornell University, Hod Lipson and Michael Shmidt, computer scientists, have developed an Artificial Intelligence program for Robotics, using data considered by them too large and complex for humans to study.[24]

## Data collection, before the hypothesis?

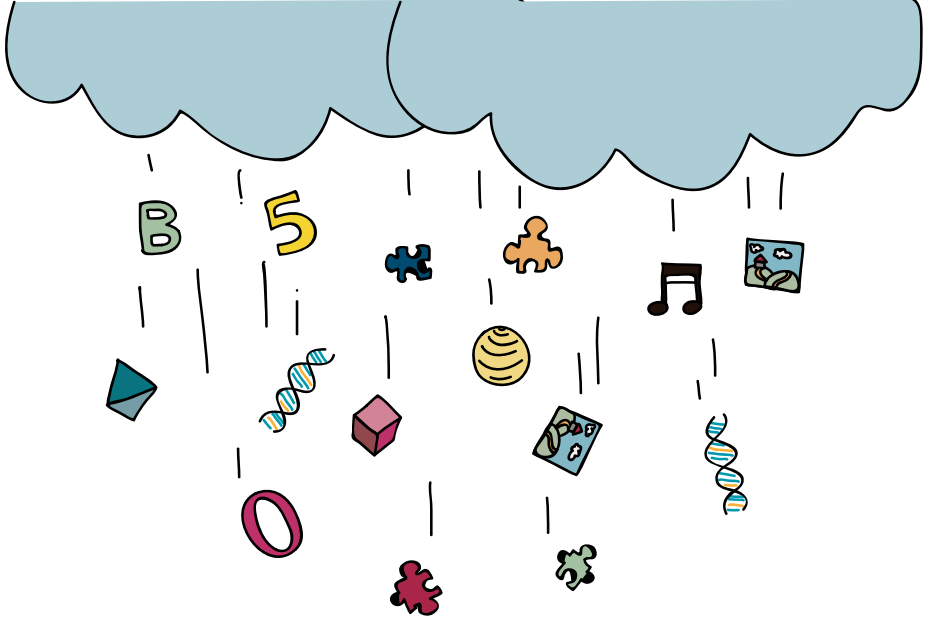
Simon DeDeo, mathematician, brings up another significant impact on research. "Now we have this new multimodal data [gleaned] from biological systems and human social systems, and the data is gathered before we even have a hypothesis. The data is there in all its messy, multi-dimensional glory, waiting to be queried, but **how does one know which questions to ask when the scientific method has been turned on its head?"**[25]

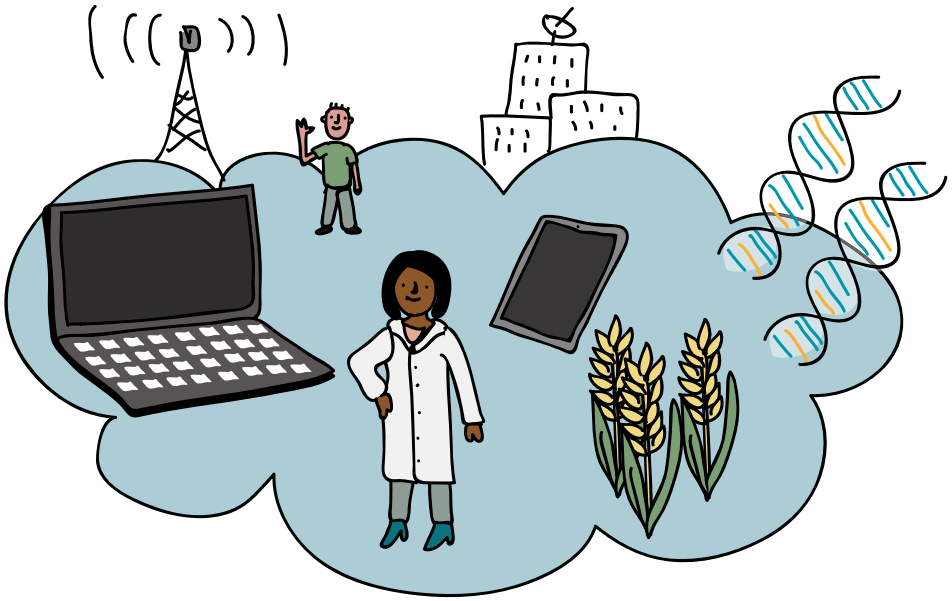
# Structural Changes

The latest technological developments, specially the cloud-based servers and the development of data networks, are enabling and enlarging the potential of Big Data. Harvey Newman, physicist, points to the significant structural changes Big Data require. If current trends hold, the computational needs of Big Data analysis will place considerable strain on the computing fabric. "It requires us to think about a different kind of system. The preferred architecture no longer features a single central processing unit (CPU) augmented with random access memory (RAM) and a hard drive for long-term storage. Even the big, centralized parallel supercomputers that dominated the 1980s and 1990s are giving way to distributed data centers and cloud computing, often networked across many organizations and vast geographical distances."[26]

Cloud computing has provided to the general public easy, scalable access to computing resources. Amazon Web Services is, today, the largest public cloud provider.[27]

**According to Shonberger and Cukier[28] three technological advances are leading Big Data to change the way we live, work and think: increased datafication of things, increased memory storage capacity and increased processing power.**



# Visualization:
## Universalizing Big Data literacy

**The digital world has very quickly understood that visualization needed to be a major feature in order to deliver information in a universal manner, making it simple to share ideas with others.** Parallel to the development of data analytics there is a very creative and intelligent development of digital literacy through visualization, bringing information, traditionally hard to understand and exclusive to experts, to anyone interested.

In 2006,[29] Hans Rosling mesmerized his audience at a TED conference, bringing statistics to life. Worried about sending an important message about health and economics in the developing world, Rosling developed software in which moving bubbles and flowing curves transformed heavy data into a clear and intuitive form.

In 2011, Deb Roy stunned his audience at another TED conference, showing the results of a research about language acquisition, and how he communicates visually the analysis of the complexity of the data collected from 90,000 hours of several different cameras filming the movements of a child and his family.[30]

# Skepticism

Skepticism is also an important part of the buzz about Big Data. Experts agree that **there is still a long road until the complexity of the data being generated today can be easily transformed into meaningful information.** "Today's big data is noisy, unstructured, and dynamic rather than static. It may also be corrupted" (Alessandro Vespignani).[31]

"Much of the recent data frenzy, from the physical and life sciences to the user-generated content aggregated by Google, Facebook and Twitter, has come in the form of largely unstructured streams of digital potpourri that require new, flexible databases, massive computing power and sophisticated algorithms to wring out bits of meaning from them" (Matt LeMay, Bitly).[32] "Big data is not magic, it doesn't matter how much data you have if you can't make sense of it."

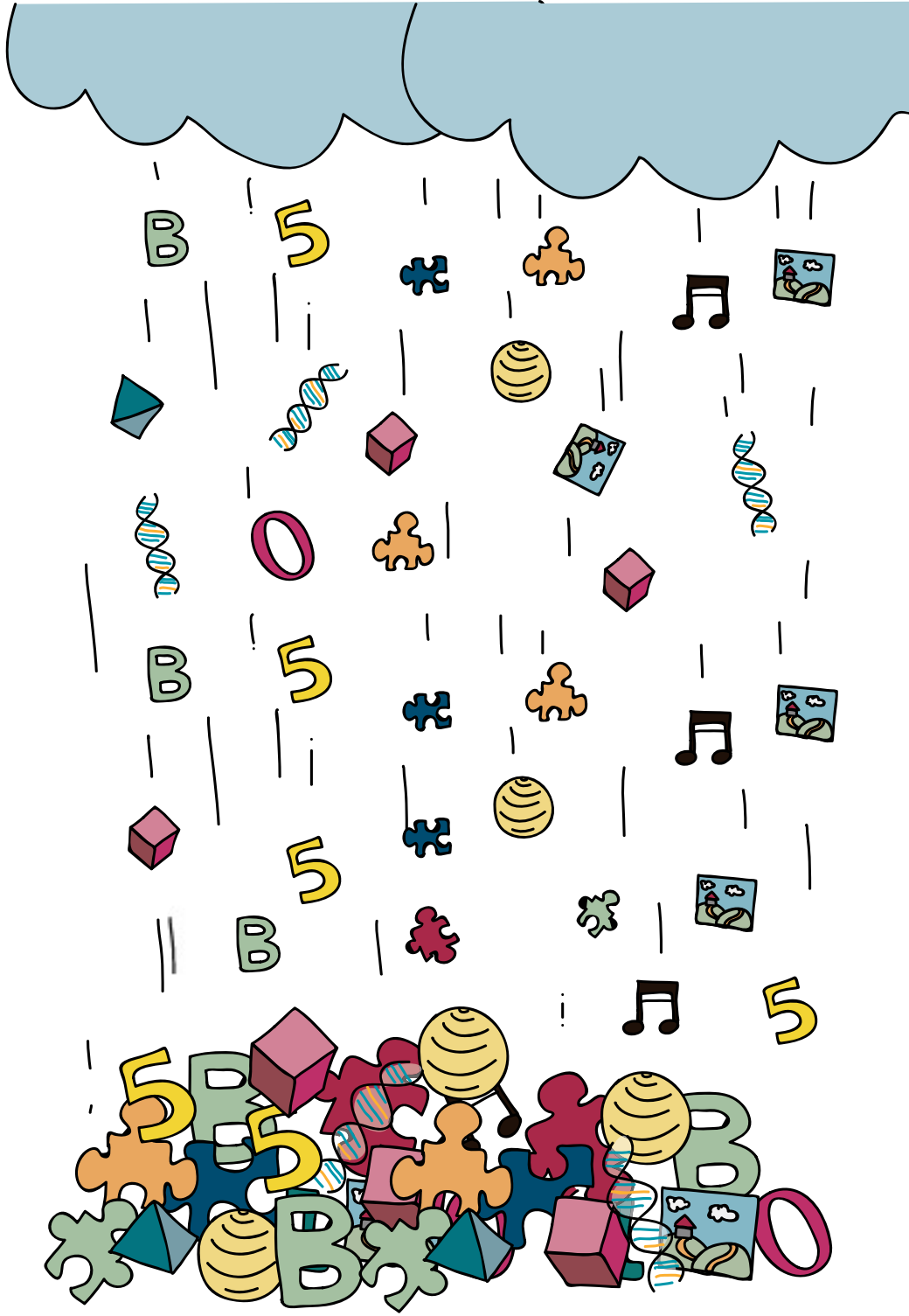
Steven Rosenbaum, entrepreneur (Magnify.net), alerts that we need "superheroes" and super-fast to make sense of the rising tide of data and information, maintaining that the fact that we are getting better at making data does not mean we are any better at making sense of it. "While devices struggle to separate spam from friends, critical information from nonsense, and signal from noise, the amount of data coming at us is increasingly mind-boggling."[33]

# Proposed solutions to make sense of Big Data Complexity

Pioneering promising tools are currently being explored in order to handle this brave new world of data. Ronald Coifman, mathematician,[34] suggests that what is needed is the Big Data equivalent of a Newtonian revolution: "It is not sufficient to simply collect and store massive amounts of data; they must be intelligently curated, and that requires a global framework." Coifman believes that modern mathematics –notably geometry – can help identify the underlying global challenges.[35]

Alessandro Vespignani, mathematician, uses everything from network analysis (creating networks of relations between people, objects and documents in order to uncover the structure within the data) to machine learning, and old-fashioned statistics. "In the end, data science is more than the sum of its methodological parts," and the same is true for its analytical tools. **"When you combine many things you create something greater that is new and different."**

Harvey Newman foresees a computational future for Big Data that relies on a type of automation through well-coordinated armies of intelligent agents that track the movement of data from one point in the network to another. Each might only record what is happening locally but would share the information in such a way



as to shed light on the network's global situation. "Thousands of agents at different levels are coordinating to help human beings understand what's going on in a complex and very distributed system." The scale would be even greater in the future, when there would be billions of such intelligent agents, making up a vast global distributed intelligent entity. "It's the ability to create those things and have them work on one's behalf that will reduce the complexity of these operational problems. At a certain point, when there's a complicated problem in such a system, no set of human beings can really understand it all and have access to all the information."[36]

Among the existing projects to deal with Big Data, one of the most significant, widely used across the industry, is Apache Hadoop, an open source software project that enables the distributed processing of large data sets across clusters of servers. Hadoop facilitates the analysis of the unprecedented volumes and velocity of unstructured data being currently produced as video, audio, social media posting, images, etc. **"In today's hyper-connected world where more and more data is being created every day, Hadoop's breakthrough advantages mean that businesses and organizations can now find value in data that was recently considered useless."**

## Unprepared and insufficient professionals

The lack of sufficient knowledge and professionals to deal with and interpret these complex and multi-varied data systems raises a big question mark. A panel of higher education experts (2013 Campus Computing Project) has expressed concern about the sky-high expectations and big investments to collect, manage and analyze data to improve student retention and guide students more efficiently. "Big Data may be transformational, but expecting that transformation to be immediate is unfair."[37] "The biggest problem with big data is that when people hear the term now in higher education, they're desperate to play catch-up, and think they can be where everyone else in the market is within a month" (Phil Ice, vice president of research and development at the American Public University System).[38] As in any field trying to benefit from the advantages of new technologies, the implementation and professionalization process is often overlooked. Many organizations that have implemented business intelligence and analytics initiatives without the proper preparation, software or personnel have not seen the expected results. Companies can get carried away with the possibilities of these tools while simultaneously failing to develop the right strategies for their best possible application.[39] The same has happened, during the last decade, with the implementation of computers in educational systems with no concurrent infrastructure preparation of broadband or professional training, resulting in waste of resources and disappointing results.

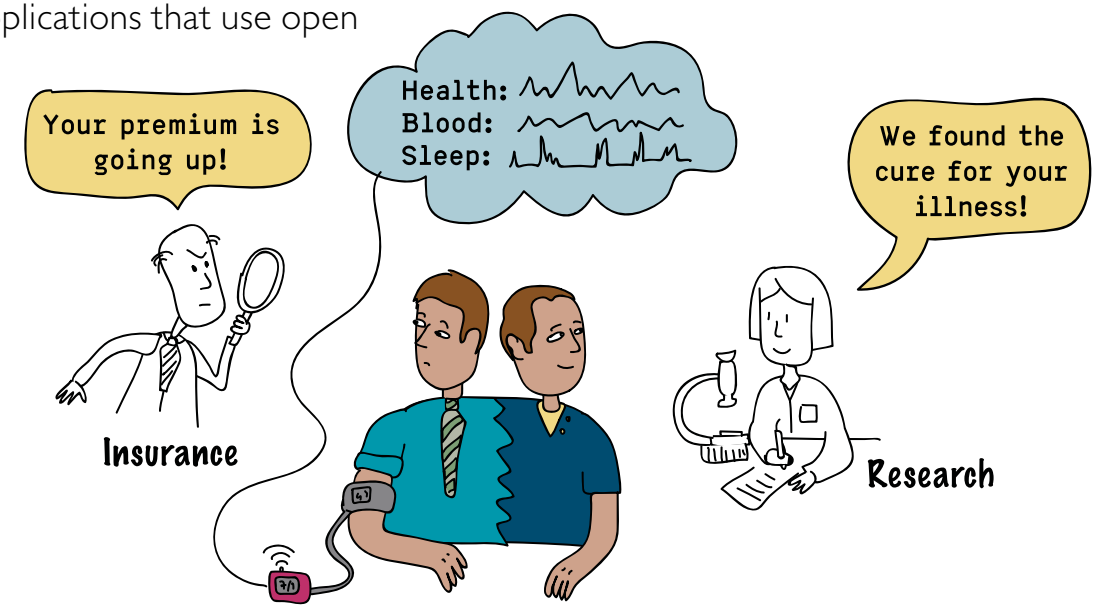
Slowly these concerns are being echoed within the education community. There is a clear rise in investment in higher education in Big Data-related infrastructure. The University of Rochester has spent more than US $100 million; Indiana University spent more than $30 million ($7.5 million on the Big Data-crunching super computer called Big Red II); the Gordon and Betty Moore Foundation and the Sloan Foundation have pledged $37.8 million to the University of California at Berkeley, the University of Washington and New York University for a Big Data collaboration.[40]

# Privacy and Awareness

**"Big data has become a way of bringing the whole world into focus at once:** capable of deriving correlations between gigantic data sets, its revelations are set to prove as valuable to scientists, educators and health professionals as they clearly already are to the NSA (US Intelligence) and GCHQ (GB Intelligence)" ( Professor Viktor Mayer-Schönberger, Oxford Internet Institute).

The benefits and highly cost-effective use of open sources by companies have to be weighed against privacy concerns. **Consumers may gain by having access to more information; however, awareness has to be raised about use and misuse of individual information.** Once aware and informed, individuals can and should use for their advantage the value potential of applications that use open data and provide feedback to feed and improve the efficiency of these tools. Individuals can become not only receivers but also active, conscious providers of information that only then could be transformed into a personal advantage.

