JOURNAL OF LEARNING ANALYTICS
SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH
(2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. http://dx.doi.org/10.18608/jla.2016.31.8

De-Identification in Learning Analytics

Mohammad Khalil and Martin Ebner
Educational Technology
Graz University of Technology, Austria
mohammad.khalil@tugraz.at

ABSTRACT: Learning analytics has reserved its position as an important field in the educational sector. However, the large-scale collection, processing, and analyzing of data has steered the wheel beyond the borders to face an abundance of ethical breaches and constraints. Revealing learners' personal information and attitudes, as well as their activities, are major aspects that lead to identifying individuals personally. Yet, de-identification can keep the process of learning analytics in progress while reducing the risk of inadvertent disclosure of learners' identities. In this paper, the authors discuss de-identification methods in the context of the learning environment and propose a first prototype conceptual approach that describes the combination of anonymization strategies and learning analytics techniques.

Keywords: Learning analytics, anonymization, de-identification, ethics, privacy

1 INTRODUCTION

Learning analytics is an active area of the research field of online education and Technology Enhanced Learning (TEL). It applies analysis techniques to the education data stream in order to achieve several objectives. These objectives mainly aim to intervene and predict learners' performance in pursuance of enhancing the learning context and its environment. Higher Education (HE) and online course institutions are looking at learning analytics with an interest in improving retention and decreasing the total dropout rate (Slade & Galpin, 2012). However, ethical issues emerge while applying learning analytics in educational data sets (Greller & Drachsler, 2012). At the first International Conference on Learning Analytics and Knowledge (LAK '11), held in Banff, Alberta, Canada in 2011, participants agreed that learning analytics raises issues relevant to ethics and privacy and "it could be construed as eavesdropping" (Brown, 2011). The massive data collection and analysis of these educational data sets can lead to questions related to ownership, transparency, and privacy of data. These issues are not unique to the education sector only, but can be found in the human resource management and health sectors (Cooper, 2009). At its key level, learning analytics involves tracking students' steps in learning environments, such as videos of MOOCs (Wachtler, Khalil, Taraghi & Ebner, 2016), in the interest of identifying who are the students "at risk," or to help students with decisions about their futures. Nevertheless, tracking interactions of students could unveil critical issues regarding their privacy and their identities (Boyd, 2008).

Ethical issues for learning analytics fall into different categories. We mainly summarize them as the following (Khalil & Ebner, 2015b): 1) transparency of data collection, usage, and involvement of third parties; 2) anonymization and de-identification of individuals; 3) ownership of data; 4) data accessibility and accuracy of the analyzed results; 5) security of the examined data sets and student records from any

(2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. http://dx.doi.org/10.18608/jla.2016.31.8

threat. These criteria point to the widely based security model CIA, which stands for Confidentiality, Integrity from alteration, and Availability for authorized parties.

The learning analytics community needs to deal carefully with the potential privacy issues while analyzing student data. Educational data analysis techniques can reveal personal information, attitudes, and activities related to learners (Bienkowski, Feng, & Means, 2012). However, there has been limited research, and there are still numerous unanswered questions related to privacy, personal information, and other ethical issues in the context of learning analytics (Bienkowski, Feng, & Means, 2012; Greller & Drachsler, 2012; Slade & Galpin, 2012; Slade & Prinsloo, 2013). For example, some educators claim that educational institutions are using applications that collect sensitive data about students without sufficiently respecting data privacy and how the data will eventually be used (Singer, 2014). Thus, data degradation (Anciaux et al., 2008), de-identification methods, or deletion of specific data records, may be required as a solution to preserve learners' information. In this paper, we will mainly focus our discussion on the de-identification process in the learning analytics atmosphere and afford a first prototype conceptual approach that combines learning environment, de-identification techniques, and learning analytics.

The paper is organized as follows: Section 2 covers the de-identification in general and the current laws associated with education, as well as the drivers linked with learning analytics. In Section 3, we propose the de-identification–learning analytics approach. The last section discusses the limitations of the de-identification process in learning analytics.

