

# Preface

This book grew out of the work of the authors on the problem of preserving the privacy of patient data, which started when they were postdoctoral researchers in the Health Information Privacy Laboratory, Department of Biomedical Informatics, Vanderbilt University. Part of their work was to understand the privacy threats that disseminating clinical data entails and to develop methods to eliminate these threats. The use of data derived from the Electronic Medical Record (EMR) system of the Vanderbilt University Medical Center enabled the authors to realize and appreciate the complexity of the problem and to gain valuable insights that led to developing practical solutions.

The structure of the book closely follows the order in which the authors undertook this research, and some of their works formed the basis of the material presented in this book. We started by realizing that disseminating EMR data requires addressing many important, and often unexplored, privacy issues. One of the main issues was to examine the re-identifiability of diagnosis codes. Towards this goal, we studied a range of attacks that may lead to patient re-identification and performed empirical studies to demonstrate the feasibility of these attacks. Using several large EMR datasets, we showed that the attacks we considered pose a serious privacy threat, which cannot be addressed by popular approaches. The work won a Distinguished Paper Award from the American Medical Informatics Association (AMIA) Annual Symposium in 2009 and appeared in a revised and extended form in the Journal of the American Medical Informatics Association (JAMIA). Having realized the importance of guaranteeing both data privacy and the usefulness of data in biomedical applications, we designed the first approach which guarantees that the disseminated data will be useful in validating Genome-Wide Association Studies. These studies attempt to find clinically meaningful associations between patients' diagnosis and genetic variations, and are considered as the holy grail of personalized medicine. The work was published in the Proceedings of the National Academy of Sciences in 2010 and was reported by the National Human Genome Research Institute (NHGRI) among the important advances in the last 10 years of genomic research (Eric D. Green, et al. in *Nature*, vol. 470, 2011). We also gained useful insights on the problem and future directions, when we were preparing two

tutorials that were presented at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD) in 2011, and at the SIAM International Conference on Data Mining (SDM) in 2012. The slides of these tutorials serve as a helpful companion to the book and can be found at <http://www.zurich.ibm.com/medical-privacy-tutorial/> and <http://www.siam.org/meetings/sdm12/gkoulalas.loukides.pdf>, respectively.

This book is primarily addressed to computer science researchers and educators, who are interested in data privacy, data mining, and information systems, as well as to industry developers, and we believe that the book will serve as a valuable resource to them. Knowledge of data mining or medical methods and terminology is not a prerequisite, and formalism was kept at a minimum to enable readers with general computer science knowledge to understand the key challenges and solutions in privacy-preserving medical data sharing and to reflect on their relation with practical applications. By discussing a wide spectrum of privacy techniques and providing in-depth coverage of the most relevant ones, the book also aims at attracting data miners with little or no expertise in data privacy. The objective of the book is to inform readers about recent developments in the field of medical data privacy and to highlight promising avenues for academic and industrial research.

Dublin, Ireland  
Cardiff, UK

Aris Gkoulalas-Divanis  
Grigorios Loukides

Anonymization of Electronic Medical Records to Support  
Clinical Analysis

Gkoulalas-Divanis, A.; Loukides, G.

2013, XV, 72 p. 23 illus., Softcover

ISBN: 978-1-4614-5667-4