Red alarms about data accessibility and the real value of privacy exploded following the publication of Wikileaks and Edward Snowden's documents exposing the activities of governments' military, diplomatic and secret services. Online Educa Berlin 2013 dedicated a panel discussion to this subject: "The end of secrecy and what it means." Dr Harold Elletson asserted that with the growing accessibility to data, total secrecy is becoming impractical in the modern age and in the connected future "both secrecy and security will be impossible without consent."[41]



# Understanding the ways Big Data is being used/ affecting our daily life:

1. Google Search Engine is the best example of Big Data intelligently being offered to help anyone, using a digital device, to find information of all sorts, and it has determinately changed our information gathering habits. According to Forbes, Google has the "largest database on the planet."[42]

2. Personal wearable gadgets, such as smart watches or smart bracelets, generate data and inform us about our body functioning by analyzing collective data. Take the Up band from Jawbone as an example: the armband collects data on our calorie consumption, activity levels and our sleep patterns. The company analyzes huge volumes of data collected for 60 years of data on sleeping patterns, bringing information which is fed back to individual users.
3. Most elite sports have now embraced Big Data analytics: the IBM SlamTracker tool for tennis tournaments; video analytics track the performance of every player in a football or baseball game; sensor technology in sports equipment such as basket balls or golf clubs (providing feedback via smart phones and cloud servers); athletes use smart technology to track nutrition and sleep, as well as social media conversations to monitor their emotional well-being.
4. Most online dating sites apply Big Data tools and algorithms to find the most appropriate matches.
5. Google Ngram Viewer allows us to understand cultural trends over time through the search of specific words. The tool is based on a humongous data set based on the millions of books Google has digitized overs the years.
6. Big Data analytics allows for the monitoring and prediction of the developments of epidemics and disease outbreaks. For example, flu outbreaks can be detected in real time by integrating data from medical records with social media analytics (from what people are typing, e.g., "Feeling rubbish today – in bed with a cold").
7. Reg4ALL is an initiative that promotes citizen action for a common good. It is a platform that allows individuals and communities to aggregate data creating a public health open database. Individuals are invited to voluntarily donate results on specific health issues to open databases which allows clinicians to study the variations and leads to greater insight and ultimately effective treatments.[43]

8. Optimization of traffic flows is based on real-time traffic information as well as social media and weather data. Currently there are pilot projects using Big Data analytics for the development of Smart Cities, where the transport infrastructure and utility processes are all joined up: a bus would wait for a delayed train and traffic signals predict traffic volumes and operate to minimize jams. An example is the implementation of Intel's Apache Hadoop to help overcrowded cities in China.[44]

9. The US government is investing heavily in improving security by enabling law enforcement, as for example the NSA, to use Big Data analytics to foil terrorist plots or to detect and prevent cyber attacks.

10. Social media data, browser logs, text analytics and sensor data are being used to get a picture of customers and understand their behaviors and preferences in order to create predictive models. Department stores are now able to very accurately target their marketing. There is a famous illustration case of a father who got angry at Target because his daughter was receiving pregnancy-related advertisements; he then found out that Target "knew" she was pregnant before he did, based on data of her recently changed cosmetic buying habits.[45]

11. Optimization of business processes is possible based on predictions generated by data such as social media data, web search trends and weather forecasts, which helps, for example, retailers to adapt their stock. Another example is geographic positioning and radio frequency identification sensors which are used to track goods or delivery vehicles and optimize routes by integrating live traffic data.

12.. Professor Sebastian Thrun (Stanford University) and Peter Norvig (data scientist, Google) are leading a project to build a self-driving car that relies on Artificial Intelligence algorithms and all the data collected from the recording and measurement of Google's street view vehicles.[46]

# Why is Big Data, BIG?

**What is really big about Big Data is the size of the impact it is having on our society; the growing possibilities it is giving to us as digital users, strengthening the value of information in our daily lives.** Data is already a source of power in the modern world, and a huge valuable commodity for those who can analyze it. Moreover, it gives opportunities to all to have access to knowledge previously considered beyond reach or the privilege of a few. We are all digital data users and contributors.

It has, however, also created expectations and fears that have to be dealt with, as with most other revolutions of our century, fast. Data allows us to know more but not to know it all! Data's true value comes from what we make of it. **The tsunami has arrived and we need to be smart enough to transform its challenges into opportunities, especially, in areas related to a public right, like Education!**

1. http://www.wired.com/insights/2013/08/why-big-is-blinding-us-to-the-real-value-of-big-data/
2. Undefined By Data: A Survey of Big Data Definitions. Jonathan Ward, Adam Baker, Sept 2013.
3. http://openthoughtsmarter.blogs.uoc.edu/rethinking-the-approach/
4. http://smartdatacollective.com/bernardmarr/141351/what-really-big-data-and-why-it-will-change-world#!
5. http://smartdatacollective.com/bernardmarr/141351/what-really-big-data-and-why-it-will-change-world#!
6. http://www.wired.com/wiredscience/2012/10/big-data-is-transforming-healthcare/
7. http://www.youtube.com/watch?v=CO2mGny6fFs
8. The Intersection of Big Data and Leadership: Lessons from Sir Terry Leahy, Stern Speakers, 10 Dec 2013.
9. http://www.ecampusnews.com/featured/featured-on-ecampus-news/big-data-bang-344/2/
10. MetaGroup, 3D data management: Controlling data volume, variety and velocity. 2001
11. http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/
12. http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/
13. http://thenewinquiry.com/blogs/marginal-utility/dumb-bullshit/
14. http://www.wired.com/wiredscience/2012/10/big-data-is-transforming-healthcare/
15. http://dmlcentral.net/blog/lyndsay-grant/understanding-education-through-big-data
16. Delatorre, Christopher. "Chasing Innovation—On data, disciplines, and ditching the rules." Urbanmolecule., 20 Jul. 2013.
17. http://kogodnow.com/2013/03/big-data-ignites-revolution-in-decision-making/
18. http://www.wired.com/wiredscience/2012/10/big-data-is-transforming-healthcare/
19. McKinsey Report 2013: Open Data: Unlocking Innovation and Performance with Liquid Information
20. http://openthoughtsmarter.blogs.uoc.edu/rethinking-the-approach/
21. http://www.hastac.org/blogs/slgrant/2013/01/15/socializing-big-data-collaborative-opportunities-computer-science-social-sc
22. McKinsey Report 2013: Open Data: Unlocking Innovation and Performance with Liquid Information
23. http://www.wired.com/wiredscience/2013/10/big-data-science/all/
24. http://www.forbes.com/pictures/lmm45emkh/7-hod-lipson-and-michael-schmidt-computer-scientists-cornell-university/
25. http://www.wired.com/wiredscience/2013/10/topology-data-sets/all/
26. https://www.simonsfoundation.org/quanta/20131009-the-future-fabric-of-data-analysis/
27. EdTech Powered by Big Data.Report by Astra. 2013
28. BIG DATA: A REVOLUTION THAT WILL TRANSFORM THE WAY WE LIVE, WORK AND THINK, book by Shonberger and Cukier, 2013
29. http://www.ted.com/speakers/hans_rosling.html
30. http://www.ted.com/talks/deb_roy_the_birth_of_a_word.html
31. http://www.wired.com/wiredscience/2013/10/topology-data-sets/all/
32. http://www.wired.com/wiredscience/2013/10/big-data-science/all/
33. http://www.fastcompany.com/1834177/content-curators-are-new-superheros-web
34. http://www.wired.com/wiredscience/2013/10/topology-data-sets/all/
35. http://www.wired.com/wiredscience/2013/10/computers-big-data/
36. https://www.simonsfoundation.org/quanta/20131009-the-future-fabric-of-data-analysis/
37. http://www.campuscomputing.net/item/2013-campus-computing-survey-0
38. http://1776dc.com/2013/12/13/how-big-data-is-changing-the-educational-frontier/
39. http://smartdatacollective.com/roman-vladimirov/167801/governance-helps-perfect-big-data-initiatives
40. http://www.ecampusnews.com/featured/featured-on-ecampus-news/big-data-bang-344/?ps=CeciliaW@cet.ac.il-0013000000135NyG-0033000001CbsTP
41. http://www.online-educa.com/OEB_Newsportal/e-learning-takes-the-limelight/?goback=%2Egde_1891552_member_581974372674598 5026#%21
42. http://www.forbes.com/pictures/lmm45emkh/1-larry-page-ceo-google/
Big Data is good for your health. Sharon Terry, Genetic Alliance
44. http://www.intel.com/content/dam/www/public/us/en/documents/case-studies/big-data-xeon-e5-trustway-case-study.pdf
45. http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/
46. http://www.forbes.com/pictures/lmm45emkh/3-sebastian-thrun-and-peter-norvig-data-scientists-google/

# Education & Big Data

MindCet
Changing Education Mindset

# Education & Big Data

Let us, for a moment, envision an educational system where all players - students, teachers, parents, politicians, publishers, developers, researchers - all are active participants, and none merely a receiver. Would it be a much more efficient and relevant system? Is the Big Data movement an enabler of such a system?


MIT Media Lab founder, Nicholas Negroponte, saw the computer as a medium for empowering communication between people and machines, "being connected is the key." Joi Ito, Media Lab current director, follows this vision through exploring and expanding Big Data-related possibilities to the educational world. Ito looks for "a world with seven billion teachers, where smart crowds, adopting a resilient approach and a rebellious spirit, solve some of the world's great problems; a world of networks and ecosystems, in which unconstrained creativity can tackle everything from infant mortality to climate change. We want to take the DNA [of the lab], the secret sauce, and drop it into communities, into companies, into governments. It's my mission, our mission, to spread that DNA. You can't actually tell people to think for themselves, or be creative. You have to work with them and have them learn it themselves."[1] "Big data is the foundation on which education can reinvent its business model and build the coalition of governments, businesses, and social entrepreneurs that can bring together the evidence, innovation and resources to make lifelong learning a reality for all" (Andreas Schleicher, Special Advisor on Education Policy, OECD)[2].

The Big Data wave is slowly reaching the educational system, especially through entrepreneurial initiatives that are offering a wide variety of different learning and systemic solutions. The development of personalized and adaptive learning systems is the current hit of educational gatherings advocating that this may be the key, not only to engage the student, but also to have a system that can respond to each student's real learning needs. Moreover, Big Data software providers are building systems that provide information to all members of the education community, which could lead to an efficient collaboration. Policy makers foresee a chance of basing their educational decisions on real-time data gathered from as many places as they wish.

Traditional educational big players are strongly investing in Big Data. Major educational publishers such as Pearson and McGraw-Hill are turning their efforts towards dynamic online platforms that are equipped to collect data from students who are interacting with them, providing adaptive and tailored responses.They have recently joined forces with younger adaptive learning companies such as Knewton[3] and Aleks[4], respectively. Infrastructural software vendors such as Blackboard[5], that reaches to a wide spectrum of the educational system, and Ellucian[6], to higher education, base their systems on data analytics tools to predict student success based on data logged by their clients' software systems. Foundations such as Bill & Melinda Gates[7] are promoting the use of Big Data to measure and help improve student learning outcomes; they have recently invested US$ 100 million in a non-profit personalized learning company, inBloom[8].

Educators are starting to claim that the benefits already in practice in other industries need to be promptly implemented by the educational system. "The average retail store knows more about a box of cereal on their shelves than we know about our students. Looking at what it takes to get someone to buy something, and at what level you want them to buy, and the campaigns you present them with, is essentially no different than planning learning outcomes"[9] (Phil Ice, American Public University System).

# Learning Analytics and Educational Data Mining

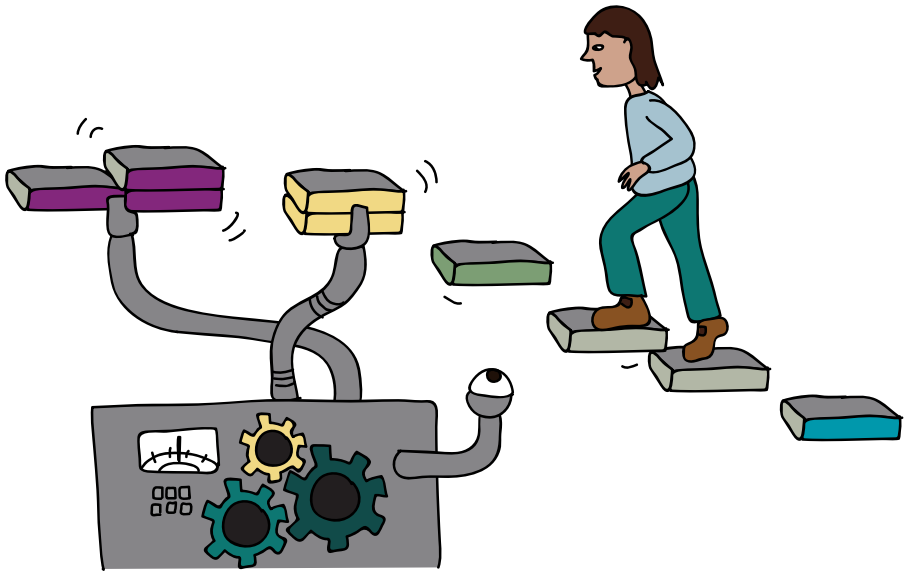
Education is trying to bridge the development that data science reached in other areas, through the development of Learning Analytics and Data Mining: processes of generating actionable knowledge from huge amounts of data.
Horizon Report 2013[10] describes Learning Analytics as the "field associated with deciphering trends and patterns from educational big data, or huge sets of student-related data, to further the advancement of a personalized, supportive system of education." The essential idea behind Learning Analytics is to use data analyses to adapt instruction to individual learner needs in real time, in the same way that Amazon, Netflix and Google use metrics to tailor recommendations and advertisements to consumers. Learning Analytics allow for the prediction of future student performance (based on past patterns of learning across diverse student bodies), recommendation and provision of feedback tailored to the student's answers, personalization of the learning options, and adapting teaching and learning styles. Often overlapping with the concept of Learning Analytics, Educational Data Mining is oriented to developing ways to discover patterns in data through exploration, searching for new knowledge – trying to identify interesting educational phenomena. Researches on both Learning Analytics and Data Mining are looking for applications that benefit learners as well as informing and enhancing the learning sciences[11]. The definition by the International Educational Data Mining Society is of "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. Whether educational data is taken from students' use of interactive learning environments, computer-supported collaborative learning, or administrative data from schools and universities, it often has multiple levels of meaningful hierarchy, which often need to be determined by properties in the data itself, rather than in advance. Issues of time, sequence, and context also play important roles in the study of educational data." Startups such as inBloom and Knewton offer services which draw together existing data from a wide range of sources, as well as data produced as a by-product through students' use of technology. Individual data is analyzed together with the data from hundreds of thousands of students, creating learning profiles, diagnosing their strengths, weaknesses and challenges as well as offering tailored learning paths[12].

# Personalized and Adaptive Learning

The concept of personalized learning has entered the EdTech market in strength and is forcing traditional instruction and content providers to search for ways to focus on the student experience. A significant number of startups are developing the means to use data analytics to allow a much more personalized educational system, tailored to the needs of each student. Today, the systems working towards the development of personalized learning have their basis in vast amounts of data generated by students while they interact with online learning environments, detecting what they know and how they learn best. The systems are able to analyze this data and recommend in real time what and how the student should study next.

Adaptive learning is being broadly used to refer to "adaptive" programs that offer different content to learners, based on an assessment of what they seem to know (Edsurge). Its specificity from the general concept of personalized learning is that adaptive learning entails an ongoing process of the system to adapt based on constant feed of new data. Adaptive learning platforms continuously collect data from the student in order for the system to learn and adapt the student's learning pathway that changes and improves over time. Personalized learning can also include systems derived from a rules-based method of decision trees that leads to pre-determined paths.

Adaptive learning provides students with modular learning environments, meaning that the curriculum is broken up and individualized. Each student sees a different curriculum adapted and adjusted in accordance with his/her learning capabilities and pace, by capturing student data from every keystroke, and offering what the student should be doing next. "By recalibrating with every interaction



to maintain appropriate challenges, learners stay in their optimal learning zone and are enabled to meet their full learning potential. This exciting advance in education has the potential to be the 'equalizer' that provides greater access and opportunity for students in our society, regardless of their backgrounds or zip codes" (Tom Vander Ark, Getting Smart CEO)[13].

Dreambox Learning[14] is introducing an adaptive learning environment to teach Math for Primary School. The student interacts with an immersive game-like adventure environment where the students show their work and thinking process virtually, encouraging them to explain, discuss and defend their mathematical thinking. Dreambox advocates that they developed an Intelligent Adaptive Learning system introducing a new generation of education technology that enables new learning experiences, adjusts path and pace to stay within the kid's zone of optimized learning, helping to accelerate understanding and critical thinking. The system also provides formative and summative data to the student's teacher to enable a more personalized experience in the classroom.

"When we refer to adaptive learning, we mean a system that is continuously adaptive - that responds in real-time to each individual's performance and activity on the system. It maximizes the likelihood a student will understand a certain concept by recommending the right instruction, at the right time, about the right thing" (David Liu, Knewton COO). Knewton, currently one of the major players in this field, uses Big Data to provide adaptive learning and analytics to students, teachers, districts and publishers. Its main educational value is in data analytics to map each student's strengths and weaknesses over time, in order to enable teachers to personalize and tailor instruction and content. "Knewton personalizes digital courses so every student is engaged and no student slips through the cracks[15]." Knewton couples personalization capabilities with tools to manage the class, alerts on required interventions and recommendations on how to form homogeneous working teams and provides to the teacher a holistic view of the class.

Another important player is inBloom which provides to states and districts technological support analyzing students and teaching data in order to allow teachers to personalize their work, districts/states to help detect weaknesses of the educational system and to inform parents.

