INTERNATIONAL COUNCIL FOR OPEN AND DISTANCE EDUCATION

October 16-19, 2017
Sheraton Centre, Toronto, Canada

WORLD CONFERENCE ON ONLINE LEARNING
icde2017

Privacy-Preserving Learning Analytics

Vassilios S. Verykios [3]

Professor, School of Sciences and Technology

A joint work with Evangelos Sakkopoulos[1], Elias C. Stavropoulos[2], Vasilios Zorkadis[4]

[1] Ministry for Migration Policy, Immigration Information Systems Department, Greece. e.sakkopoulos@ypes.gr

[2] Educational Content, Methodology & Technology Laboratory, Hellenic Open University, Greece. estavrop@eap.gr

[3] Hellenic Open University, Greece. verykios@eap.gr

[4] Hellenic Data Protection Authority, Greece. zorkadis@dpa.gr

20 YEARS HELLENIC OPEN UNIVERSITY

Outline

- Social genome & big data
- Data Protection Regulatory Framework
- Educational Data & Learning Analytics
- Privacy, Protection, and Anonymity
- Side effects and solutions

Social Genome

- Any activity we engage in, leaves behind an imprint
- posting on social media,
- doing some online shopping,
- applying for a credit card or a travel visa
- This info creates a unique social genome for every one

Big Data and Analytics

- Huge data banks
- Numerus pathways we collect data from
- An unprecedented power to analyze this data for the benefit of society

Pros and cons

- Huge improvement in our daily lives
- Verification of research results and reduction in the costs of research projects
- Strict regulations and laws are required, to protect individual and civil rights

Regulatory Frameworks

- HIPAA, protection of health related data
- FERPA, protection of data concerning educational rights
- EU Data Protection Directive, covers any personal data held by data administrators
- EU GDPR, 25 May 2018

General Data Protection Regulation

- Deals with consent, data governance, audit, and transparency regarding data breaches
- Organizations must have technology that demonstrates the usage of data
- In case of a breach, involved individuals must be informed within 72 hours

Data Ownership Issues

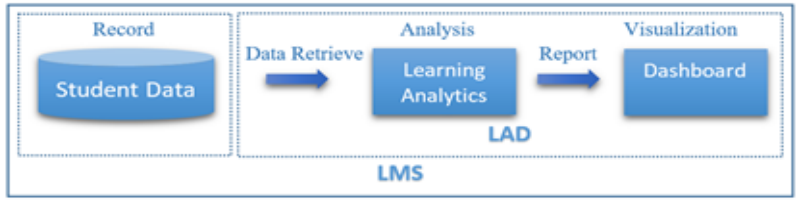
- Companies collect their own data for their benefit
- For *research purposes*, data are gathered by different bodies having their own ways of managing and storing them
- Access and sharing requires tools, technology and practices

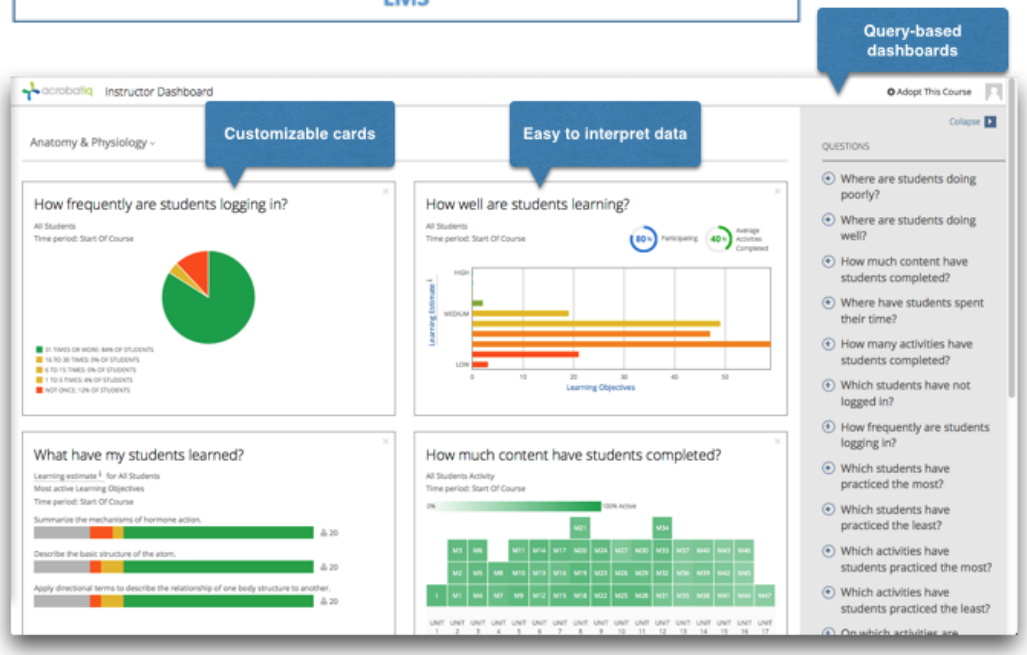
Educational Data

- Institutions traditionally obtain and store information
- Student grades & attendance
- Parental status & student health
- Technology allow us now to monitor students' activities

Learning Analytics

- Information is predominantly collected to improve the educational system
- Assessing the usefulness of educational material in relation to students' learning capabilities
- Personalizing the way the teaching method used to convey knowledge and offer support to students





Synergies of scale

- Institutions need to take initiatives
- The careless use & analysis of data belies many dangers
- Disclosure of simple demographic data may lead to the identification of individuals

Data Protection and Confidentiality

- *Data protection*: data need to be protected so that it cannot be accessed by an intruder in any way, shape or form.
- *Data confidentiality*: data can be accessed by legally authorized individuals who are prohibited to share the information with anyone else

