

Aris Gkoulalas-Divanis
Grigorios Loukides *Editors*

Medical Data Privacy Handbook

Medical Data Privacy Handbook

Aris Gkoulalas-Divanis • Grigorios Loukides
Editors

Medical Data Privacy Handbook

Editors

Aris Gkoulalas-Divanis
IBM Research - Ireland
Mulhuddart
Dublin, Ireland

Grigorios Loukides
Cardiff University
Cardiff, UK

ISBN 978-3-319-23632-2 ISBN 978-3-319-23633-9 (eBook)
DOI 10.1007/978-3-319-23633-9

Library of Congress Control Number: 2015947266

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

*Dedicated to my parents, to Elena, and to the memory of my
grandmother Sophia*
–Aris Gkoulalas-Divanis

Dedicated to the memory of my grandmother
–Grigorios Loukides

Preface

The editors started working on medical data privacy in 2009, when they were postdoctoral researchers in the Health Information Privacy Laboratory, Department of Biomedical Informatics, Vanderbilt University. Their work on the topic involved understanding the privacy risks of medical data publishing and developing methods to prevent these risks. Protecting medical data privacy is a challenging problem, since a large volume of complex data must be protected in a setting that involves multiple parties (patients, physicians, carers, researchers, etc.). To address the problem, it is important to develop principled approaches that are specifically geared towards medical data. In addition, it is equally important to increase the awareness of all parties, involved in managing medical data, about privacy risks and approaches for achieving medical data privacy. Thus, the overarching aim of this book is to survey the field of medical data privacy and to present the state-of-the-art approaches to a wide audience.

The structure of the book closely follows the main categories of research works that have been undertaken to protect medical data privacy. Each such category is surveyed in a different part of the book, as follows. Part **I** is devoted to medical data *sharing*. Part **II** focuses on medical data privacy in *distributed and dynamic settings*. Following that, Part **III** examines privacy preservation in *emerging applications* featuring medical data, and Part **IV** discusses medical data privacy through *policy, data de-identification, and data governance*.

Privacy-preserving data sharing requires protecting the identity of patients and/or their sensitive information. For instance, attackers may use external data or background knowledge to learn patients' identity, even though attributes that directly identify patients (e.g., SSNs, phone numbers) have been removed. The problem has been studied extensively in the context of medical data, by the computer science, medical informatics, and statistics communities. However, there is no one-size-fits-all solution and various challenges remain. The purpose of Part **I** of this book is to survey the main research directions in the area of privacy-preserving medical data sharing and to present state-of-the-art approaches, including measures, algorithms, and software tools, that have been designed to solve this problem.

The protection of medical data privacy is particularly challenging, when multiple interrelated parties are involved. For example, medical data practitioners often need to link or exchange different parts of data about a patient, in the context of patient treatment. In addition, medical researchers or insurers may need to access patient information, according to the patient's privacy requirements. In this case, both the objectives of the parties accessing the data and the patient's requirements may change over time. Furthermore, data that are stored or processed in the cloud are vulnerable to a multitude of attacks, ranging from malicious access to intentional data modification. Part II of this book presents approaches focusing on privacy protection in such distributed and dynamic settings. These include approaches for linking data (record linkage), managing data access and patient consent, as well as exchanging health information. Furthermore, a comprehensive survey of privacy concerns and mitigation strategies for medical data in the cloud is presented.

Advances in medical devices and ubiquitous computing enable the collection and analysis of many complex data types, including genomic data, medical images, sensor data, biomedical signals, and health social network data. These data are valuable in a wide spectrum of emerging applications, either alone or in combination with data such as patient demographics and diagnosis codes, which are commonly found in Electronic Health Record (EHR) systems. For example, genomic studies have strong potential to lead to the discovery of effective, personalized drugs, and therapies. However, genomic data are extremely sensitive and must be privacy-protected. Part III of this book surveys privacy threats and solutions for all the aforementioned types of data that are central in emerging applications.