# Educational Data

Today, when more and more learning is done online, we can record student learning activity in high granularity, from many different sources such as student records, Learning Management Systems (LMS), courseware published and shared online and a whole world of educational-content data currently available online. Data is managed at all levels (individual, school, district, state), in many different systems, and in all forms (structured and unstructured through document texts, pictures, videos, interactive actions, etc.). The main challenge is to integrate the different pieces of data in order to create a coherent view. "Right now, all sorts of student data are being kept in everything from testing programs and instructional software to grade books and learning management systems. But the data are often trapped in the program and not easily extracted or combined with other data on the same student, creating the educational equivalent of the Hotel California: data can check in any time it likes, but it can never leave. Or be used effectively by teachers" (Frank Catalano, Edtech analyst)[16].

Most legacy software systems in education have been constructed with little consideration of data portability. Many initiatives today focus on providing a common language or vocabulary and structure to enable seamless sharing of data among different systems and applications. More and more, companies are starting to push for the aggregation of student data into analytics tools that can be sold in turn back to the school.

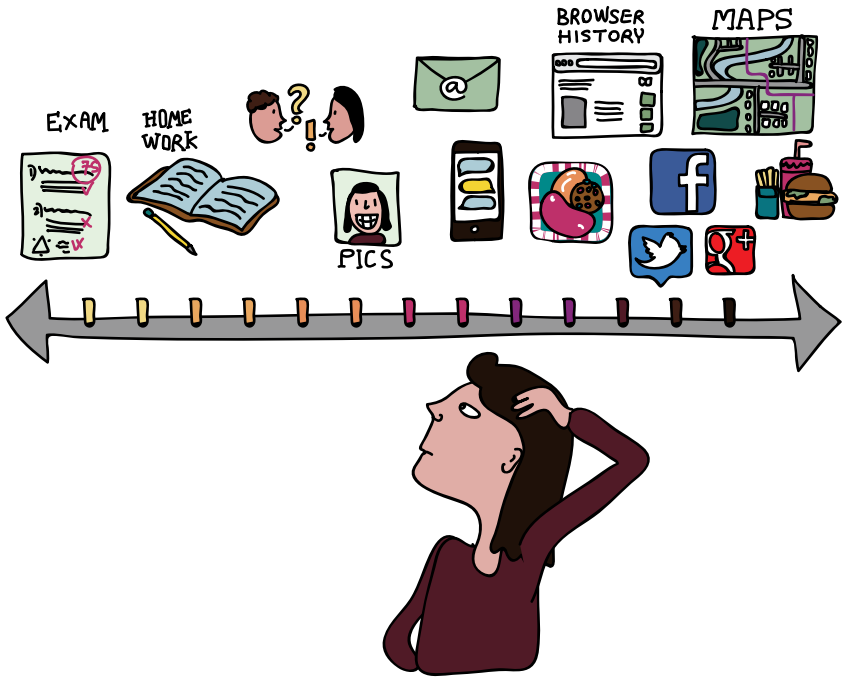
# How broad should Educational Data be?

How much information does the system need to understand the student's academic performance?

Systems are starting to include students' life activities as part of educational data: library check-outs, gym visits, inter-mural sports participation, cafeteria and bookstore purchases, minutes from student meetings, times in and out of students' dormitory, LMS logins and sessions, blog and forum comment history, internet usage while on campus, e-mails sent and received via university email accounts, pages students read in digital textbooks, the passages they highlight, their social media profiles, videos watched on MOOCs, their Wikipedia visits, etc[17].

"Is it our responsibility to monitor social media sites to help protect students from the dangers of bullying, drug use, violence, and suicide?" asks a principal of a Middle School, in a debate about the invasive social media in school and learning settings.

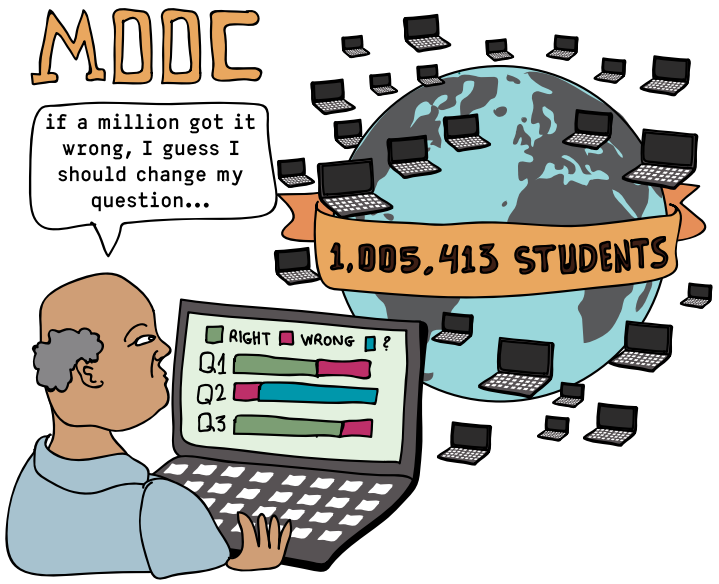
Educational platforms' use of social networks creates a dilemma about the boundaries of educational data[18].



# Educational massive collection of data

The MOOC (massive open online course) market has grown exponentially in the last couple of years, and the major MOOC providers are Coursera, Udacity, EdX & Khan Academy. MOOCs, aimed at unlimited participation and open access via the web, use short video lectures coupled with a set of assessments adapted to the masses (over 100K students in popular courses), automated feedback through online assessments (e.g., quizzes and exams) and peer-review and group collaboration activities. Coursera, for example, collect data for every action (or inaction) performed by a student – when a student pauses a video, increases playback speed, answers a quiz question, revises an assignment or comments in a forum. This microscopic level of data, when collected at the scale that MOOCs operate, facilitates the identification of defaults in the system. As Daphne Koller, co-founder of Coursera, points out: "If two students in a university class of a hundred give a wrong answer, you would never notice, but when two thousand people give the same wrong answer, it's kind of hard to miss[19]."

The massive amounts of data generated from course enrollments, ranging from 10,000 to 100,000 students, enable providers to improve outcomes such as the optimization of course material. If a test question is answered incorrectly, or if students lose focus during a specific point in the course, data can direct the course creators to go back into the curriculum to add or modify. MOOC providers are taking advantage of this scale to experiment with course materials, presentation methods and communication with the students. For example, Sebastian Thrun, founder of Udacity, A/B tested a color lesson vs. a black and white lesson, and found that "Test results were much better for the black-and-white version…that surprised me." Andrew Ng used A/B testing at Coursera to experiment with e-mail reminders to increase engagement. This methodology of data-driven education is only possible when you have substantial scale – hundreds or even thousands of users. "More than anything, data and scale will enable teachers and instructors to have actionable feedback on what is, and what is not, working" (Salman Khan, founder of Khan Academy)[20].

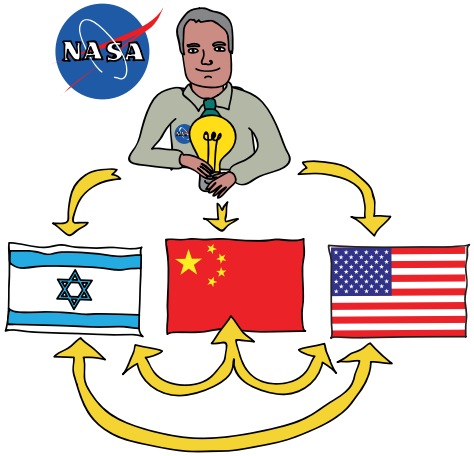


# Interchanging data between educators

Our relation with the business world has gone from trusting people to provide information, to willingly handing over credit card data, to connecting trustworthy strangers in all sorts of marketplaces. Worried about the lack of a similar trend in education, Project "MyPISA" tries to build a team of educators who actively interchange and share information; as they say it, "big data is building big trust."[21] Principals and teachers are beginning to see themselves as teammates – not just spectators – on a global playing field.

# Sharing Research Data with educators and students

NASA and Amazon Web Services Inc. (AWS) are making a large collection of NASA climate and Earth science satellite data available to research and educational users through the AWS cloud. The system enhances research and educational opportunities by promoting community-driven research, innovation and collaboration. "NASA continues to support and provide open public access to research data, and this collaboration is entirely consistent with that objective," said NASA Chief Scientist Ellen Stofan[22]. By using the cloud, research and application users worldwide gain access to an integrated Earth science computational and data management system they can use on their own. "We are excited to grow an ecosystem of researchers and developers who can help us solve important environmental research problems," said Rama Nemani, principal scientist for the NEX project. "Our goal is that people can easily gain access to and use a multitude of data analysis services quickly through AWS to add knowledge and open source tools for others' benefit."



# Housing Educational Data

Collaborative initiatives to create shared data repositories and standardization of systems that collect, manage and integrate educational data is the goal of various organizations. Schoolzila[23] offers data warehouse and hosting to house all data from the school or district with relevant source systems (SIS, HRIS, surveys, assessment systems, etc.) plus a set of reports and exploration tools.

In March 2013, inBloom was given the responsibility of maintaining a data warehouse containing the files of millions of students in the US public school system, a collaborative project between the Bill & Melinda Gates Foundation, the Carnegie Corporation of New York and school officials from various states. inBloom develops portals to allow mining of those data for a variety of purposes (NHM Horizon Report 2013: K-12 edition).

# Visualization: Focusing on the user experience

Part of the impact of Big Data on the general user is its visual expression. Companies providing products and services to the educational industry are increasingly aware of the importance of focusing on the user experience (UX), developing visuals which are friendly and easy to understand. "Visualization serves us because it puts the tools of understanding business directly in the hands of those needing to make decisions" (Chris Taylor, Wired)[24]. Because data analytics offers insights for every tier of the educational system, from the student to governing bodies, its expression has to offer clear and different possibilities of understanding the outcome.

A lot of thought and resources are being invested to develop the optimum visualizations, as a major feature of educational products. The capacity of a teacher to visualize on one screen, in real time, what is happening with every kid in the class, can significantly change the teacher's performance. The learning maps of Knewton, for example, show the unique sequence a student takes across content modules to attain a learning objective. At a district level, Blackboard offers interactive dashboards for monitoring and analyzing college activity.

# Preventive and predictive measures

The possibility of predicting student outcomes can be a valuable resource for the educational industry, not only allowing the system to provide resources to prevent undesirable outcomes, but also providing information about the students' suitable academic future. "If you look at state assessment reports for K-12s, you can see how easy it is to use this data. The best states will have navigable websites that export data, highlight issues surrounding income, and in turn, impact higher ed as they start to get a clear picture on which students have difficulty succeeding" (Barbara Dreyer, CEO of Connections Education). Blackboard Analytics Services GM, Jim Hermens, demonstrates that educators' access to the adequate data analysis can positively affect students' retention. "Taking what you know about a student before he or she matriculates, and then using that info to plan his or her overall success, has now been proven to be an effective tactic."[25] Blackboard released in 2013 a "Retention Center" tool, within its LMS, for educators to quickly identify students who are falling behind, based on research outcome that led to four important indicators: student login history, course activity, missing deadlines and grade drop.

Due to the accessibility to data and the possibilities of analyzing it, we are starting to see a wave of studies and products trying to predict students' academic performance and behavior. We can also find educational institutions already using these methods to identify students at risk as early as possible. EdWeek.org has recently published the case of Maryland educators who are finding that the early-warning signs of a student at risk of dropping out may become visible at the very start of their school careers. The affluent and tech-savvy 149,000-student Montgomery County public schools, in a suburb of Washington, is building one of the first early-warning systems in the US that can identify red flags for 75 percent of future dropouts as early as the second semester of 1st grade[26].

Professor Viktor Mayer-Schönberger, of Oxford University, warns about the overweight educational institutions may give to Big Data analysis in order to predict students' academic performance creating what he calls a "dystopian future."[27] He illustrated the concept by comparing it to a science fiction movie where a person is sentenced for crimes yet to be committed. The belief in empirical data as the truth may take institutions to wrongly disregard students' personal and qualitative input, which today still bears value.

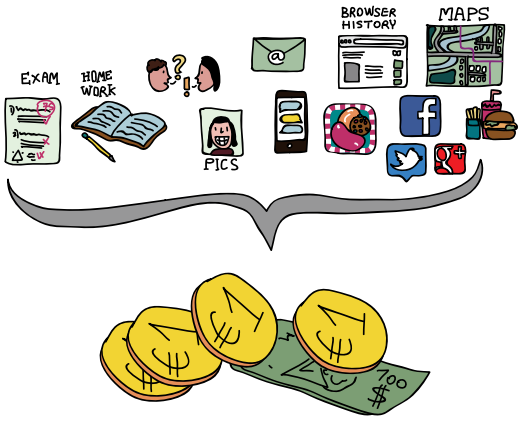
# Protecting students or playing G-D?

Access to information, and consequent knowledge acquisition, has for centuries been used and abused by the dominant power. Today, the possibilities Big Data are bringing to help us understand students' achievements and difficulties have to be carefully weighed against premature conclusions. Results can be reached and decisions taken on students' future based on misleading understanding of data. Data has to be given meaning, and predictions are only expectations even if they are based on data, especially when we are dealing with human behavior.

"When a learner's identity is something they define in their relationships with teachers and peers they have an element of choice in determining what kind of learner they are, and what kind of learner they might want to become. They can provide the context that makes sense of their data. They can challenge or resist others' interpretations of their actions and motives. In short, they have some control and voice over who they want to be as a learner.... But we need to consider the implications and consequences of using big data analytics as our main way of knowing about education. It tends to simplify big social and political questions about what kinds of learners we are and want to be, or how education should respond to major social and economic challenges, to a simple process of prescribing the next piece of educational software to download."[28]

# Mechanizing Education

Critics point to data-driven learning, not traditional learning, as a threat to turn schools into factories, due to the increasing digitization consequence of the agreements with for-profit companies that push their products on teachers and students.
Skepticism also arrives from the capacity of technology to assume functions such as diagnosing a student's strengths and weaknesses and adjusting materials and approaches to suit individual learners. Critics talk about the overweight being given to data instead of spending more on human resources[29].
"So it may be that children's sense of themselves as learners comes to be more dominated by visualizations of their educational data through apps, web profiles and infographics than through processes of reflection and dialogue. The ancient maxim to 'know thyself' becomes instead: 'measure thyself.' If the reliability of our knowledge rests on the extent that it can be backed up by big data, our learning profiles may be seen – both by others and ourselves – as more robust and objective descriptions of who we 'really' are, supplanting and dismissing our own messy, subjective self-knowledge[30]."




# Data as currency

Major internet players and communication service providers are already openly discussing "trading" the user profile data they own and manage. B. Shear, Innovation Insights, warns about the commercial by-products of educational software. "Of Google's $37.9 billion in 2011 revenue, 96 percent was earned from advertising. Is Google providing schools free access to its Google Apps for Education software in the hopes that it will eventually earn advertising revenue from data mining our children's digital school assignments and education-related interactions?"[31] At the beginning of 2013, Massachusetts became the first US state to ban companies that provide cloud computing services from processing student data for commercial purposes.
Non-profit funded platforms like inBloom are expecting to begin charging districts for their infrastructure usage starting 2015 (US$2-US$5 per student per year). In addition, application providers that will be riding their infrastructure and data cloud will be looking to gain their share of the value chain.
"Who owns the learning experience? Who owns all this education data? Companies? Schools? Instructors? Students? Do students know what data is being collected about them? How can we make sure that learning analytics and data mining aren't about extracting value but adding value? How do we make sure that in our rush to uncover insights from all this education data we now capture, that the student isn't just the object of analysis? How do we make sure the student has subjectivity, agency and control — over their data and their learning"?[32]

# Looking ahead

**Entrepreneurial initiatives are pulling the educational industry to be innovative and use data analysis to provoke significant pedagogical changes.** Many research groups, in Israel, are trying to influence the educational system by exploring the latest analytical and technological developments. As for example the possibilities brought by sensors which track the students' movements white interacting with the device, or the latest artificial intelligence models to develop learning systems which are relevant to the newest generations' new ways of interacting, communicating and sharing information.

Trying to go a step further in adaptive learning and base the system on the student's knowledge rather than attainment levels, Sr. S.Hershkovitz of the Center for Educational Technology together with Ernest Lyubich, are currently exploring machine learning models to develop an interactive Math course where instruction can be adapted according to each individual student's responses.

Trying to integrate the newest generations learning contexts as collaboration and social media, Professor Koby Gal and his team at Ben Gurion University, are initiating a project through exploring the techniques and models from both artificial intelligence and the learning sciences. The multi-disciplinary project develops technologies to analyze and support collaborative learning across different technology-enhanced environments, both inside and outside the classroom, in the context of different types of ubiquitous social media (e.g., social networking sites such as Facebook and Wikipedia), and scaling-up the benefits of group learning from very small groups to large group sizes and long-term interactions.

Trying to integrate the newest techniques of perceptual computing (visual - eye tracking or 3-d gesture recognition, speech, emotion recognition, etc.) in learning adaptive systems, a research-project at Intel, led by Shahar Shpiegelman, explores the possibilities of new technologies on big data, predictive analytics and perceptual computing as source of information to understand better students' learning performance and learning patterns. The project intends to create a system that combines data related to the physical interaction of the students with a computer and contextual data from the tutoring system to learn specific academic units. By using learning analytics, the system will provide information that can help orient the way the computerized lesson responds to each student. The personalization is based on a real-time data gathering and response while the student is interacting with the lesson.