2 BACKGROUND

2.1 Personal Information and De-Identification

Personal information is any information that can identify an individual. In fields such as the health sector, it is named Personal Health Information or PHI. While in other fields, such as the education sector, this information is named Personal Identifiable Information or PII. The National Institute of Standards and Technology (NIST) defines PII as "any information about an individual maintained by an agency, including 1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and 2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information" (McCallister, Grance, & Scarfone, 2010). The personal information of learners can be categorized into details such as name, sex, photograph, date of birth, age, address, religion, marital status, e-mail address, insurance number, ethnicity, et cetera, or educational details such as qualifications, courses attended, degrees, and study records. As a criterion, a leak of individuals' personal information can induce misuse of data, embarrassment, and loss of reputation. However, organizations may be required to publish details extracted from personal information. For instance, some educational institutions are required to provide statistics about student progress; likewise, health organizations may need to report special cases from their patient records, such as communicable diseases. As a result, de-identification helps organizations to protect privacy

(2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. http://dx.doi.org/10.18608/jla.2016.31.8

while still informing the public. The de-identification process is used to prevent revealing individual identity and keeping the PII confidential.

In learning analytics, it is common for stakeholders to request additional information about the results extracted from educational data sets. Educational data mining and learning analytics mainly aim to enhance the learning environment and empower learners and instructors (Greller & Drachsler, 2012). Therefore, the analysis of these data may have interesting trends that could lead to further and deeper analysis by other institutions or researchers. Requests for more extensive analysis may involve the use of student-level data. Accordingly, ethical issues arise, such as privacy disclosure, and the need to de-identify the data becomes paramount.

Recently, Harvard and MIT universities released de-identified data from 16 courses offered in 2012–2013 from their well-known edX Massive Open Online Course (MOOC) (MIT News, 2014). The Harvard and MIT edX ensures that the anonymity of the released data complies with the Family Educational Rights and Privacy Act (FERPA).[1] Furthermore, Prinsloo and Slade (2015) suggested different approaches that inform students in higher education of the implications of learning analytics on their private data.

2.2 De-Identification Legislation

De-identification of student records has been regulated in the United States and the European Union. The United States adopted FERPA regarding the privacy of student educational records. In the European Union, the Data Protection Directive (DPD; 95/46/EC[2]) regulates the processing of personal data and the movement of such information. FERPA §99.31(b) deals with the de-identification of data rule. It clearly states that institutions "may release, without consent, education records, or information from education records, that has been de-identified through the removal of all Personally Identifiable Information (PII)." This section of FERPA requires institutions to use reasonable methods to identify the other parties who disclose education records. On the other hand, the most explicit citation of de-identification in the European DPD is Article 26 on anonymization, in which "principles of data protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable." Moreover, parties are encouraged to use de-identification techniques to render identification of data subjects impossible. It is not obvious, however, what level of de-identification is required to anonymize education records under European law. However, the Article 29 Data Protection Working Party has an opinion on the identification of data: "Once a data set is truly anonymized and individuals are no longer identifiable, European data protection law no longer applies" (2014, p. 5).

2.3 Drivers of De-Identification in Learning Analytics

A study by Peterson (2012), addressed the need to de-identify data used in academic analysis before making it available to institutions, to businesses, or for operational functions. Peterson (2012) pointed

[1] http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html (last access January 2015)
[2] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri= CELEX:31995L0046:EN:HTML(last access January 2015)

(2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. http://dx.doi.org/10.18608/jla.2016.31.8

to the idea of keeping a unique identifier in case a researcher may need to study the behaviour of a particular individual. Slade and Prinsloo (2013), however, drew attention to the ambiguity of data mining techniques in monitoring student behaviour in educational settings. The authors linked de-identification with consent and privacy and stressed the need to guarantee student anonymity in their education records in order to achieve learning analytics objectives such as interventions based on student characteristics. An example of the link between consent and de-identification would be a questionnaire or survey that those filling it out are told will be used for research only. In that case, clearly the limitation of using their data will be just the one study. If the survey includes personal information, however, then assurances of anonymizing their data should be considered.

