



Quantified Educational Life: Finding Needles in the Campus Haystack



A hypothetical big data privacy case study

In 2015, MIT decides to embrace the potential of data-powered, analytics-driven systems in all aspects of campus life, from education to health care to community sustainability. The panel discussion will explore how this decision unfolds through five phases.

1. Recognizing the value of online education but dissatisfied with the disconnect between the MOOC courses offered to anyone on the planet and MIT's campus course offerings, four departments -- Civil and Environmental Engineering (Course 1), Electrical Engineering and Computer Science (Course 6), Physics (Course 8), and Mathematics (Course 18) -- have decided to require that all students use the edX platform for doing problem sets, viewing preparatory video lectures, seeking help from teaching assistants (through online discussion forums) and as in collaboration on class projects. Academic departments believe that they can improve the quality of instruction by studying details of student activity down to the click-stream level. Some instructors try to understand how to improve lectures and problem sets by comparing student usage patterns with final grades. There is concern that some who have access to this data could be trying to assess when students are working and sleeping; who they study with; and where they sleep at night. However, there is not yet a policy on who can access the log files associated with MOOC data nor are there tools to analyze those logs. Hence, some students wonder about what they might be exposing and to whom.

Going further, by analyzing student food purchases and campus health records, some researchers on campus seek correlations between health status and educational achievement. As MIT believes that student learning data should be held in the public trust and available for general research purposes, with appropriate privacy protection, the Institute seeks to make available a certain amount of this data on an open access basis. Community response is generally quite positive. Courses subjected to intensive study see notable improvements in student achievement after just one year. A small controversy erupts on campus when the FBI serves MIT with a subpoena for all student-produced source code produced for problem sets and exams in Course 6 classes for the last five years. The FBI and the US Secret Service are investigating an insider attack on the BigCo infrastructure. They have found some of the malicious code used in the attack but have no leads on the identify of the coders. On the theory that lots of MIT students go on to work at BigCo, the investigators seek to match code signatures from the malicious code against the corpus of recent MIT students to see if there are sufficiently similar styles to warrant further investigation. MIT complies with the subpoena. In the end, the police apprehended four suspects (two of whom were MIT students) by first identifying one non-MIT student through other means, and then connecting this individual into a social graph of course 6 students together with publicly available social network data.

2. One year into this quantified educational life program, private industry begins to take note and wants to get in on the act. MIT receives a generous gift from Fitbyte, Inc., providing a free wireless activity tracker for every member of the campus community, students, faculty and staff. This version of the Fitbyte device contains not only activity monitoring but also location detection accurate to 10 meters. Anyone who uses these devices and agrees to have their data uploaded to the MIT cloud for research purposes will receive free repair and loss replacement for the device, and the MIT health plan is considering a 10% discount for those who participate in this research study as well. By 2016, 75% of the campus community is using the devices and 45% share their data for research purposes. Armed with data on student sleep and activity levels, professors frustrated with students nodding off in class begin sending reminders to underperforming students with low participation rates when their sleep patterns take them under the optimal 6.75 hours/night of sleep. And, the MITMedical integrates this data into their Electronic Medical Records system so that clinicians can better advise their patients. Fitbyte promises not to release any personally identifiable information to any third party.

3. The collaboration between Harvard and MIT in edX brings the schools' curricula closer together, resulting in a much larger number of students registering for classes on the other campus. One unfortunate side effect is that students completely overwhelm the Red Line between MGH and Porter Square. Reacting to public pressure, the MBTA puts in place a drastic congestion pricing regime with fares increasing by up to 300% on travel between 9:30am - 4pm. (This time window is calculated to avoid impact on rush-hour commuters.) Not wanting to impede educational progress, MBTA gives students the option to ride at normal fares when they can demonstrate a bona fide need to travel between the Kendall Square and Harvard Square campuses. In order to satisfy this requirement, students must share their individual academic schedules with the MBTA. After one year, this new scheme substantially improves the overcrowding problem but the public is still not satisfied. Transportation planners conclude that they need to implement dynamically-adjusting train schedules tuned to student class schedules. Therefore, Harvard and MIT agree to provide the MBTA real time access to all student course schedules.

With the ascendancy of Boston's Silicon-Carbon Alley, the region is increasingly in the sights of the second wave of al Qaeda terrorists as well as domestic white supremacists with Ted Kaczynski-like Luddite tendencies. Unbeknownst to Boston-area residents, the United States Intelligence Community is receiving a steady stream of threats targeting Boston in general and university campuses in particular. Based on a few intercepted communications, some analysts develop a theory that science and engineering students at risk of failing out of their academic programs are vulnerable to being recruited into terrorist plots. Hence, the FBI seeks access to all transcripts for students with lower than a C average. This information is combined with foreign signals intelligence traffic from the NSA to identify students who may be in contact with foreign terrorist organizations. The NSA also asks for the Fitbyte data and medical records to detect students whose metabolic functions indicate unusually high stress levels.

4. The wildfire success of big data analytics on many campuses in the edX consortium has created a nationwide interest in realizing these benefits more broadly. It has also whet the appetite of the commercial sector. A new commercial venture called InsideHealthyLearnersX.com is launched with the intention to aggregate data from 50-100 of the top universities in the United States. Each university already collects and stores data sets roughly equivalent to what MIT and Harvard have. That data will now be contributed to IHeLX. University participants will have the chance to mine this large universe of data in order to compare themselves to broader national education patterns. IHeLX has committed to providing privacy-preserving access to a data stream for research by bone fide scholars, whether or not they are affiliated with one of the member universities. IHeLX has not yet disclosed any further commercial plans through US News & World Report is both interested in and threatened by rumors of future IHeLX services.

5. Even in light of the great success of this effort, the faculty at MIT realizes that the France is still in possession of indispensable intellectual culture, so they approach the École Normale Supérieure (ENS) in Paris about an educational alliance. ENS is interested in the alliance but requires an explanation about how the online educational component of MIT and its relationship with IHeLX is consistent with the fundamental right of privacy as described in the European Union Charter of Fundamental Rights.

None of what is written here bears any relationship to known MIT or edX plans