Parts I–III of this book focus on technical solutions that allow data owners (e.g., a healthcare institution) to effectively protect medical data privacy. On the other hand, Part IV focuses on the legal requirements for offering data privacy protection, as well as on the techniques and procedures that are required to satisfy this requirement. More specifically, this part examines key legal frameworks related to medical data privacy protection, as well as data de-identification and governance solutions, which are required to comply with these frameworks. A detailed presentation of the data protection legislation in the USA, EU, UK, and Canada is offered.

This book is primarily addressed to researchers and educators in the areas of computer science, statistics, and medical informatics who are interested in topics related to medical privacy. This book will also be a valuable resource to industry developers, as it explains the state-of-the-art algorithms for offering privacy. To ease understanding by nonexperts, the chapters contain a lot of background material, as well as many examples and citations to related literature. In addition, knowledge of medical informatics methods and terminology is not a prerequisite, and formalism was intentionally kept at a minimum. By discussing a wide range

of privacy techniques, providing in-depth coverage of the most important ones, and highlighting promising avenues for future research, this book also aims at attracting computer science and medical informatics students to this interesting field of research.

Dublin, Ireland
Cardiff, UK
July, 2015

Aris Gkoulalas-Divanis
Grigorios Loukides

Acknowledgements

We would like to thank all the authors, who have contributed chapters to this book, for their valuable contributions. This work would not have been possible without their efforts. A total of 63 authors who hold positions in leading academic institutions and industry, in Europe (France, Germany, Greece, Italy, Luxembourg, Switzerland, and UK), North America, Asia, Australia, and New Zealand, have contributed 29 chapters in this book, featuring more than 280 illustrations. We sincerely thank them for their hard work and the time they devoted to this effort.

In addition, we would like to express our deep gratitude to all the expert reviewers of the chapters for their constructive comments, which significantly helped towards improving the organization, readability, and overall quality of this handbook.

Last but not least, we are indebted to Susan Lagerstrom-Fife and Jennifer Malat from Springer, for their great guidance and advice in the preparation and completion of this handbook, as well as to the publication team at Springer for their valuable assistance in the editing process.

Contents

1	Introduction to Medical Data Privacy	1
	Aris Gkoulalas-Divanis and Grigorios Loukides	
1.1	Introduction	1
1.1.1	Privacy in Data Sharing	2
1.1.2	Privacy in Distributed and Dynamic Settings	3
1.1.3	Privacy for Emerging Applications	3
1.1.4	Privacy Through Policy, Data De-identification, and Data Governance	4
1.2	Part I: Privacy in Data Sharing	5
1.3	Part II: Privacy in Distributed and Dynamic Settings	8
1.4	Part III: Privacy for Emerging Applications	9
1.5	Part IV: Privacy Through Policy, Data De-identification, and Data Governance	11
1.6	Conclusion	13
	References	13
 Part I Privacy in Data Sharing		
2	A Survey of Anonymization Algorithms for Electronic Health Records	17
	Aris Gkoulalas-Divanis and Grigorios Loukides	
2.1	Introduction	17
2.2	Privacy Threats and Models	19
2.2.1	Privacy Threats	19
2.2.2	Privacy Models	19
2.3	Anonymization Algorithms	21
2.3.1	Algorithms Against Identity Disclosure	21
2.4	Directions for Future Research	29
2.5	Conclusion	31
	References	31

3	Differentially Private Histogram and Synthetic Data Publication	35
	Haoran Li, Li Xiong, and Xiaoqian Jiang	
3.1	Introduction	35
3.2	Differential Privacy	36
3.2.1	Concept of Differential Privacy	36
3.2.2	Mechanisms of Achieving Differential Privacy	37
3.2.3	Composition Theorems	39
3.3	Relational Data	39
3.3.1	Problem Setting	39
3.3.2	Parametric Algorithms	42
3.3.3	Semi-parametric Algorithms	42
3.3.4	Non-parametric Algorithms	43
3.4	Transaction Data	48
3.4.1	Problem Setting	49
3.4.2	DiffPart	49
3.4.3	Private FIM Algorithms	50
3.4.4	PrivBasis	50
3.5	Stream Data	51
3.5.1	Problem Setting	51
3.5.2	Discrete Fourier Transform	52
3.5.3	FAST	52
3.5.4	w-Event Privacy	53
3.6	Challenges and Future Directions	54
3.6.1	Variety of Data Types	55
3.6.2	High Dimensionality	55
3.6.3	Correlated Constraints Among Attributes	55
3.6.4	Limitations of Differential Privacy	56
3.7	Conclusion	57
	References	57
4	Evaluating the Utility of Differential Privacy: A Use Case Study of a Behavioral Science Dataset	59
	Raquel Hill	
4.1	Introduction	59
4.2	Background	62
4.2.1	Syntactic Models: k -Anonymity	62
4.2.2	Differential Privacy: Definition	64
4.2.3	Applications	66
4.3	Methodology	67
4.3.1	Utility Measures	69
4.4	Results	70
4.4.1	Variable Distributions	71
4.4.2	Multivariate Logistic Regression	74
4.5	Discussion	79
4.6	Conclusion	80
	References	80