1. http://www.wired.co.uk/magazine/archive/2012/11/features/open-university?page=all
2. http://www.huffingtonpost.com/andreas-schleicher/big-data-and-pisa_b_3633558.html
3. http://www.knewton.com/
4. http://www.aleks.com/
5. http://uki.blackboard.com/sites/international/globalmaster/Platforms/
6. http://www.ellucian.com/
7. http://dmlcentral.net/blog/lyndsay-grant/understanding-education-through-big-data
8. https://www.inbloom.org/
9. http://1776dc.com/2013/12/13/how-big-data-is-changing-the-educational-frontier/
10. New Media Consortium Horizon Report 2013.
11. http://www.columbia.edu/~rsb2162/BakerSiemensHandbook2013.pdf
12. http://dmlcentral.net/blog/lyndsay-grant/understanding-education-through-big-data
13. http://www.dreambox.com/white-papers/the-future-of-learning
14. http://www.dreambox.com/
15. http://www.eltjam.com/big-data-and-adaptive-learning-in-elt-knewton-interview-part-1/?utm_source=linkedin&utm_medium=social&utm_content=3190976#%21
16. Frank Catalano, How Will Student Data Be Used? GeekWire, July 3, 2012.
17. http://hackeducation.com/2013/10/17/student-data-is-the-new-oil/
18. http://www.eschoolnews.com/2013/12/23/schools-monitor-media-400/2/
19. http://www.ted.com/talks/daphne_koller_what_we_re_learning_from_online_education.html
20. http://www.skilledup.com/blog/mooc-data/
21. http://www.huffingtonpost.com/andreas-schleicher/big-data-and-pisa_b_3633558.html
22. http://www.nasa.gov/press/2013/november/nasa-brings-earth-science-big-data-to-the-cloud-with-amazon-web-services/#.UoNKIfIT5ZA
23. https://schoolzilla.org/
24. Visualization: The simple way to simplify Big Data. Chris Taylor. Wired. 8.26.13
25. http://1776dc.com/2013/12/13/how-big-data-is-changing-the-educational-frontier/
26. Dropout Indicators Found for 1st Graders, By Sarah D. Sparks, edweek.org, 07/29/2013
27. http://www.timeshighereducation.co.uk/news/big-data-could-create-dystopian-future-for-students/2010061.article
28. http://dmlcentral.net/blog/lyndsay-grant/understanding-education-through-big-data
29. Scientific American August 2013
30. http://dmlcentral.net/blog/lyndsay-grant/understanding-education-through-big-data
31. http://insights.wired.com/profiles/blogs/bill-to-ban-data-mining-of-student-email#axzz2oelHtmKm
32. http://hackeducation.com/2012/12/09/top-ed-tech-trends-of-2012-education-data-and-learning-analytics/

# New Data: Privacy and Awareness of online environments

# New Data: Privacy and Awareness of online environments


"Google goes to court over Gmail scanning" (The Telegraph, Sept. 2013)[1] ;"Facebook sued for scanning 'private' messages for profit" (Wired, Jan. 2014)[2] "LinkedIn is breaking into user emails, spamming contacts – lawsuit" (GigaOm, Sept. 2013)[3]; "We have sensors that track us everywhere we go. Think about what this means for the privacy of the average person" (Edward Snowden, TV, Dec. 25, 2013)[4]; "Did you know that your 'likes' in Facebook could expose intimate details about you as well as personality traits you might not want to share with anyone?" (How Big Data Analytics reveal your most intimate secrets[5]). The year 2013 was full of striking headlines about online privacy and the use and trade of individual information without consent. **Not less dramatic are the headlines affecting the educational world as the massive cyber-attack in California involving its universities[6], or the recent recognition by Google that it does data mine student emails for ad-targeting purposes in its Google Apps for Education[7]** .At the same time that there is a significant increase in investments and products (massive flood of educational apps becoming a major learning resource[8]) based on Big Data systems in educational settings. Occurrences that have raised a red alert in the entire educational community. Teachers and parents, especially, are worried about the use and misuse of students' data. "Student Data is the New Oil"[9] is a statement gaining popularity among the educational media. On the other hand, the undeniable benefits for the entire educational community, as we saw in the previous chapters, of

learning systems based on Big Data, creates a state of uneasiness and doubt!

In the beginning of 2013 a big fuss arose against inBloom, which has turned into a legal suit[10] and the withdrawal of several US states from the project of creating an important educational data storage based on a cloud run by Amazon.com, with an operating system created by News Corporation – Amplify. inBloom declared its plans to share the data with non-profit as well as for-profit vendors with state and district consent. "Parents, teachers, advocacy groups and privacy experts throughout the country have protested this unprecedented plan to share children's sensitive information with private corporations and for-profit vendors. New York organizations opposing this data mining include Class Size Matters, the Learning Disability Association of New York, Alliance for Quality Education, New York State Allies for Education and the Coalition for Educational Justice. These groups have pointed out that a breach of this highly sensitive information, or its inappropriate use, could put children's safety at risk."[11]

A study carried out by Common Sense Media, an organization that rates EdTech products for their usefulness and appropriateness, showed that most of the mobile apps for kids collect personal information and share it with commercial providers without parents' knowledge, which led to a pledge to major companies offering EdTech such as Google, Pearson, Scholastic, and Samsung, to make sure student data is used for educational purposes only and not for marketing.[12]

The development of ways for tracking physical and emotional related data and its use by the education system raises even more questions about invasion of students' privacy. "The conversation on privacy will need to change dramatically in the near future. It will not be long before you will be able to take a picture of someone with your phone camera and have software that can impute regions of that person's genomic DNA…self-trackers want to use these sensors…, to equip us with new ways to hear our bodies."[13]

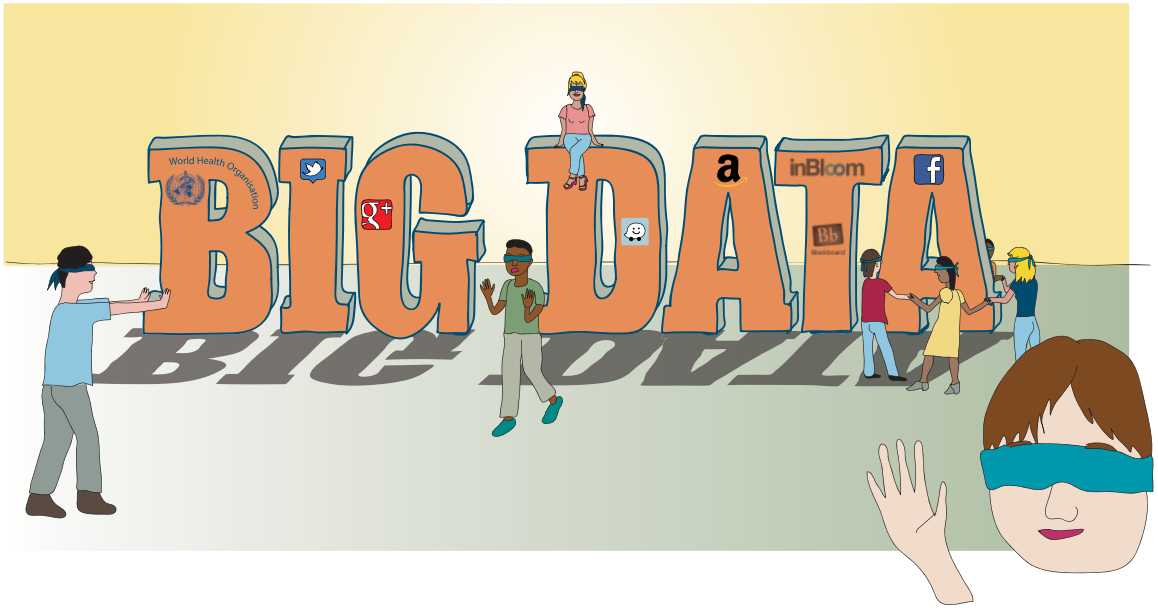
**Will the educational community worries obstruct the implementation of data-driven developments in educational settings?** Researchers and educational technology experts maintain that awareness about the real use of data is the key issue. A major concern about the implementation of programs which require the use of students' data is the "confusion" or "lack of specific knowledge" about all the technology advancements in Big Data. According to David Rubin, attorney for the US Council of School Attorneys, one of the biggest challenges to protecting data privacy in the cloud is the lack of understanding by school boards and district superintendents. "You start to talk to them about data privacy and cloud warehousing and you see their eyes glaze over. With so much jargon it's easy to say 'it's a problem for IT,' but everyone should be well-versed in data privacy."[14]

The growing use of Big Data in educational settings and concomitant lack of awareness is an even higher concern with the invasive online environments such as social networks and mobile apps. "AT&T, Verizon, Facebook and Google sell their customers usage data (location, web browsing history, etc.). They also provide ways of 'opting out,' if the customer is aware and knows how to do it."[15]

Digital users provide personal data when they go online, sometimes knowingly and other times without realizing that they are providing it to third parties, and quite often, they do not realize

that they are part of an online information industry. A lack of trust and understanding among users of the destination of their data could become a barrier to the continued development of innovative ventures. This is especially true within educational settings dealing with minors. A study on online privacy and awareness in the UK suggests that the decisions consumers make are influenced by how direct they perceive the risks and benefits to be, strengthening the importance of awareness and privacy attitudes when taking decisions about online data.[16]
There are many myths about misuse of data,[17] and only if the educational community members are well informed can they choose when and where students' information should be used, or even when it is relevant to fight for protection of educational data by policy makers. The type of data used by the systems specializing in education should be carefully selected by the relevant players in order to protect the student's individual privacy.

# Survey on Online Data Privacy and Awareness




The education community's awareness about the use and misuse of online data, and the importance they give to privacy online, strongly influence the decisions they make about the implementation of systems based on students' data. In order to shed some light on this matter, a survey was conducted looking at the general population as well as sub-groups of teachers and students. The sample was asked about their knowledge of and concern over online information of popular online environments such as social networks, mobile apps, search engines, and specific areas such as health and educational systems.

**A total of 1,877 Israelis were surveyed and the results showed that the majority are concerned about their privacy online,** with the commercial use of private information by mobile companies being the area of greatest concern.
The results showed a lack of awareness by the respondents in most areas surveyed, except on questions related to the information used by social networks.

# Methodology

The survey was carried out in two phases: the first looked at the general population, and the second focused on the educational community (students and teachers).
Subjects answered a group of questions about awareness (true or false statements), and another group of questions about privacy concerns (5-point scale from strongly agree to strongly disagree statements).
Further details about the methodology can be requested from MindCET@cet.ac.il or ceciliaw@cet.ac.il.

## How often do you connect to social networks?

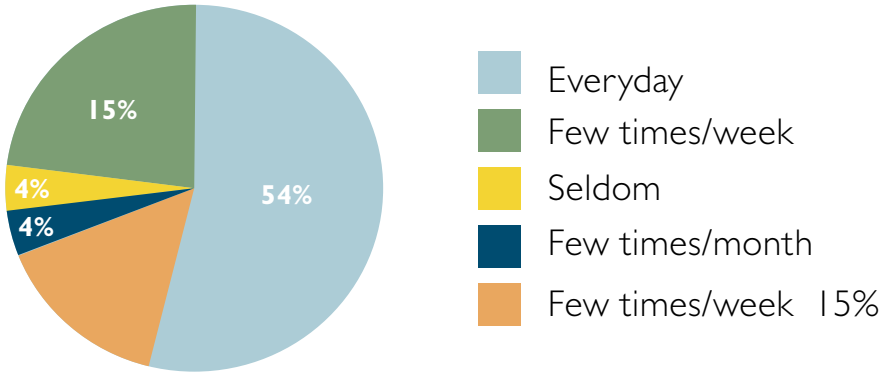
| Frequency of use of Social Networks | | | | | | |
|---|---|---|---|---|---|---|
| gender | | AGE | | | | |
| MALE | FEMALE | 18-29 | 30-39 | 40-49 | 50-60 | |
| 69% | 70% | 85% | 65% | 62% | 55% | Every day or a few times per week |

# General Population

Telephone interviews were conducted from 10 December 2013 to 1 January 2014 among Israelis aged 18-60, forming a sample of 1,000 subjects. To be representative of the Israeli population, the data was weighted in sex and age according to the true proportion in the Israeli population. The sample included 21% of non-Jews and it was proportionally distributed throughout all areas of Israel.
Fifty-four percent of the sample used social networks every day, 23% never, and, as expected, the younger the age the higher the frequency of use.

## Men and women reported the same frequency of use of social networks.

The results suggest that age and gender are independent variables that significantly affect many of the variables surveyed.



# Age

Age affected significantly 7/10 awareness statements; however, different trends were found in different questions.The younger subjects showed less awareness about questions related to the commercial use by mobile companies; however, they showed higher awareness about Google's policies on pictures and personal information, about information exposure on social networks like Facebook, and about information shared in online messengers such as WhatsApp.

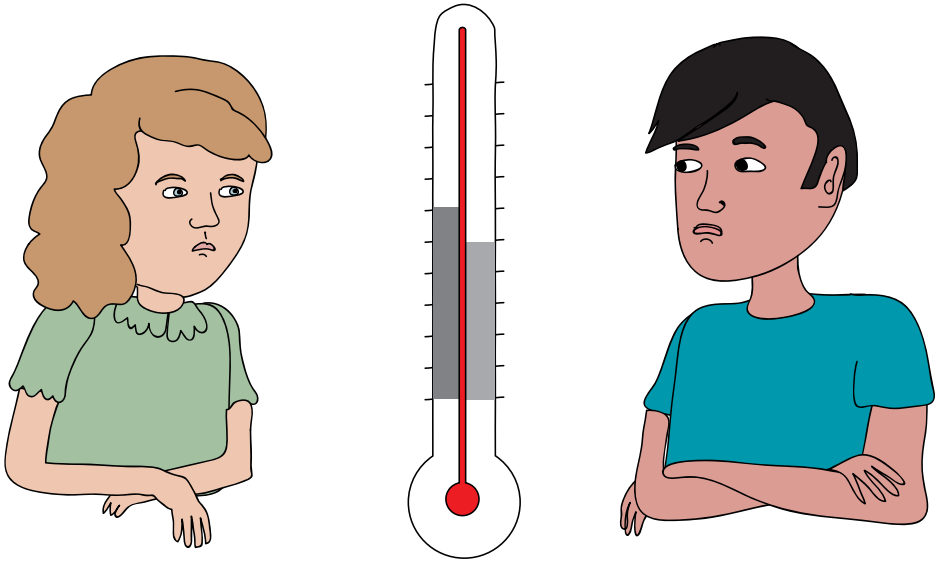
Younger people (under 29 years of age) were less concerned about online privacy issues.
Age affected significantly the perception that one can be anonymous online – the older you are the less you trust you can be anonymous.

# Gender

Men were more aware than women in 5 /10 questions about the use of online data, especially on questions related to mobile companies and apps. Gender affected significantly 8/10 privacy concern statements, with the women always the ones showing greater concern in the different areas surveyed.

Women responded that they are more concerned (60%) about their privacy online compared to male subjects (50%). A higher percentage of men said they do not care if anyone has access to their content on social networks; they believe significantly more than women that the opportunity to share a network environment with other people extends their horizons; and they are also less worried about information they share in WhatsApp, Facebook and Twitter. Significantly, more women said they comment online only if they can do it anonymously; they make significantly more use of their privacy settings on Facebook to limit access to their posts; they are much more concerned about networking than men are; and they are more concerned about uploading their pictures.
Among the group of respondents who play online games, sixty-five percent (65%) of women do not play with players they do not know compared to forty-four percent (44%) of men.

Women responded that they are much more concerned about their privacy online compared to male subjects



## Awareness

Respondents showed higher awareness on the questions related to the commercial use of information by social networks (77% of subjects showed awareness about privacy settings functions and 60% about customizing advertisements based on personal data), and by mobile companies/apps (where 55% and 60% of subjects showed awareness of data being transferred to third parties). In all other areas a large part of respondents did not show awareness, either by believing a false statement to be true, or by answering "don't know."

The questions about online environments where users have a personalized entry (username, telephone number) were the ones about which respondents were less aware: only 25% were aware that the information sent through online messenger companies (such as WhatsApp) or emails is not exclusive to the intended target; only 36% were aware of the personalized information at websites such as Google (pictures, personal data, etc.).

On the other hand, only 22% of respondents believe it is possible to be completely anonymous online.

## Privacy

55% of respondents expressed their strong concern about privacy online and 24% said they are not worried.

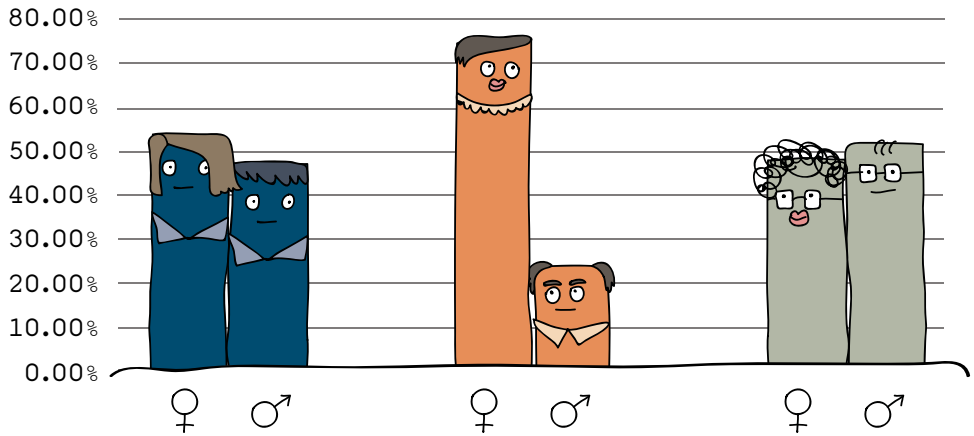
On the questions related to the educational system, 34% agree and 48% disagree that it should use or have free access to students' personal information or to the students' social networks. The majority (61%) totally disagree that the health system should be allowed to use their personal data, even if it is for research related to health improvements. Seventy-two percent disagree on the use of their cellphone data by companies, even if it is to offer them good deals. On questions related to sharing personal information, 68% said they restrict the access of their pictures uploaded online, 71% said they restrict the access to their Facebook posts, 50% said they are concerned about sharing information on social networks. Only 36% said they only comment online (blogs, videos, etc.) if it is anonymously.