Ryan Baker (2013) discussed the demands of de-identifying educational data sets in his "Learning, Schooling, and Data Analytics" chapter in the *Handbook on Innovations in Learning for States, Districts, and Schools*. De-identification of these data sets means being able to share them among other researchers without violating FERPA regulations. Baker stressed that educational policies should include rules for anonymizing data in order to prevent identifiable information from being leaked without authorization. Furthermore, Drachsler and Greller covered the topic of anonymization in their DELICATE approach (Drachsler & Greller, 2016). A "strictly guarded key" should be held so that researchers may link their results from learning analytics and educational data mining with individual students in order to benefit the students. De-identification techniques have been reviewed as a right of access principle in learning analytics deployment (Pardo & Siemens, 2014). In addition, Pardo and Siemens further suggest that semantic analysis might be required to detect identifiable records in anonymized data sets.

3 PROPOSED APPROACH

In this section, we propose a conceptual de-identification–learning analytics framework as shown in Figure 1. The framework begins with learners involved in learning environments. Currently, a large number of learning environments support online learning, such as MOOCS, Learning Management Systems (LMS), Immersive Learning Simulations (ILS), mobile learning, and Personalized Learning Environments (PLE). These platforms offer environments with rich, vast amounts of data that can be quantitatively/qualitatively analyzed to benefit learners and enhance the learning context.

(2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. http://dx.doi.org/10.18608/jla.2016.31.8

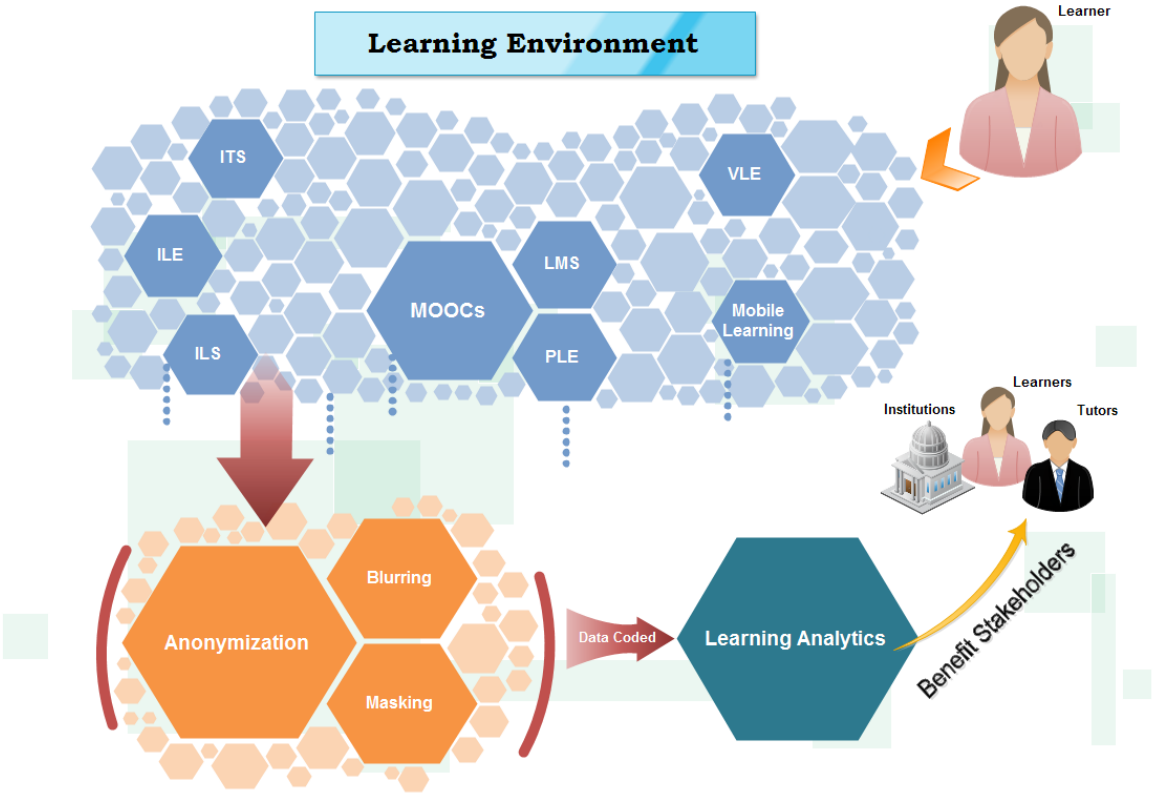

Figure 1: The proposed conceptual de-identification–learning analytics framework

The next step is the de-identification process where techniques to convert personal and private information into anonymized data take place. De-identification techniques include such methods as anonymization, masking, blurring, and perturbation. The last step includes the de-identified data linked with a unique descriptor that may be examined by learning analytics researchers and benefit stakeholders, but ultimately must be used only to the advantage of students.

