Problem Statement

- The current focus in learning environments seems to tend towards learning outcomes more than on the learning process to achieve the learning outcome.
- The goal of learning analytics is to provide real-time assessment of the *learning process* towards providing students with adaptive, real-time support in improving their learning skills.
- Learning analytics relies on ubiquitous sensing or collection of huge quantities of data about students' interactions with learning tools, which tools are adapted to fit in the learning analytics world.
- Bigdata is applied in the realm of learning analytics to support the analysis of the continuous and simultaneous flow of learning activity data from students all over the world.

PROBLEM: major ethics/privacy concerns rise up from the fact that huge quantities of data are collected from students' learning experiences. Any compromising of students' data can lead to dire consequences in terms of harming a student's career, reputation, etc., which cannot be tolerated.

Problem Statement *(continued …)*

PII (Personally Identifiable Information)

- Most learning environments nowadays allow or require students' accounts to include Personally Identifiable Information (PII) such as full names, email addresses, phone number, social network IDs, etc., that can violate a student's anonymity and privacy in the case of a data leak, which is quite widespread these days.

Data Ownership

- Any learning data captured from students belongs to the student, neither to the institution this student is part of nor to the organization building the learning analytics system. Learning environments today do not offer such flexibility in the ownership of the students' data.

Our Solution

What should the SOLE goal of learning analytics be?

- Learning analytics should ONLY be concerned with HELPING students improve their learning skills towards becoming the BEST students while at the same time protecting their privacy and anonymity.
- Therefore, every learning analytics tool should be designed and built with the privacy and anonymity of students' data as its top priority.

Our Proposed Solution:

- We do propose simple breach prevention techniques and measures:
    1. a four-layer secure, fully anonymous authentication layer
    2. encryption of data in transit and at rest
    3. data access control layer

Authentication Layer

- Anonymity starts at the account creation level. Therefore, we implement the following measures:
  - Do NOT enable students to enter PII in their account or profile (e.g. names, address, phone, etc.).
  - Each student account comprises the four following System Identifiable Information (SII) tokens:
    - [public] institution id (unique key per institution – system-generated)
    - [public] student id (unique key per student – system-generated)
    - [student-private] PIN number (student's password generated by the system)
    - [system-private] internal system id for that user (known ONLY to the system)
  - The combination of all four tokens is what is needed to uniquely identify each user without unveiling anything personal about students, even in the case where data would be compromised.
- CON: Anonymity comes at the cost of being impersonal towards students. Students may like to see in their dashboards labels like "Hi, John, you've scored 90%", and teachers may want to see their students' names instead of IDs. These are aspects we are definitely considering and want to address without risking privacy and anonymity at all cost.

Data Encryption

- Even though we fully anonymize students' learning data, it is still important to minimize risks of leaks by protecting the data in transit from the client side to the server side and at rest in databases on the server.
- VPN could have been an option, but the huge quantity of data constantly flowing into the system and the disparity of the students' locations and work environments makes it difficult to use a VPN.
- Therefore, we propose to use programmatic data encryption techniques to encrypt learning events on the client side and then send these data encrypted to the server side, where they are decrypted for analysis.
- Here are tools, mechanisms, or techniques we propose to use:
  - JCryption (JavaScript encryption library) - http://www.jcryption.org/
  - JSEncrypt - http://travistidwell.com/jsencrypt/
  - Chilkat Software - http://www.chilkatsoft.com/encryption-features.asp
  - Java Cryptography Architecture (JCA) - https://docs.oracle.com/javase/8/docs/technotes/guides/security/crypto/CryptoSpec.html
  - Database encryption - http://docs.mongodb.org/master/core/security-introduction/
  - SSL (HTTPS)
  - Disk encryption

Data Access Control Layer

- It is extremely important that students have full control over what data they accept to share, when they want to share them, with whom to share data.

- We are building a learning analytics system where the data **really** belong to the student. To accomplish this, we build a central student management console where they can go and select with an extremely fine level of granularity who has access to their data and to some or all data.

- Each student can choose not to have his/her data collected. They are allowed to go in their management console, view all of their data collected so far, decide to withdraw all or some of their data, etc.

- In other words, using this transparent approach, we make students the sole owners of their data.