# Education Community

## How do teachers and students compare to the general population in terms of awareness and privacy?

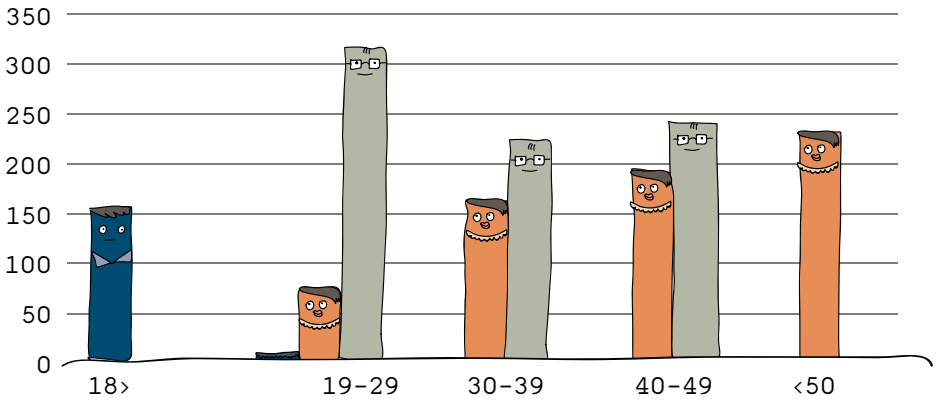
In order to answer this question, data was collected from students and teachers, separately, and the results compared to the general population data, forming a new sample of 1,887 questionnaires, sub-divided into: students (N=156), teachers (N=721) and general population (N= 918[18]). The data was collected

### Gender distribution



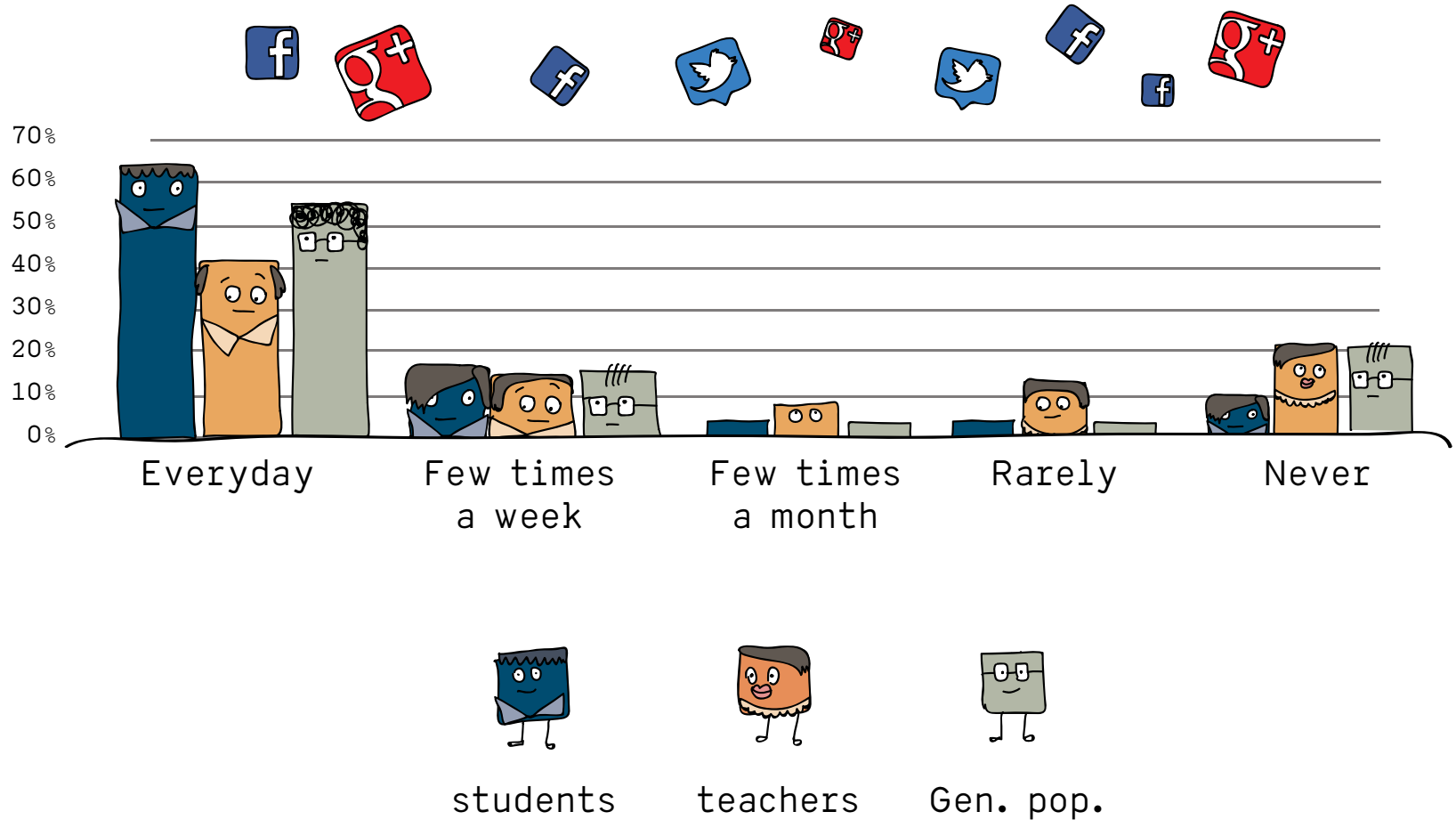

students teachers Gen. pop.

### Age distribution

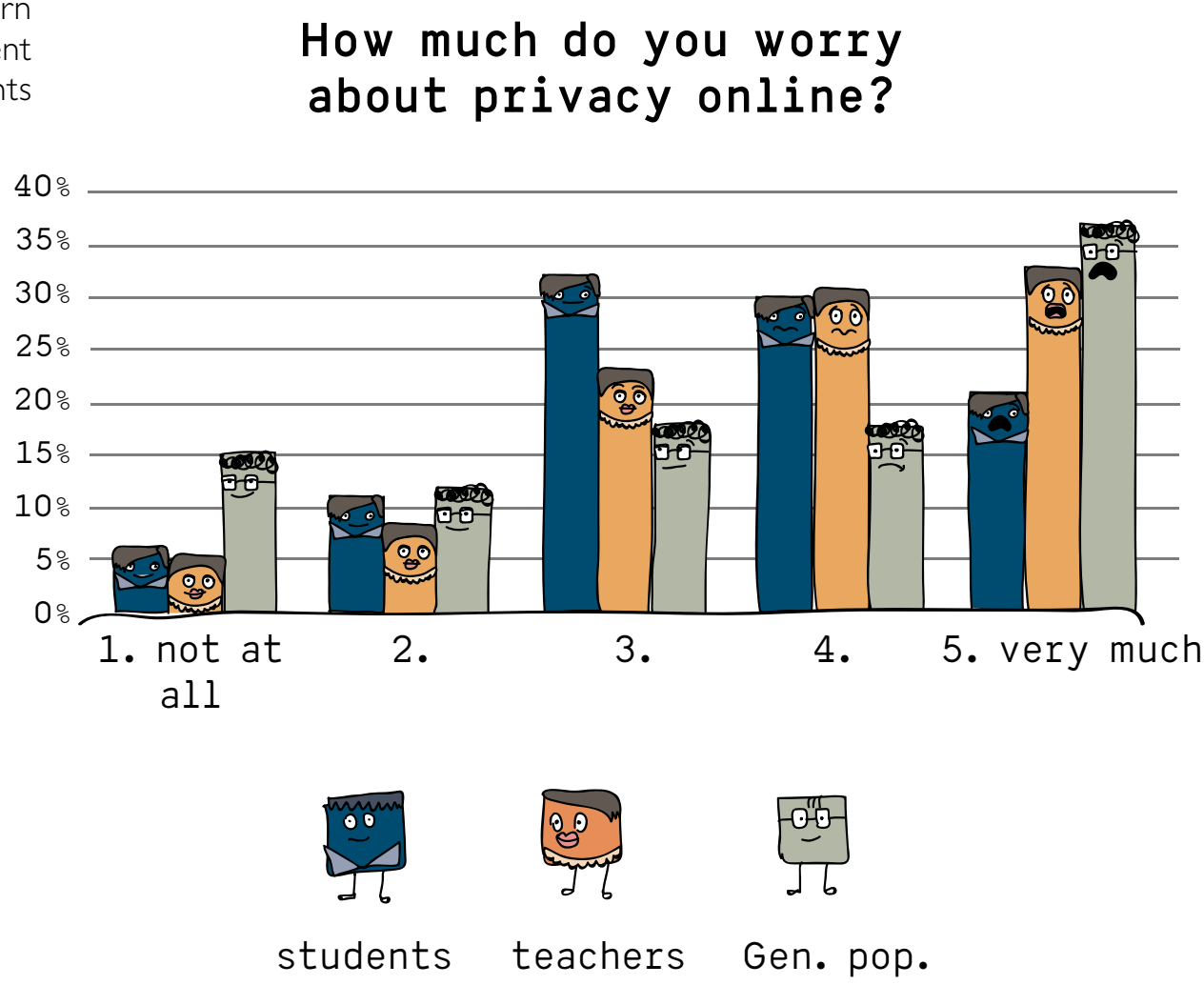



students teachers Gen. pop.

by phone interviews (general population), by email (teachers) or by voluntarily clicking on a call published on an educational publisher platform (students). The age (youngest 14 and oldest



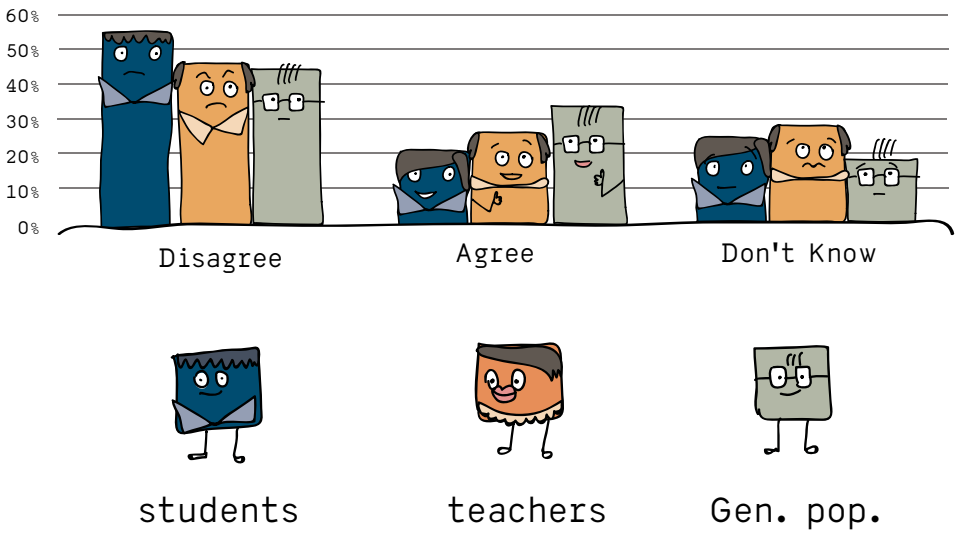
63) and gender distributions are shown on the graphs. The differences in the age and gender distribution are justified by the sampling.
Teachers showed the lowest percentage of social network frequent users (22% never use it) and students the highest (82% every day or weekly).
The three groups, separately, expressed concern over most of the areas surveyed, giving different priorities to each of them. Comparatively, students



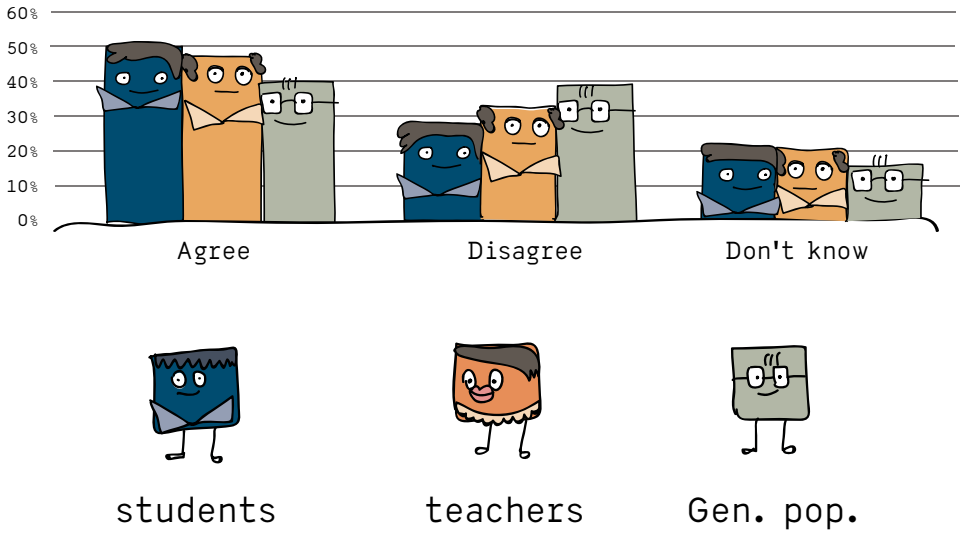
were significantly less concerned about privacy online, except on school-related use of data. As shown on the graph about privacy online in general, 50% of the students are not worried at all to moderately worried, while 63% of teachers and 60% of the general population are very worried.
Students and teachers answered differently from the general population on the questions related to education.

**The majority of students disagree that schools should use their social networks information even if it is to improve their learning;**
teachers and students show a similar response trend on this question, which is different from the general population. Students

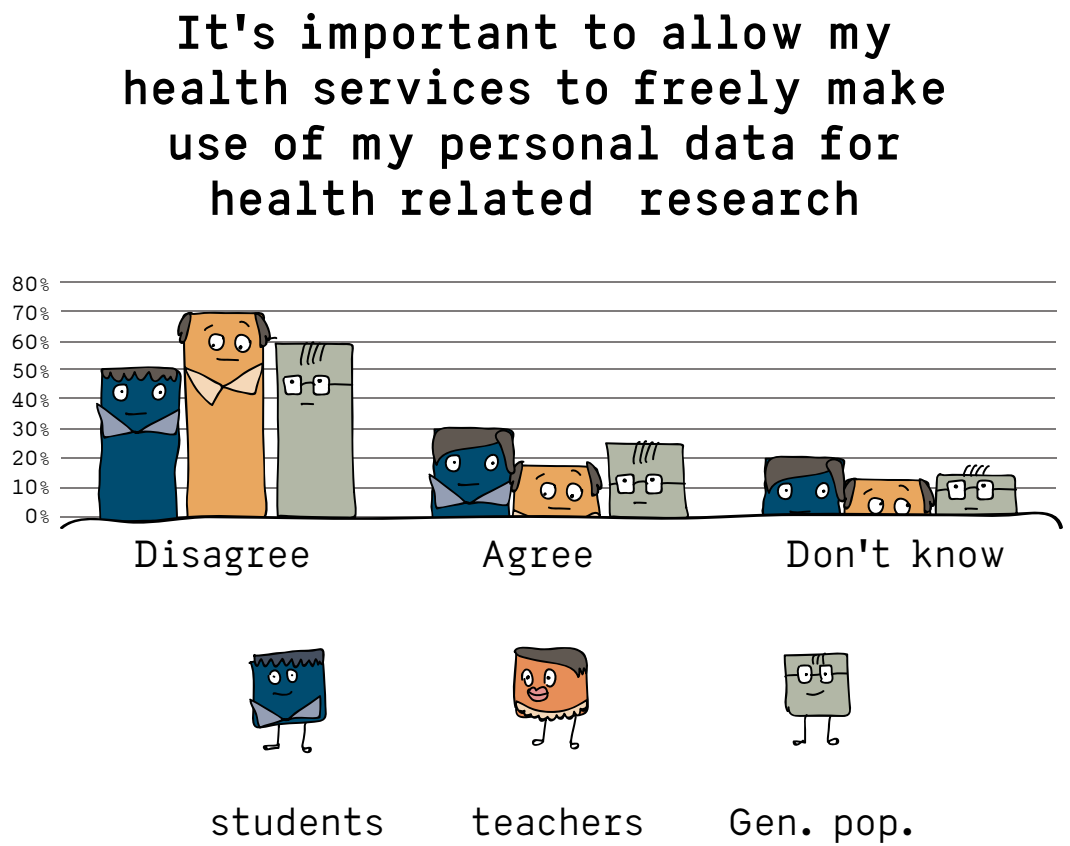
**Schools should use the information that students publish on social networks to improve their learning**




**The educational system should not be allowed free access to student's personal data held by the school**

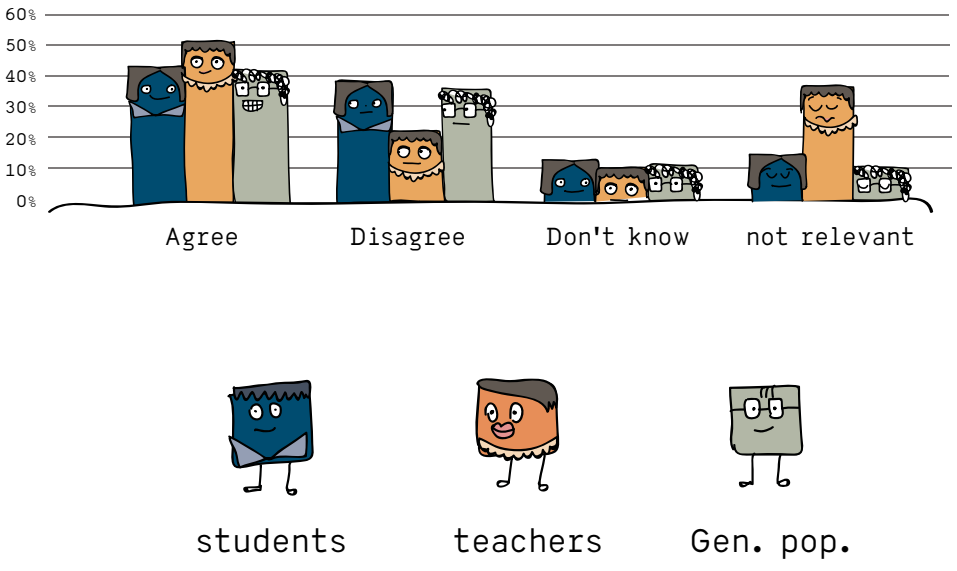


and teachers also have similar attitudes towards allowing the educational system free access to student information, to which they mainly disagree. It is interesting to note the amount of "don't know" answers, especially from students and teachers who are currently significantly involved with the educational system.
Neither of the groups agrees to give freedom to the health care system to use personal information, even if it is for research purposes. **A significantly greater number of teachers were more concerned with the use of data by the health system (70%) than the educational system (47%). Students, on the other hand, were consistent in showing similar attitude for both systems;** 50% do not agree with free use of their personal information by the Health system and 55% by the Educational system.