3.1 De-Identification Techniques

In our proposed de-identification–learning analytics conceptual framework, there are several techniques available to de-identify student data records. Figure 3 lists several methods of de-identification and provides examples (based on Article 29 Data Protection Working Party, 2014; Cormode & Srivastava, 2009; Eurostat, 1996; Petersen, 2012).

Anonymization

Data anonymization techniques have recently been keenly researched in different structured data records with the goal of guaranteeing the privacy of sensitive information against unintended disclosure and a variety of attacks (Cormode & Srivastava, 2009). Ohm (2010) defined reasons behind anonymization when organizations want to release the data to the public, sell the information to third parties, or share the information within the same organization. The difference between anonymization

(2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. http://dx.doi.org/10.18608/jla.2016.31.8

and de-identification, however, is quite misunderstood. Anonymization principles are a subset of holistic de-identification methodologies. Data anonymization is the process of de-identifying data while preserving its original format (Raghunathan, 2013). In the educational context, anonymization refers to different procedures to de-identify student data in such a way that it cannot be re-identified (the opposite of de-identification) unless there is a record code. Anonymization is not reserved only for tabular data records, but can also be applied to other types of data — such as visualized data or graphs — where institutions intend to present their outcomes without revealing sensitive information.

On the other hand, in addition to anonymization, de-identification includes masking, randomization, blurring, and so on. For instance, replacing "Bernard" with "$$$$$$$" is a method of masking while altering "Bernard" to "Wolfgang" would be an example of anonymization. However, masking and blurring are not as well known as anonymization. By any means, de-identification, pseudonymization, and anonymization are interchangeable topics under the information concealing umbrella. To clarify the differences in simple terms, pseudonymization means cloaking the original data with false information with the ability to track it back to its original formation; anonymization, conversely, cannot be reversed (Raghunathan, 2013).

As previously mentioned, educational data records may include private information, such as name or student ID, which singularly are called direct identifiers. Removing or hiding these identifiers does not assure a true data anonymization. Identifiers could be linked with other information that would allow identification of individuals (see Figure 2). However, quasi-identifiers can be used to ensure better de-identification of data. "Date of Birth + Sex + Name" is an example of a quasi-identifier. In 2006, AOL released the search records of 500,000 of its users. Several days after AOL's database release, *New York Times* journalists were able to reveal the identity of a 62-year-old widow using a similar process to that shown in Figure 2 (Soghoian, 2007). AOL admitted that the data release was a mistake and the research team responsible for sharing the data was fired.

Course_ID	Course_Name	User_ID - deidentified	Grade
001	GOL	0005	70%
002	MEK	0009	90%
001	GOL	0006	60%

Course_ID	Grade	Name	Country
002	90%	Sabrina	Austria
001	60%	Michael	Germany
004	75%	Rebecca	Austria

Figure 2: Linking data sources leads to name identification

Another example of identifying individuals was reported in 2000 when demographic information led to retrieving the names and contact information of patients whose medical data had been released in the United States (Sweeney, 2000).

(2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. http://dx.doi.org/10.18608/jla.2016.31.8

Samarati and Sweeney (1998) provided a well-known anonymization technique, namely *k*-anonymization. This method addresses the problem of linking records to identify the individual's information when releasing data, thus safeguarding anonymity. The *k*-anonymity technique focuses on avoiding a data record from being identified with *k* individuals (Cormode & Srivastava, 2009).