Privacy and Anonymity

- *Privacy*: we know the identity of the person but are not aware of the specific attributes or properties of this person
- *Anonymity*: we know the attributes or properties but are unaware of whom they belong to

Data Processing Regulations

- Clear legal guidelines are required
- Individuals must have the consent and the prerogative to be exempt from the collection
- Be aware of exactly how, where and for how long their data will be stored and utilized

Data Access Regulations

- Third parties who are able to access data banks must be authorized and comply with all the legal requirements
- Information systems storing their data must be accessible only by authorized individuals
- If data is to be made available to the public, it must be de-identified

Personal Information

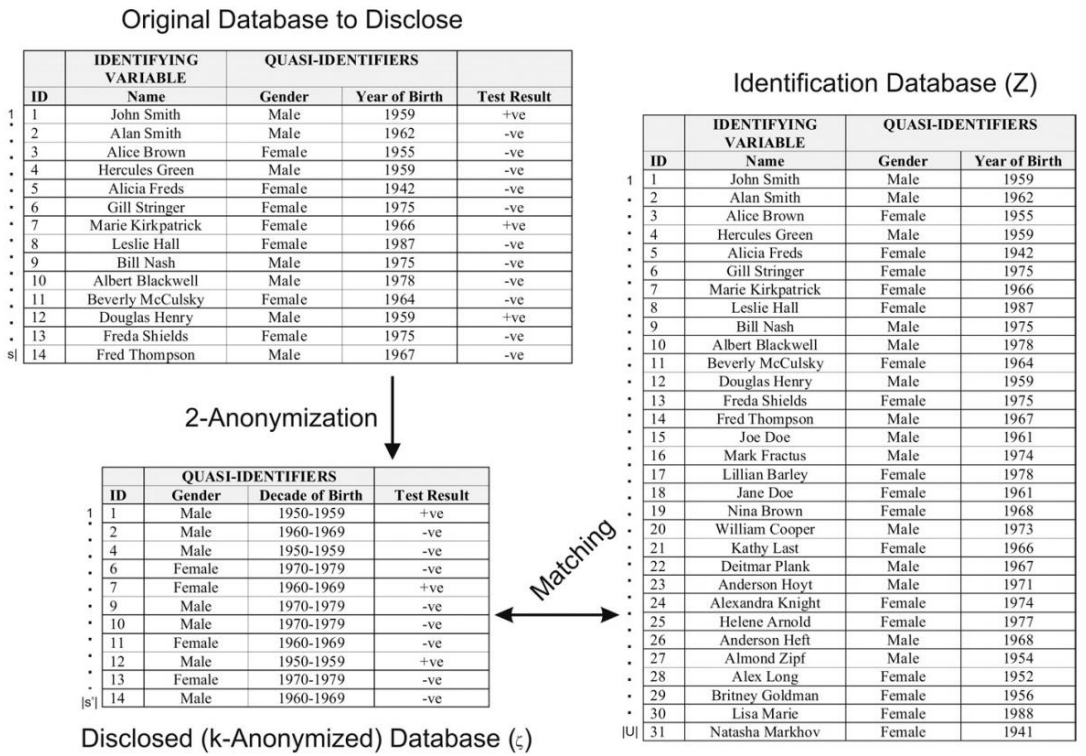
- Unique identifying traits (social security number, passport id, fingerprint)
- Pseudo (quasi) identifiers, a combination of relatively common attributes (sex, age, area code)
- Sensitive data (medical or criminal status)
- Leisure activities or qualifications

Data De-identification

- Omittance of unique identifiers to avoid person's identity being revealed within a collection of data
- Detect quasi identifiers
- Generalize or distort the data values so that the data is rendered anonymous

K- Anonymity

- Information is altered so that it is virtually impossible for the individuals' identity to be disclosed
- Ensure that each record within the set is similar to at least k-1 other records
- The higher the value of k the safer the identity of the data subject is



Side Effects

- Large number of records is lost
- The original data set is distorted & altered
- It is not always feasible to identify the deviation of the results of the tampered data sets from the original
- The greater the degree of anonymization, the poorer the quality of the information itself

Solutions

- Use big data to its full advantage, but ensure that there is compliance with rules and regulations of privacy
- Big open data are transparent and accurate
- Important scientific pathways for protection must be followed

References

- K. el Emam and F.K. Dankar, *Protecting Privacy Using k-Anonymity,* J. American Medical Informatics Association, 15(5), pp. 627-637, 2008
- P. Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, UCLA Law Review, Vol. 57, p. 1701, 2010
- H.-C. Kum, A. Krishnamurthy, A. Machanavajjhala, and S.C. Ahalt, *Social genome: Putting big data to work for population informatics*, Computer 47(1), PP. 56-63, Jan 2014
- J. Newman and S. Oh, *8 Things You Should Know About MOOCs*, June 13, 2014,
- J.P. Daries, J. Reich, J. Waldo, E.M. Young, J. Whittinghill, D.T. Seaton, A.D. Ho, and I. Chuang, *Privacy, Anonymity, and Big Data in the Social Sciences*, ACM Queue, 12(7), 2014
- E. Young, *Educational Privacy in the Online Classroom: FERPA, MOOCS, and the Big Data Conundrum*, Harvard Journal of Law & Technology 28(2), pp. 549-592, Spring 2015
- O. Angiuli, J. Blitzstein, and J. Waldo, *How to De-identify Your Data*, ACM Queue, 13(8), 2015
- M.E. Gursoy, A. Inan, M.E. Nergiz, and Y. Saygin, Yucel, *Privacy-Preserving Learning Analytics: Challenges and Techniques*, IEEE Trans. Learn. Technol., 10(1), pp. 68-81, 2017
- *GDPR: How to win the data privacy war*, April 10, 2017