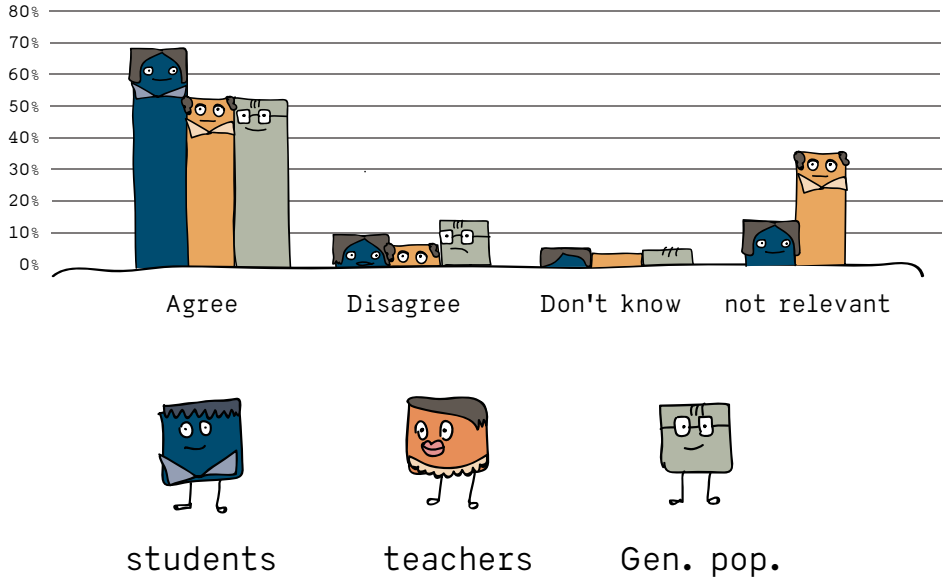


Students showed a certain ambivalence in their level of concern about their privacy online. They expressed lack of concern about playing with players they do not know or about providing personal comments online; however, they were divided in their concern over sharing information on social networks. Sixty-three percent of students showed concern about uploading their pictures and most of them (69%) use Facebook privacy settings to restrict the information to be shared.

**When I contact friends on social networks I do not worry about who else can have access to that content**




**I use Facebook privacy settings to restrict who can read my post**

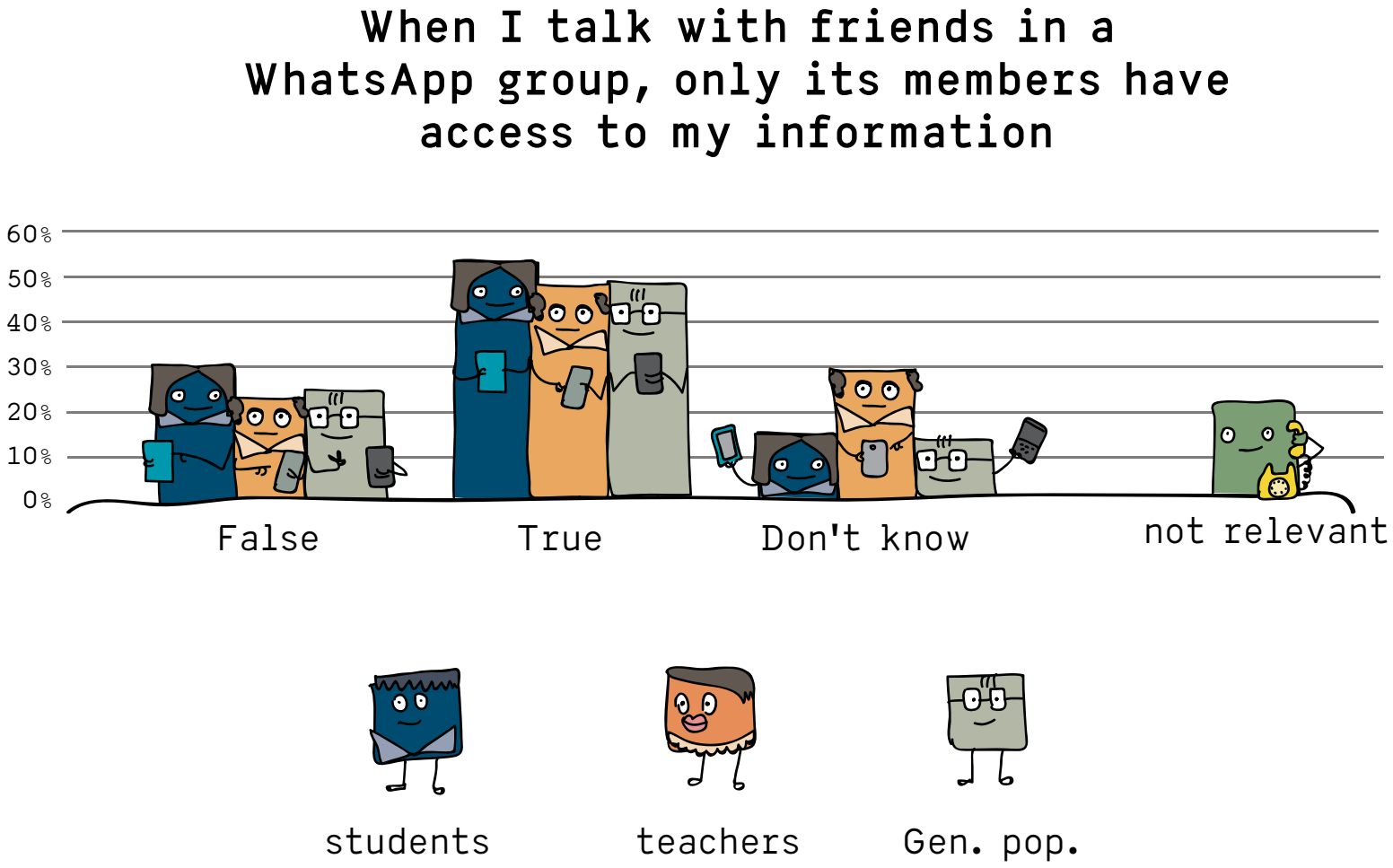


Teachers' greatest concern (78%) is the use by mobile phone companies of their personal information for commercial purposes, a concern shared by the majority of the other two groups (67% of students and 71% of the general population).
It is interesting that the results showed that there is no specific trend about being anonymous when commenting online, being very similar in the amount of people who care and who do not care.

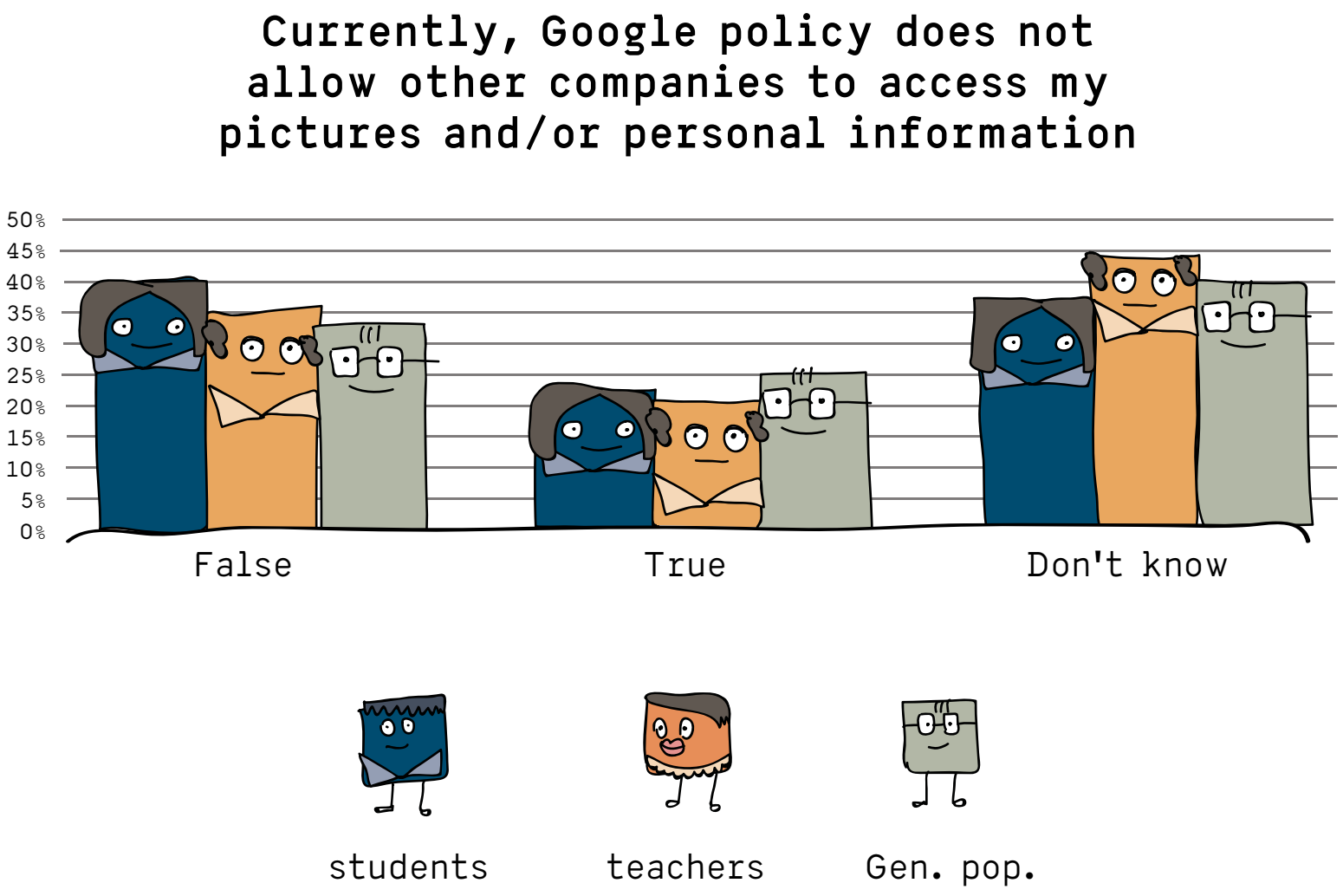
## Teachers showed a greater level of awareness about use of data online in 5/10 questions compared to the other two groups.

On the other 5 questions, students showed highest awareness compared to the other two groups, taking into consideration that all three groups comprised a low percentage of individuals who showed awareness.
Teachers were significantly more aware than the other two groups on the questions related to the commercial use of the data on mobile phones; for example, 73% of teachers and only 59% of the general population and 49% of students were aware that phone companies can make use of personal data for their own use or to transfer to third parties.

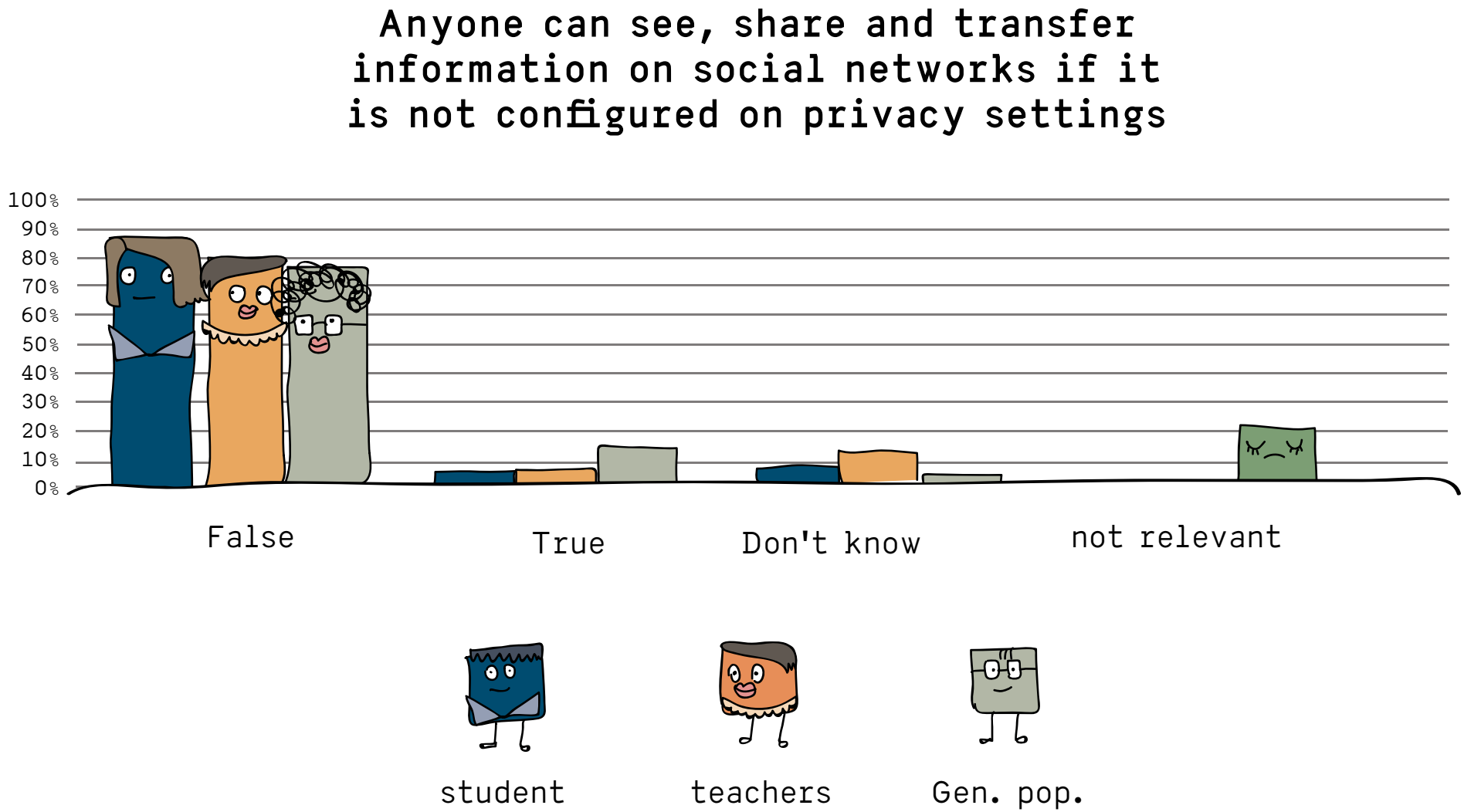
Teachers and students showed high awareness compared to the general population on the questions related to social network exposure and use of information. Students showed a slightly higher awareness than teachers about disclosure of the information transmitted through online messages and emails.



A significant number of the subjects answered "don't know" on the question about Google policies allowing other companies to have access to pictures or other personal information, and **teachers were significantly more aware that the information provided to websites which require personal entrance are not exclusive.**



The questions where all groups showed high awareness were the ones related to social networks (Facebook use of the information to customize advertisements, and the use of privacy settings).