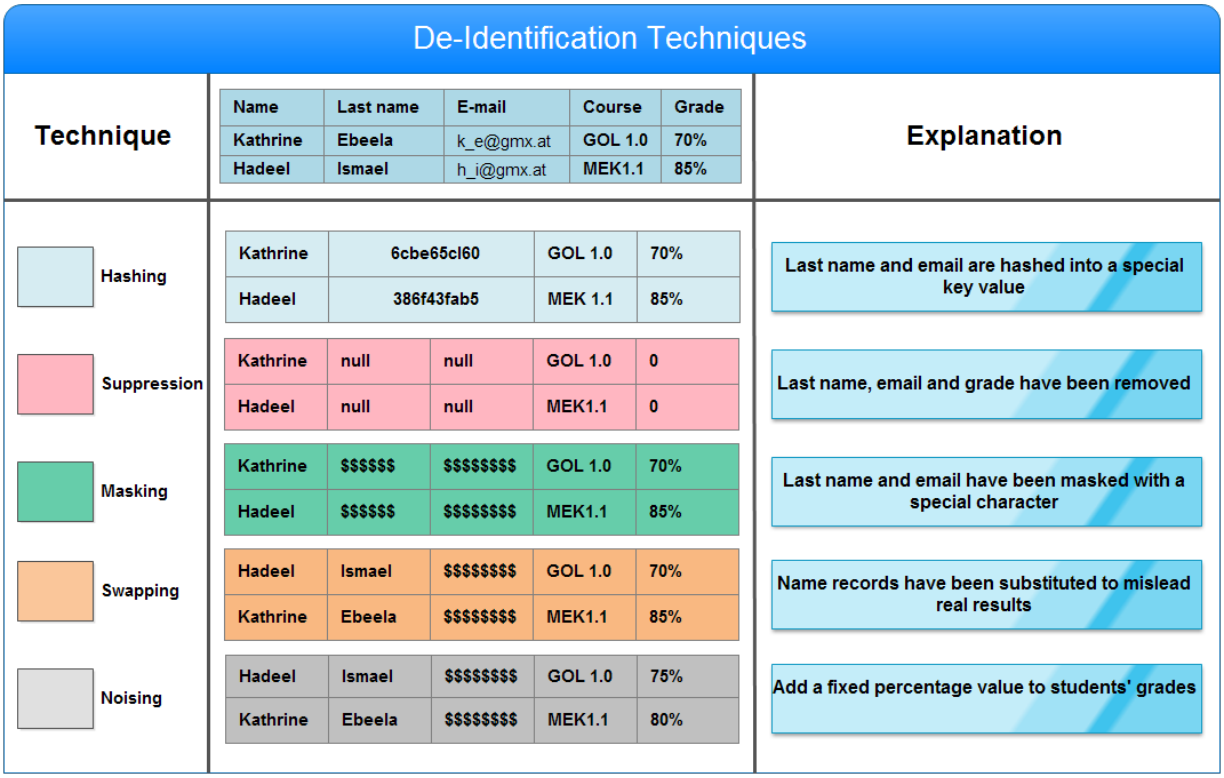

Figure 3: Examples of de-identification techniques

Masking

Masking is a de-identification technique that replaces sensitive data with fictional data in order to disclose results outside the institution. Data masking can modify the data records so that they remain usable while keeping personal information confidential. For instance, character masking replaces a string with special characters.

Blurring

Blurring involves reducing precision to minimize the identification of data. There are several ways to achieve blurring, such as dividing the data into subcategories, randomizing the data fields, or adding noise to data records.

3.2 Coding Data Records

In scientific research, data usually requires further investigation with researchers looking deeper into the details. Having de-identified data might be insufficient for these purposes; researchers may require

(2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. http://dx.doi.org/10.18608/jla.2016.31.8

additional information in order to do more analysis. The American federal Health Insurance Portability and Accountability Act (HIPAA), which is responsible for protecting the confidentiality of patient records, authorizes using an "assigned code" that can be appended to the records in order to permit the information to be re-identified for research purposes.[3] Based on that HIPAA rule, we found that FERPA 99.31(b) allows for using a unique descriptor for student data records in order to match an individual's information for research and institutional use. Accordingly, we conclude that assigning a code to student records in our proposed framework can grant learning analytics researchers the ability to study behaviours of specific students and, therefore, can benefit learners. Despite the fact that learning analytics poses ethical challenges, the main goal is still to benefit learning environments and students, such as making recommendations, classifying students into profiles or predicting their performance (Ebner & Schön, 2013; Greller & Drachsler, 2012; Slade & Prinsloo, 2013; Khalil & Ebner, 2015a; Khalil, Kastl & Ebner, 2016).

