

sets for national security or law enforcement purposes, such as in the [Total Information Awareness](#) Program or in [ADVISE](#), has raised privacy concerns.^{[26][27]}

Data mining requires data preparation which can uncover information or patterns which may compromise confidentiality and privacy obligations. A common way for this to occur is through [data aggregation](#). Data aggregation involves combining data together (possibly from various sources) in a way that facilitates analysis (but that also might make identification of private, individual-level data deducible or otherwise apparent).^[28] This is not data mining *per se*, but a result of the preparation of data before – and for the purposes of – the analysis. The threat to an individual's privacy comes into play when the data, once compiled, cause the data miner, or anyone who has access to the newly compiled data set, to be able to identify specific individuals, especially when the data were originally anonymous.^{[29][30][31]}

It is recommended that an individual is made aware of the following **before** data are collected:^[28]

- the purpose of the data collection and any (known) data mining projects;
- how the data will be used;
- who will be able to mine the data and use the data and their derivatives;
- the status of security surrounding access to the data;
- how collected data can be updated.

Data may also be modified so as to *become* anonymous, so that individuals may not readily be identified.^[28] However, even "de-identified"/"anonymized" data sets can potentially contain enough information to allow identification of individuals, as occurred when journalists were able to find several individuals based on a set of search histories that were inadvertently released by AOL.^[32]

The inadvertent revelation of [personally identifiable information](#) leading to the provider violates Fair Information Practices. This disclosure can cause financial, emotional, or bodily harm to the indicated individual. In one instance of privacy violation, the patrons of Walgreens filed a lawsuit against the company in 2011 for selling prescription information to data mining companies who in turn provided the data to pharmaceutical companies.^[33]

Situation in Europe^[edit]

Europe has rather strong privacy laws, and efforts are underway to further strengthen the rights of the consumers. However, the [U.S.-E.U. Safe Harbor Principles](#) currently effectively expose European users to privacy exploitation by U.S. companies. As a consequence of [Edward Snowden's global surveillance disclosure](#), there has been increased discussion to revoke this agreement, as in particular the data will be fully exposed to the [National Security Agency](#), and attempts to reach an agreement have failed.^[citation needed]

Situation in the United States^[edit]

In the United States, privacy concerns have been addressed by the [US Congress](#) via the passage of regulatory controls such as the [Health Insurance Portability and Accountability Act](#) (HIPAA). The HIPAA requires individuals to give their "informed consent" regarding information they provide and its intended present and future uses. According to an article in *Biotech Business Week*, "[i]n practice, HIPAA may not offer any greater protection than the longstanding regulations in the research arena," says the AAHC. More importantly, the rule's goal of protection through informed consent is approach a level of incomprehensibility to average individuals."^[34] This underscores the necessity for data anonymity in data aggregation and mining practices.

U.S. information privacy legislation such as HIPAA and the [Family Educational Rights and Privacy Act](#) (FERPA) applies only to the specific areas that each such law addresses. Use of data mining by the majority of businesses in the U.S. is not controlled by any legislation.

References^[edit]

1. ^ Jump up to: ^a ^b ^c ^d ^e "Data Mining Curriculum". *ACM SIGKDD*. 2006-04-30. Retrieved 2014-01-27.
2. **Jump up** Han, Kamber, Pei, Jaiwei, Micheline, Jian (June 9, 2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann. ISBN 978-0-12-381479-1.
3. **Jump up** Clifton, Christopher (2010). "Encyclopædia Britannica: Definition of Data Mining". Retrieved 2010-12-09.
4. **Jump up** Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2009). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". Retrieved 2012-08-07.
5. ^ Jump up to: ^a ^b ^c ^d ^e Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases" (PDF). Retrieved 17 December 2008.
6. **Jump up** Han, Jiawei; Kamber, Micheline (2001). Data mining: concepts and techniques. Morgan Kaufmann. p. 5. ISBN 978-1-55860-489-6. Thus, data mining should have been more appropriately named "knowledge mining from data," which is unfortunately somewhat long
7. **Jump up** See e.g. OKAIRP 2005 Fall Conference, Arizona State University About.com: Datamining
8. **Jump up** Witten, Ian H.; Frank, Eibe; Hall, Mark A. (30 January 2011). Data Mining: Practical Machine Learning Tools and Techniques (3 ed.). Elsevier. ISBN 978-0-12-374856-0.
9. **Jump up** Bouckaert, Benno; Frank, Eibe; Hall, Mark A.; Holmes, Geoffrey; Pfahringer, Bernhard; Reutemann, Peter; Witten, Ian H. (2010). "We have experiences with a Java open-source package". *Journal of Machine Learning Research*. **11**: 2533–2541. the original title, "Practical machine learning", was changed ... The term "data mining" was [added] primarily for marketing reasons.
10. **Jump up** Ména, Jesús (2011). Machine Learning Forensics for Law Enforcement, Security, and Intelligence. Boca Raton, FL: CRC Press (Taylor & Francis Group). ISBN 978-1-4398-6069-4.
11. **Jump up** Piatetsky-Shapiro, Gregory; Parker, Gary (2011). "Lesson: Data Mining, and Knowledge Discovery: An Introduction". Introduction to Data Mining. KD Nuggets. Retrieved 30 August 2012.
12. **Jump up** Fayyad, Usama (15 June 1999). "First Editorial by Editor-in-Chief". *SIGKDD Explorations*. **13** (1): 102. doi:10.1145/2207243.2207269. Retrieved 27 December 2010.
13. **Jump up** Kantardzic, Mehmed (2003). Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley & Sons. ISBN 0-471-22852-4. OCLC 50055336.
14. **Jump up** Gregory Piatetsky-Shapiro (2002) *KNuggets Methodology Poll*, Gregory Piatetsky-Shapiro (2004) *KNuggets Methodology Poll*, Gregory Piatetsky-Shapiro (2007) *KNuggets Methodology Poll*, Gregory Piatetsky-Shapiro (2014) *KNuggets Methodology Poll*
15. **Jump up** Óscar Marbán, Gonzalo Mariscal and Javier Segovia (2009); *A Data Mining & Knowledge Discovery Process Model*. In Data Mining and Knowledge Discovery in Real Life Applications, Book edited by: Julio Ponce and Adem Karahoca, ISBN 978-3-902613-53-0, pp. 438–453, February 2009, I-Tech, Vienna, Austria.
16. **Jump up** Lukasz Kurgan and Petr Musilek (2006); *A survey of Knowledge Discovery and Data Mining process models*. The Knowledge Engineering Review. Volume 21 Issue 1, March 2006, pp 1–24, Cambridge University Press, New York, NY, USA doi:10.1017/S0269888906000737