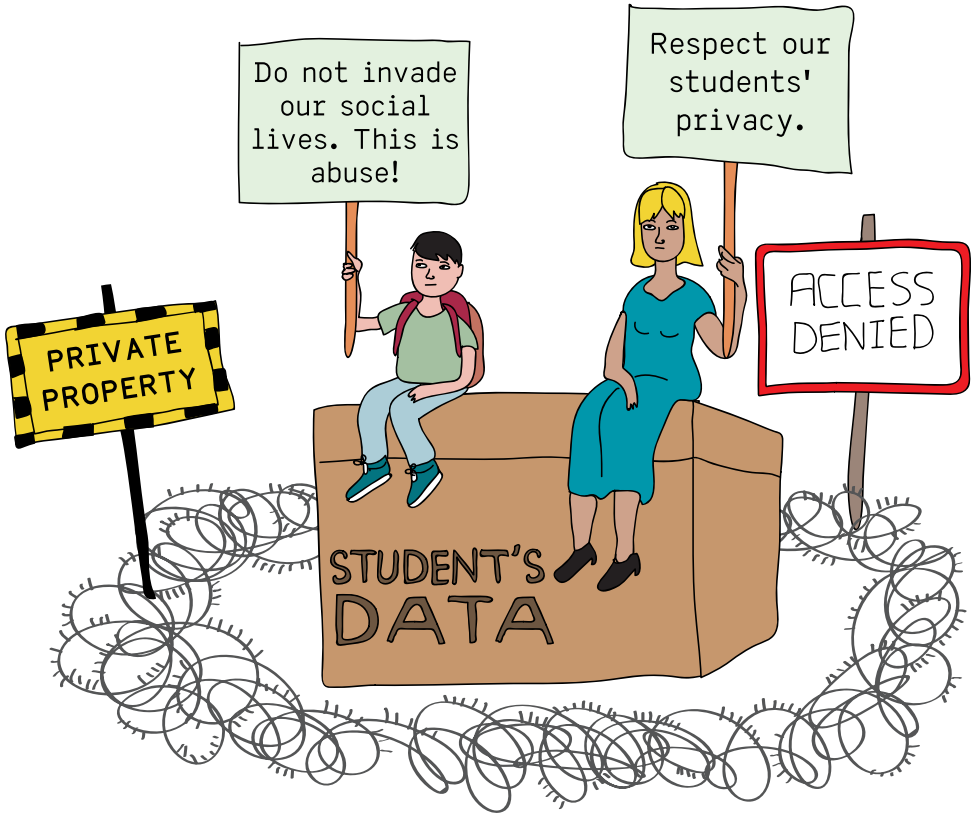


# Conclusions:

The survey clearly demonstrates the concern of the Israeli population over online use of personal data; compared to a study[19] of 10,354 adults from all over the world, Israelis showed a much higher percentage of people concerned (55% of Israelis compared to 37% of subjects from around the world). Women were consistently more concerned than men, and age did not show a clear trend in all privacy concern questions, a result reproduced in other surveys around the world. This result suggests that the impact and complexity of the online world on any user's life, and especially the younger generation's, lead to differentiated attitudes.

All adults (over 18 years of age) are more concerned about personal data being transferred to other parties by the health services, even if it is for research, than by the educational system. Only the group of students (under 18 years of age) were consistent in not agreeing with either system freely using personal data.

The survey also suggests that there is either a lack of awareness or a significant suspicion (amount of "don't know" answers) about the use of personal information in online environments. The amount of media headlines alerting about abuse or misuse of personal online data and the lack of clear, easy to read, privacy policies, could explain the "confusion" of the respondents.



**A clear statement is made by teachers and students against the free use of students' data by the educational system, even if it is to improve learning.** This result suggests that there is a significant concern by the educational community regarding the implementation of teaching systems, such as the ones mentioned in this report, that require the collection of students' data. Moreover, the results suggest a concern about the boundaries of what the educational system can consider as students' data.

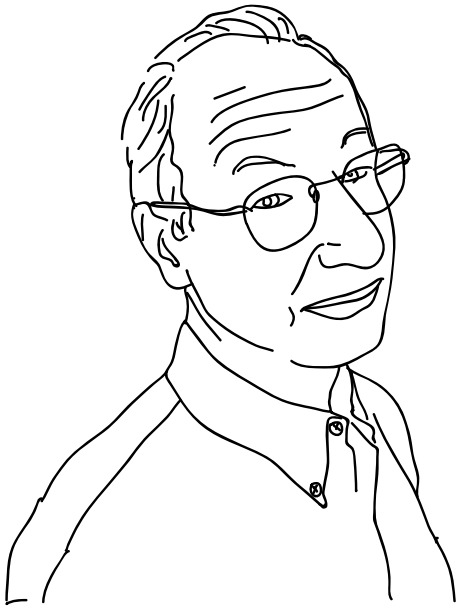
Only teachers were consistent in showing high awareness about the different online companies passing information to third parties; the rest of the sample showed an average of 45% of people aware. Interestingly, the question specifically mentioning Google, the largest online company in the world (Wired, 2014), suggests that the respondents were doubtful , since almost half answered, "I don't know." Earlier studies suggest that users are less bothered by information used by companies they trust . Therefore, this result may indicate that the extent that Google has pervaded users' online lives, as an information and communication main channel, has led to either a comfortable lack of awareness, or a worrisome awareness from the user side.
Our survey reproduced results found in other surveys such as the high awareness about privacy settings on social networks, and that the use of data by mobile phones is less known than by social networks.
The results do not suggest that the younger generation is more aware of online data use, but that teachers are!

1. http://www.telegraph.co.uk/technology/google/10289798/Google-goes-to-court-over-Gmail-scanning.html
2. http://www.wired.co.uk/news/archive/2014-01/03/facebook-private-messaging-lawsuit
3. http://gigaom.com/2013/09/21/linkedin-is-breaking-into-user-emails-spamming-contacts-lawsuit/
4. http://www.dailymail.co.uk/news/article-2529236/US-whistleblower-Edward-Snowden-delivers-alternative-Christmas-message.html
5. http://smartdatacollective.com/bernardmarr/129421/how-big-data-analytics-Facebook-likes-reveal-your-most-intimate-secrets
6. http://www.ecampusnews.com/top-news/universities-student-privacy-444/
7. http://safegov.org/2014/1/31/google-admits-data-mining-student-emails-in-its-free-education-apps?body=http://safegov.org/2014/1/31/google-admits-data-mining-student-emails-in-its-free-education-apps
8. http://www.eschoolnews.com/2014/02/19/resources-education-apps-722/?ps=CeciliaW@cet.ac.il-00130000135NyG-0033000001CbsTP
9. http://hackeducation.com/2013/10/17/student-data-is-the-new-oil/
10. http://blogs.edweek.org/edweek/marketplacek12/2013/12/new_york_battle_over_inBloom_data_privacy_heading_to_court.html
11. http://www.wnyc.org/story/307074-what-you-need-know-about-inbloom-student-database/
12. http://www.nytimes.com/2013/10/14/technology/concerns-arise-over-privacy-of-schoolchildrens-data.html?_r=0
13. http://www.wired.com/wiredscience/2012/10/big-data-is-transforming-healthcare/
14. http://www.eschoolnews.com/2013/11/20/data-privacy-cloud-133/4/
15. http://techcrunch.com/2013/07/05/att-considers-selling-your-browsing-history-location-and-more-to-advertisers-heres-how-to-opt-out/
16. Online personal data: the consumer perspective, 2011
17. http://www.eschoolnews.com/2013/11/20/data-privacy-cloud-133/
18. To be able to compare the three groups, all teachers were taken out of the original general population sample (N=1000).
19. ComRes polling and research consultancy. March 2013. Big Brother Watch Online Privacy Survey Online Personal Data the Consumer Perspective, UK, 2011

# Big Data & Education:
# Do you see big opportunities and/or big threats?

MindCet

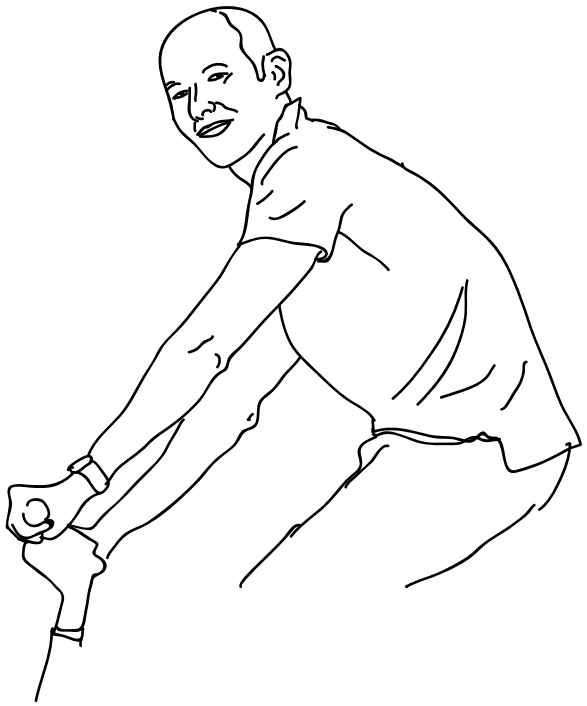
# Asking the experts!



## David Weinberger:

### Director of Harvard Media Lab

It's a tremendous opportunity for students to see further outside the Black Box of ideas that arrive as if from on high. As that data becomes Big Data, and especially as the tools for interacting with Big Data become easier and more familiar, the wall between learning and researching will be further eroded. As Big Data becomes a part of a worldwide knowledge commons, students will further see the power not only of Big Data but of social collaboration around data, ideas, and knowledge.



# Kobi Gal:

## Director of Human-Computer Decision-Making Dept., Ben-Gurion University

Thanks to Big Data we are in a position unlike ever before to revolutionize education. This is due to the following two developments. On the one hand, learning is increasingly happening online or mediated by educational software, creating a wealth of information in terms of academic content and students' interactions with software, with teachers, and with each other. On the other hand, developments in artificial intelligence and machine learning have produced algorithms with the abilities to infer intricate trends and patterns from large-scale data. Combining these two aspects will provide transformative educational tools for the benefit of students and teachers. Such technologies will consist of: (1) educational software that support model construction, exploration, and trial-and-error, providing a rich educational environment for students; (2) intelligent algorithms for analyzing students' interactions with such software in real time, inferring their activities with the software; (3) machine-generated support that guides students' interactions to maximize their learning while minimizing intervention; and (4) novel visualization tools for providing teachers with real-time assessment of students' work, enhancing their understanding of students' learning processes while minimizing the effort required.

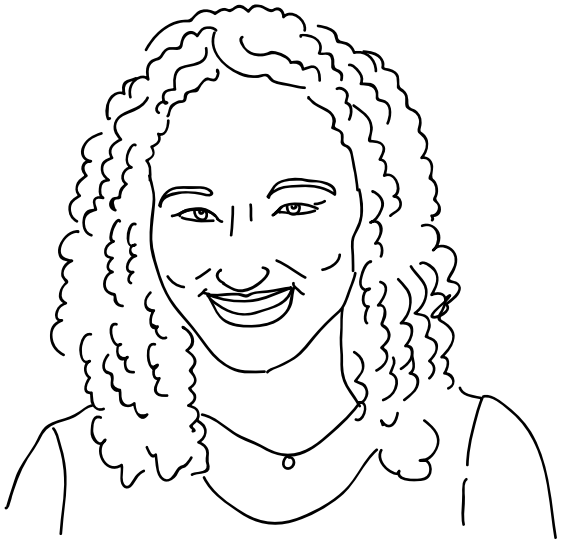
# Leon Markovitz:

## Entrepreneur, co-founder of WikiBrains

Nearly 90% of world data has been generated in the past 2 years. For better or worse, it is a revolution that has only just begun. Every company today is embracing the data behind their numbers and statistics like never before. Computers continue to get better at analyzing Big Data and finding patterns, and as more people continue to embrace it, the insights we find will continue to blow us away – in a positive way. Today, everything you do online, and most of what you do offline, is being measured. What you like, follow, and listen to, as well as your purchasing habits, daily routines, and coming soon…much, much more. The purpose of predictive technologies is to provide companies and their clients with the service they need when they need it. It is now possible to use the power of Big Data analytics to prevent problems and tackle them before they pose a risk! It is this awareness that inspired the creation of WikiBrains. A biologically inspired algorithm trying to imitate the way your brain works –by creating patterns. Each topic represents a neuron, and each connection to another topic is like a synapse. Individual brain cells are indifferent to the wider concept, but the pathway that they help trigger results in an emergent behavior we may refer to as understanding. By having millions of minds contributing associations and creating maps, we can help others quickly discover interconnections to get new insights and understanding.

Isaiah Berlin said "To understand is to perceive patterns", and WikiBrains goal is to facilitate this to the crowds –the beauty is that the more people participate, the smarter it gets. You do not need to rely on yourself, you can leverage the connections/ patterns others are creating and benefit from them in a positive feedback loop.

A website like WikiBrains has the potential to redefine creativity and lead to faster insights in individuals and organizations. WikiBrains, in essence, is a graph that finds interrelations between apparently separate points of information. It helps you quickly see the connections that you may not have known existed. If you are a writer, WikiBrains serves as a window into the world brain -helping you identify what different demographics associate to different topics across time. If you are an organization, it helps you visualize how everything and everyone is connected, while eliminating the unnecessary noise.

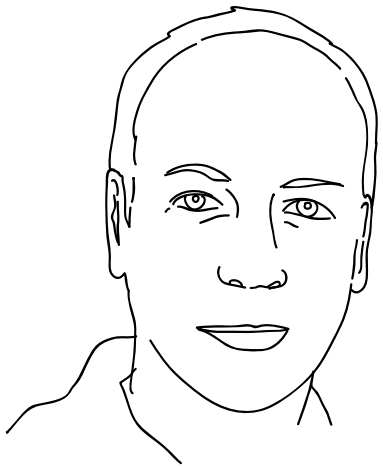


# Orly Fuhrman:

## Entrepreneur, co-founder of Lingua.ly

The Big Data revolution is touching many aspects of our online lives today, and presents an even bigger opportunity when it comes to the way we learn. There is wide agreement that it's time for a real revolution in education: schools worldwide still follow principles that were shaped in the Industrial Revolution era, with standard curriculums and teachers that are mostly engaged in conveying content to overcrowded classes. In such a setting, it is easy to overlook individual preferences, differences, and special needs. No wonder so many students lose interest in their course work: it's either too challenging, too easy, or just boring.

Big Data holds great promise in solving this problem and personalizing education. We can already see Big Data tools integrated in testing systems and courses today (as in Desire2Learn, Knewton, Benchprep, and more), allowing for a faster and more accurate evaluation of students' learning styles, strengths, and weaknesses. There is still a curriculum, with lessons and content that is carefully selected, but each student navigates it in a customized way that is far more engaging and motivating than the 'one size fits all' lesson plan. Lingua.ly takes personalized learning even further: it uses big data tools to take learning away from the limited scope of a course, textbook, or curriculum, and onto the web. There is no lesson plan or assigned texts: students can learn any time, everywhere they go on the web: online newspapers, blogs, email, Facebook, and more. Lingua.ly collects fine-grained information about each student, to keep track of his or her vocabulary, level, and progress. Then, it helps them practice and finds real content from around the web that is customized to their individual profiles. In a teaching context, this approach frees the instructor to focus on the higher-level aspects of learning, rather than on curation, marking, and tracking progress.



# Nathan Intrator:

## Professor at the School of Computer Science and Neuroscience, Tel Aviv University.

Big Data for me means the ability to efficiently create models which rely on large amounts of data and are far better than what we could achieve few years ago. Moreover, it is the ability to individualize models to different data clusters, or human populations, and therefore to correctly account for differences between groups. This is important in personalized education and in medicine for diagnosis and treatment.

# Shahar Shpigelman:

## Program manager, Advanced analytics @Intel.

Not only is Big Data not a threat, it is one of the biggest opportunities in education. Big Data will enable the personalization of education to the specific needs of the students; it will also provide the teachers with the tools they need to provide better education for each of their students. Just imagine a world where a teacher/a parent can see the whole learning path of his student/children from kindergarten and have a clear understanding of the students' knowledge and gaps.

# Omri Mendels:

## Data Scientist, Intel

Big Data, in all fields, can lead to analytical, data-driven and better decision-making. In education, it can greatly enhance the teacher's toolbox with a deeper understanding of each student's needs, ambitions, weak points, and strong points.
Its strongest ability is to distinguish the individual from the class and suit an optimal curriculum for each student.
However, without proper emotional intelligence and better understanding of the cognitive models we're using as human beings, it could never fully support a true learning process.

# Arnon Hershkovitz:

## Senior Lecturer, Science and Technology Dept., School of Education, Tel Aviv University

We are, it seems, at the beginning of a new era, one in which we can make use of the vast range of data collected in the framework of learning and teaching processes, in order to make those same processes more effective, that is, to improve the performance of students, education systems and decision makers. This is a great opportunity, whose first harbingers can already be seen in research and in application throughout the world. Thus, for example, intelligent tutoring systems (ITS) can already identify students' patterns of behavior – cognitive, metacognitive and emotional – and in some instances also respond accordingly; teachers and lecturers can today make use of existing tools to enhance their knowledge of those of their students using computer-based learning systems; educational institutions can analyze the enormous amount of administrative data that they in any event accumulate, in order to identify success (or, alternatively, dropping out) by students. And, primarily, research that makes use of methods borrowed from the world of Big Data is constantly adding to our knowledge of learning and teaching processes.

But – and it's an important but – there are some important challenges that, at the moment, constitute obstacles to realizing this wonderful vision of Big Data in education. I will briefly address the main ones (in arbitrary order). This is certainly not a full list. It is important, nonetheless, to remember that this is a relatively new field, and so the obstacles to its development are still many.

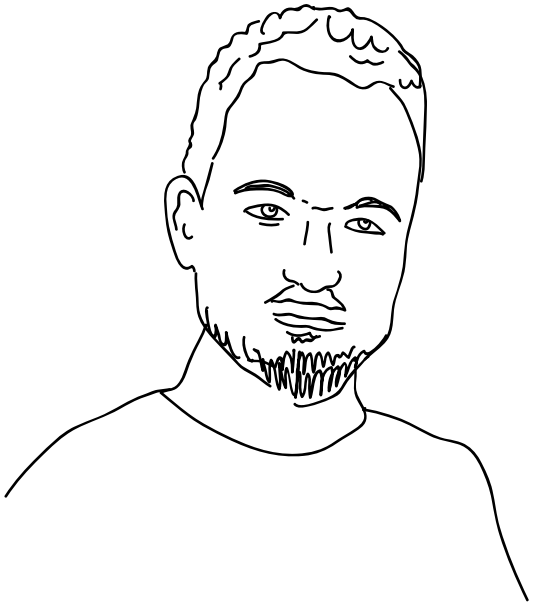
Standardization (a technical obstacle): Currently, there is no standard for collecting and storing data in educational contexts. In particular, there is no standard for collecting log files from educational software. This makes difficulties for researchers who wish to expand their perspective (the methods and approaches developed to handle data in a particular format may not easily

be adaptable to other data). It also raises difficulties for the developers of generic applications, who wish to get to the largest number of end users (students, teachers, school principals, policy makers).