4 LIMITATIONS

Despite the fact that de-identification protects confidential information and privacy, the de-identified data still poses some privacy risks (Petersen, 2012). In many cases, some attributes are capable of identifying individuals; in other cases, attackers can link records together from different sources and therefore "code break" the de-identification. On the other hand, in their paper "Privacy, Anonymity, and Big Data in the Social Sciences," Daries et al. (2014) assured that with de-identification, there is no guarantee of keeping the analysis process uncorrupted. Pardo and Siemens agree that "data can be either useful or perfectly anonymous, but never both" (2014, p. 447). The bottom line is that the stricter the de-identification guidelines, the greater the negative affect on the ultimate analysis.

5 CONCLUSION

Since learning analytics first became known in 2011, it has helped learners to improve their performance based on analyzing their educational data. Nevertheless, this field raises many issues related to ethics and ownership. The massive scale of data collection and analysis leads to questions about the consent and privacy of personal information. This paper mainly discusses one of the attainable solutions for preserving learners' sensitive information, the "de-identification of data" to facilitate learning analytics applications. We shed light on this topic via US and EU regulations regarding data privacy. We proposed a conceptual approach with examples of de-identification techniques that assist us with our "iMooX" platform (http://www.imoox.at) and can help learning analytics specialists preserve confidential learner information.

Although de-identification is not a foolproof solution for protecting learner privacy, it is an imperative consideration in examining the ethical dimensions of learning analytics.

[3] Rule 45 C.F.R. § 164.514(c).

(2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. http://dx.doi.org/10.18608/jla.2016.31.8

REFERENCES

Article 29 Data Protection Working Party. (2014). Opinion 05/2014 on Anonymisation Techniques (0829/14/EN WP216). Retrieved from http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Anciaux, N., Bouganim, L., Van Heerde, H., Pucheral, P., & Apers, P. M. (2008). Data degradation: Making private data less sensitive over time. In J. G. Shanahan, S. Amer-Yahia, I. Manolescu, Y. Zhang, D. A. Evans, A. Kolcz, K.-S. Choi, A. Chowdhury (Eds.), *Proceedings of 17th ACM International Conference on Information and Knowledge Management (CIKM 2008)* (pp. 1401–1402). New York: ACM. http://dx.doi.org/10.1145/1458082.1458301

Baker, R. S. J. d. (2013). Learning, schooling, and data analytics. In M. Murphy, S. Redding, & J. Twyman (Eds.), *Handbook on innovations in learning for states, districts, and schools* (pp. 179–190). Philadelphia, PA: Center on Innovations in Learning, Temple University.

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Retrieved from the website of the Office of Educational Technology, US Department of Education, https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf

Brown, M. (2011). *Learning analytics: The coming third wave* (*EDUCAUSE Learning Initiative Brief)*. Retrieved from EDUCAUSE library https://net.educause.edu/ir/library/pdf/ELIB1101.pdf

Cooper, N. (2009). Workforce demographic analytics yield health-care savings. *Employment Relations Today 36*(3), 13–18. http://dx.doi.org/10.1002/ert.20256

Cormode, G., & Srivastava, D. (2009). Anonymized data: Generation, models, usage. In C. Binnig & B. Dageville (Eds.), *Proceedings of the 35th International Conference on Management of Data* (pp. 1015–1018). New York: ACM. http://dx.doi.org/10.1109/ICDE.2010.5447721

Boyd, D. (2008). Facebook's privacy trainwreck: Exposure, invasion, and social convergence. *Convergence: The International Journal of Research into New Media Technologies*, *14*(1), 13–20. http://dx.doi.org/10.1177/1354856507084416

Daries, J. P., Reich, J., Waldo, J., Young, E. M., Whittinghill, J., Ho, A. D., ... & Chuang, I. (2014). Privacy, anonymity, and big data in the social sciences. *Communications of the ACM*, *57*(9), 56–63. http://dx.doi.org/10.1145/2643132

Drachsler, H. & Greller, W. (2016). Privacy and analytics – it's a DELICATE issue. A checklist to establish trusted learning analytics. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge (LAK '16)*, 89–98. http://dx.doi.org/10.1145/2883851.2883893

Ebner, M., & Schön, M. (2013). Why learning analytics in primary education matters. In C. Karagiannidis & S. Graf (Eds.), *Bulletin of the Technical Committee on Learning Technology*, *15*(2), 14–17.