Disciplinary isolation (a research obstacle): Many of those involved the various fields of educational research are not aware of the way in which methods borrowed from the world of Big Data can assist them in their research. But even if they are aware of it – they may not always be able to apply those methods, because of objective difficulties in understanding them. To successfully apply those methods, what is needed is multidisciplinary cooperation and cross-pollination. Although the international research communities dealing with Big Data in education (the two main ones being the International Educational Data Mining Society and the Society for Learning Analytics Research) are essentially multidisciplinary, when we look at the individuals who comprise them, cross-disciplinary cooperative ventures are still not the norm.

Awareness on the ground (an implementation obstacle): Are students aware that, when they use learning software, data is collected about them which could improve their use of the system and their learning? Do teachers working with learning software know that data is collected that documents the students' actions, and that this data may assist them and/or their students? Do school principals know that the enormous amount of administrative data collected in their schools' computerized systems (such as Mashov) can be analyzed, in order to obtain important, yet unseen, knowledge regarding various aspects of school administration? Increased awareness of these possibilities may increase demand for applications, and thus accelerate research and development in this area.

Ethical and legal issues (clichéd, but important to mention): Research, development and implementation in this field have to be carried out while protecting the privacy of the users of these various systems. This is in itself an enormous issue, but for the moment it's enough to just mention it.



# Ram Hadar

## Entrepreneur, Founder of Quototi

If until recently all companies had to struggle to obtain any information about their existing or potential customers, nowadays all they need do is open the virtual floodgates (from API to Screen scrape and everything in between) and the information simply "flows" in.

What is surprising is that while we hand over all our information to various companies willingly and mostly without our knowledge but with our "agreement" (I Agree) – starting with simple games which also install services which record our every activity, through specialized applications, operating systems that generate reports, to sending samples of our DNA to a Canadian company for analysis, for which we pay – we are still reluctant to provide less sensitive information to known destinations in contrast to the biometric databases.

The problem has long been transformed from how to get the information to where all the information is going, who is behind the innocent game your child has installed and what is subsequently done with the information.

From the point of view of the companies themselves, because so much diversity is available in terms of quantity and speed, the question arises how to correctly plan the various databases to deal with all of this diversity quickly (in real time) and manage to distill from it insights of real value, given the noises that exist in the information, noises that are understood and created (all the players are present in the all the arenas and affect them through a wide variety of means). And of course, after all this the question arises as to how the information is kept and who keeps it.

In our world the phenomenon of listening to the whispers of the masses is known and recognized, and therefore we can immediately assume that the users' behavior has changed with this knowledge and is thus not reflecting

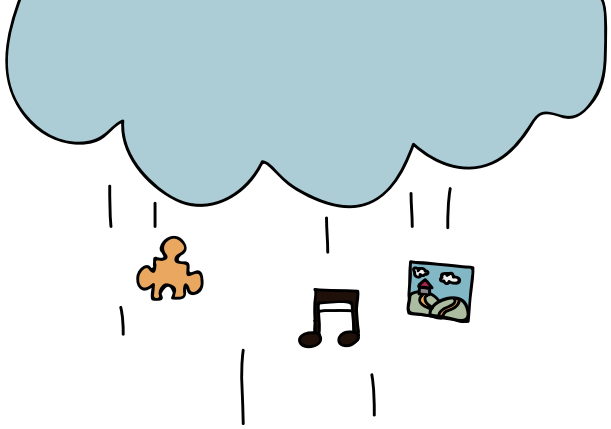
their true needs. On the other hand,"if you torture the information sufficiently, it will confess to anything." In other words, you need to ask the questions and plan the desired information before beginning the work. Otherwise, the result, like the information, will be flawed.

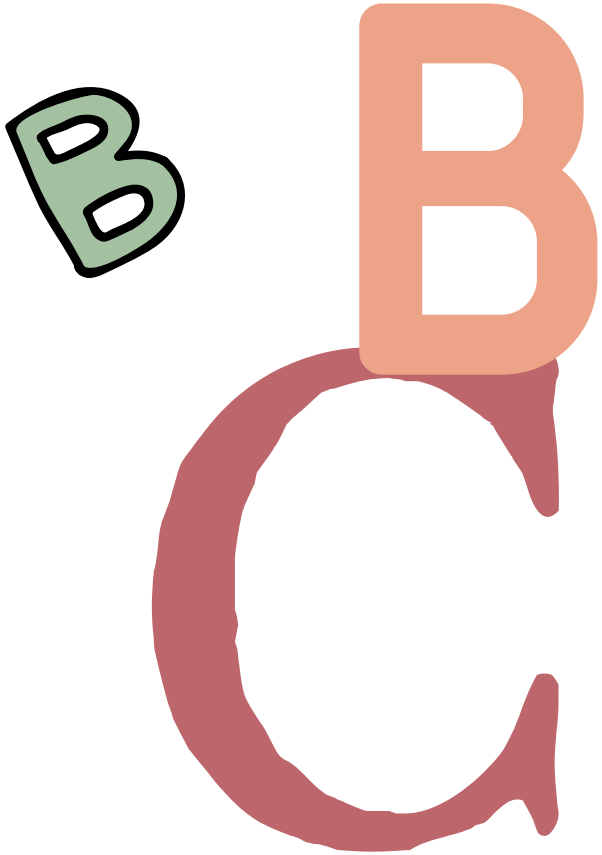
We at Acovado, which specializes in the world of "Big Data," strategy and development and analysis, help companies to forge their informational strategy for the near future while planning far ahead. Working closely together, an ad hoc and hand-on team is developing the set of tools and databases that will support this strategy, as we believe that one size no longer fits all.

When we established the Quototi platform, which helps users to share and curate text using the Mark & Share system, we were conscious of the same dangers. We achieved a solution by, on the one hand, leveraging the reader's natural activity when marking and copying texts in a lengthy process, and through this indicating to us what he regards as the essence of the article, and, on the other hand, giving the user absolute ownership of his raw data.

# Unfinished Dictionary

MindCet
Changing Education Mindset





# Unfinished Dictionary

**Business Intelligence**...a set of methodologies to transform raw data into useful inforamtion for business purposes

**Cloud**.....converged computer infrastructure providing the ability to run and host information from a remote location

**consumer genomics**.....“the consumer’s genome” - attributes that determine the consumer’s behaviour

**content curator**.....assist the process of searching, collating and sharing existing content

**cookies**.....invisible markes placed on your browser as trackers of information

**CRM**.....a model for marketing that’s based on tracking the customer’s actions for better personalization

**Culturonomics**.....a field of investigation which links cultural trends to a quantitative analysis of word use over a particular period of time

**Data Broker**.....a person or business that researches information

**data dives**.....weekend events of bringing together different professionals with data scientists

**Data exhaust**.....information by products, i.e of online activiy

**data hoarders**.....a person who keeps more data than he needs

**data monetization**.....every process in which data is converted to revenue. i.e using data for marketing purposes

**Database**.....an organized collection of data

**Datafication**.....growing dependence upon data to operate properly

**datashop**.....a large open repository of educational technology interaction data

**Educational Data Mining**..research field concerned with data from education

**Gigabytes**.....unit of computer memory or data storage capacity equal to 1,024 megabytes or 2^30 bytes (or a list of 2^33 ones and zeros)

**Hadoop**.....open-source framework for processing large sets of data.

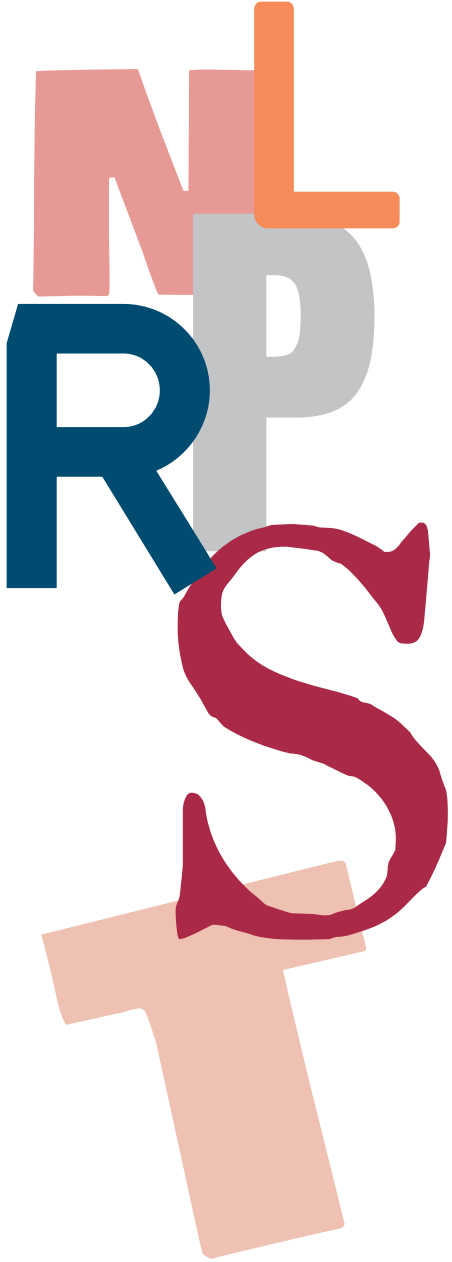
**HPC**.....any computational activity requiring more than a single computer to execute a task

**information silo**.....any information management system that is unable to communicate with other information management systems.

**intelligent lampposts**...sensor networks for energy efficient transportation systems

**JSON**.....java script object notation

**Learning Analytics**.....research field concerned with data from education



**liquid data**.....open, widely available, and in shareable

**NoSQL**.....a format for a database which providesa mechanism for storage and retrieval of data

**Petabytes**.....a unit of computer memory or data storage capacity equal to 2^15 bytes (or a list of 2^17 ones and zeros)

**RDBMS**.....a format for a database in which infromation is grouped in relation to a pre-defined key.

**siloing**.....a process or a state in which information is not communicated between different parts of a system or organization.

**simstudent**.....simulated student

**Streaming Data**.....the process of transfering unprocessed data at a steady high-speed rate

**TDA**.....topological data analysis

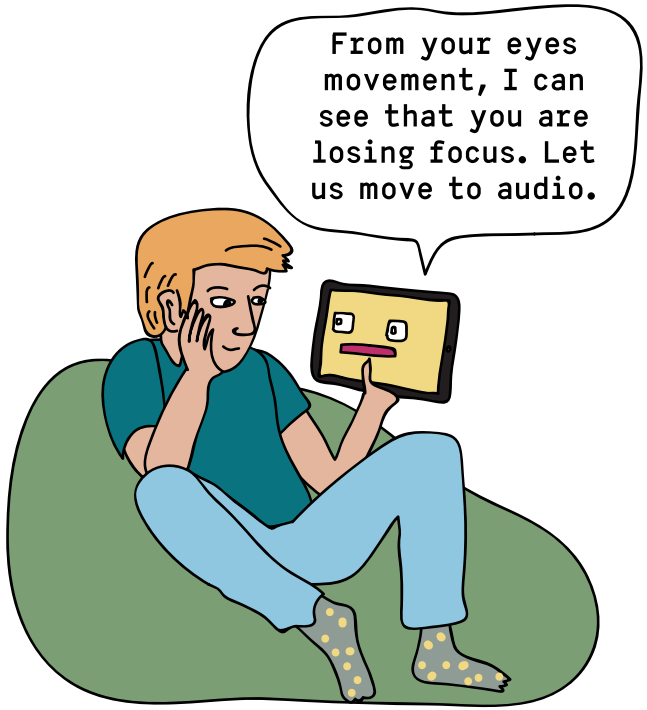
**Terabytes**.....a unit of computer memory or data storage capacity equal to 1,024 gigabytes - 2^40 bytes (or a list of 2^43 ones and zeros).

# MindCET Pitch

# MindCET Pitch

Big Data's true value comes from the growing range of possibilities it provides to us as digital users, strengthening the value of information and making data a commodity that is available to all. We are all digital data users and contributors; we only need to be aware of it to make the best use of it.

The current development of intelligent systems that understand the process of learning, adjusting the response in real time, based on an inestimable number of subjects and through a wide variety of communication and information channels, can lead to an educational revolution. The development of learning systems that optimize instruction to individual needs and offer the possibility to base policies on real-time macro information, as well as the capacity to allow the entire educational community to have access to relevant information, presents us with unique opportunities for the development of a much more relevant educational system.

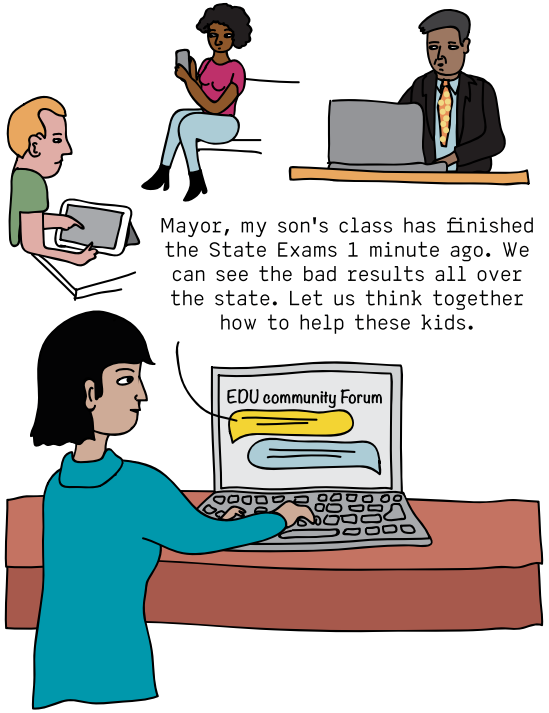




Nevertheless, expectations of Big Data as revealing the truth about human behavior may lead to misconceptions and misleading actions. Preventive and predictive measures have to be wisely pondered especially when they deal with students' future. Privacy concerns should be addressed when determining the limits of what educational data comprises. Moreover, intelligent systems need to allow for students' creativity and foster free cognitive activity that leads to new problem-solving possibilities. Personalization should enhance the capacity of the system to respond to the user's real needs and not inhibit the user's overall view of the general picture.

We believe that the rise in awareness about the significance of Big Data is the key to making it a truth enabler. Understanding the uses and possibilities of Big Data can allow for an active role by all members of the educational community towards re-constructing the current deficient educational system:

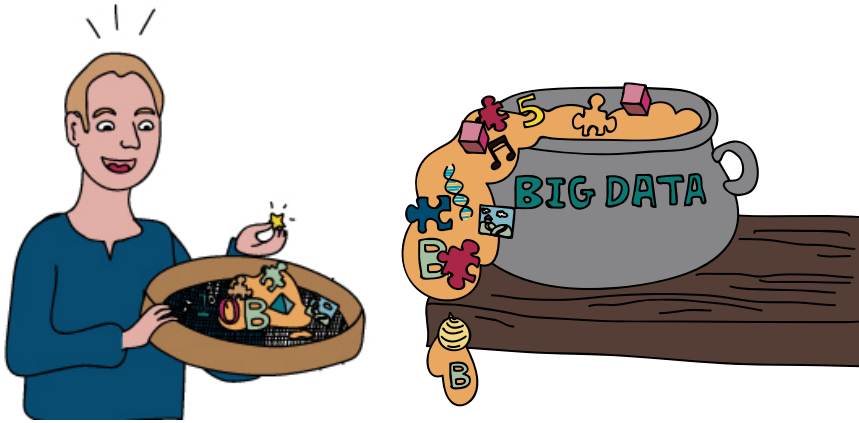
Students and parents can participate in the decisions about the boundaries of educational data, facilitating the development of meaningful learning systems while at the same time protecting their individual privacy. Lack of awareness can lead to misjudgments that can only obstruct the implementation of intelligent systems that can significantly benefit the students.





Teachers and educational policy makers' awareness that data is a means to help understand the student but not its perfect reflection, may give a more balanced and efficient weight to Big Data's predictive powers.

EdTech entrepreneurs' contributions can pave the way to innovative forms of exploring students' data, if they keep an open and curious mind about the vast unknown territory all this tsunami of complex data can uncover



# Credits:

Editor in chief.....Dr. Cecilia Waismann
Writer.....Dr. Cecilia Waismann
.....Ran Magen
Research.....Dr. Cecilia Waismann
.....Estela Melamed
.....Ran Magen
.....Idan Yitzhaky
.....Sarid: Institute for Research Services
.....Nurit Vatnik
Collaborators.....Avi Warshavsky
.....MindCET Staff
.....CET Marketing Staff
Graphic design.....Shine Little Studio
Illustration.....Noemi Fein
Introduction Video.....Nir Weiss
English editing.....Nechama Unterman

We would like to thank all the experts who
voluntarily collaborated with their knowledge.





Address: 904, Tsvi Borenstein st. Yeruham
16 Klausner St., Tel Aviv, Israel
Website: www.MindCET.org
E-Mail: MindCET@cet.ac.il