Eurostat. (1996). Manual on disclosure control methods. *Luxembourg: Office for Official Publications of the European Communities*. Retrieved from http://ec.europa.eu/eurostat/ramon/statmanuals/files/manual_on_disclosure_control_methods_1996.pdf

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology and Society*, *15*(3), 42–57.

Khalil, M., & Ebner, M. (2015a). A STEM MOOC for school children: What does learning analytics tell us? *International Conference on Interactive Collaborative Learning* (ICL 2015), (pp. 1217–1221). Florence, Italy: IEEE.

Khalil, M., & Ebner, M. (2015b). Learning analytics: Principles and constraints. In S. Carliner, C. Fulford, & N. Ostashewski (Eds.), *Proceedings of EdMedia: World Conference on Educational Media and Technology, 2015(1),* 1789–1799.

(2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. http://dx.doi.org/10.18608/jla.2016.31.8

Khalil, M., Kastl, C., & Ebner, M. (2016). Portraying MOOCs learners: A clustering experience using learning analytics. In M. Khalil, M. Ebner, M. Kopp, A. Lorenz, & M. Kalz (Eds.), *Proceedings of the European Stakeholder Summit on experiences and best practices in and around MOOCs (EMOOCS 2016*) (pp. 265-278). Norderstedt; Germany: Books in Demand GmbH.

McCallister, E., Grance, T., & Scarfone, K. (2010). *Guide to protecting the confidentiality of personally identifiable information (PII)* (Recommendations of the National Institute of Standards and Technology) Retrieved from the website of Computer Security Division of the National Institute of Standards and Technology http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf

MIT News. (2014, May 30). MIT and Harvard release de-identified learning data from open online courses. [Web post by the MIT News Office]. Retrieved from http://news.mit.edu/2014/mit-and-harvard-release-de-identified-learning-data-open-online-courses

Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, *57*, 1701.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*, 438–450. http://dx.doi.org/10.1111/bjet.12152

Petersen, R. J. (2012, July 18). Policy dimensions of analytics in higher education. *EDUCAUSE Review*. Retrieved from http://er.educause.edu/articles/2012/7/policy-dimensions-of-analytics-in-higher-education

Prinsloo, P., & Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge*, 240–244. http://dx.doi.org/10.1145/2460296.2460344

Raghunathan, B. (2013). *The complete book of data anonymization: From planning to implementation*. Boca Raton, FL: CRC Press.

Samarati, P., & Sweeney, L. (1998). *Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression* (Technical report). Retrieved from the Electronic Privacy Information Center website https://epic.org/privacy/reidentification/Samarati_Sweeney_paper.pdf

Singer, N. (2014, November 18). ClassDojo adopts deletion policy for student data. *New York Times*. Retrieved from http://bits.blogs.nytimes.com/2014/11/18/classdojo-adopts-deletion-policy-for-student-data/

Slade, S., & Galpin, F. (2012). Learning analytics and higher education: Ethical perspectives. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)*, 16–17. http://dx.doi.org/10.1145/2330601.2330610

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57,* 1509–1528. http://dx.doi.org/10.1177/0002764213479366

Soghoian, C. (2007, December 1). AOL, Netflix and the end of open access to research data. *C-Net.* Retrieved from http://www.cnet.com/news/aol-netflix-and-the-end-of-open-access-to-research-data/

Sweeney, L. (2000). Simple demographics often identify people uniquely. *Health (San Francisco)*, *671*, 1–34. San Francisco, CA.

Wachtler, J., Khalil, M., Taraghi, B., & Ebner, M. (2016). On using learning analytics to track the activity of interactive MOOC videos. In M. Giannakos, D.G. Sampson, L. Kidzinski, A. Pardo (Eds.), *Proceedings of the LAK 2016 Workshop on Smart Environments and Analytics in Video-Based Learning* (pp.8–17) Edinburgh, Scotland: CEURS-WS. Retrieved from http://ceur-ws.org/Vol-1579/paper3.pdf