5	SECRET: A Tool for Anonymizing Relational, Transaction and RT-Datasets	83
	Giorgos Poulis, Aris Gkoulalas-Divanis, Grigorios Loukides, Spiros Skiadopoulos, and Christos Tryfonopoulos	
5.1	Introduction	84
5.2	Related Work	86
5.3	Overview of SECRET	87
5.3.1	Frontend of SECRET	87
5.3.2	Backend of SECRET	93
5.3.3	Components	98
5.4	Using SECRET	101
5.4.1	Preparing the Dataset	102
5.4.2	Using the Dataset Editor	103
5.4.3	The Hierarchy Editor	104
5.4.4	The Queries Workload Editor	104
5.4.5	Evaluating the Desired Method	105
5.4.6	Comparing Different Methods	106
5.5	Conclusion and Future Work	107
	References	108
6	Putting Statistical Disclosure Control into Practice: The ARX Data Anonymization Tool	111
	Fabian Prasser and Florian Kohlmayer	
6.1	Introduction	111
6.1.1	Background	112
6.1.2	Objectives and Outline	113
6.2	The ARX Data Anonymization Tool	114
6.2.1	Background	115
6.2.2	Overview	117
6.2.3	System Architecture	120
6.2.4	Application Programming Interface	123
6.2.5	Graphical User Interface	126
6.3	Implementation Details	133
6.3.1	Data Management	134
6.3.2	Pruning Strategies	136
6.3.3	Risk Analysis and Risk-Based Anonymization	138
6.4	Experimental Evaluation	139
6.5	Discussion	142
6.5.1	Comparison with Prior Work	142
6.5.2	Limitations and Future Work	144
6.5.3	Concluding Remarks	145
	References	145

7	Utility-Constrained Electronic Health Record Data Publishing Through Generalization and Disassociation	149
	Grigorios Loukides, John Liagouris, Aris Gkoulalas-Divanis, and Manolis Terrovitis	
7.1	Introduction	150
7.1.1	Identity Disclosure	150
7.1.2	Utility-Constrained Approach	152
7.1.3	Chapter Organization	154
7.2	Preliminaries	155
7.3	Generalization and Disassociation	156
7.4	Specification of Utility Constraints	159
7.4.1	Defining and Satisfying Utility Constraints	159
7.4.2	Types of Utility Constraints for ICD Codes	162
7.5	Utility-Constrained Anonymization Algorithms	163
7.5.1	Clustering-Based Anonymizer (CBA)	164
7.5.2	DISassociation Algorithm (DIS)	165
7.5.3	Comparing the CBA and DIS Algorithms	169
7.6	Future Directions	174
7.6.1	Different Forms of Utility Constraints	174
7.6.2	Different Approaches to Guaranteeing Data Utility	175
7.7	Conclusion	176
	References	176
8	Methods to Mitigate Risk of Composition Attack in Independent Data Publications	179
	Jiuyong Li, Sarowar A. Sattar, Muzammil M. Baig, Jixue Liu, Raymond Heatherly, Qiang Tang, and Bradley Malin	
8.1	Introduction	180
8.2	Composition Attack and Multiple Data Publications	181
8.2.1	Composition Attack	181
8.2.2	Multiple Coordinated Data Publications	183
8.2.3	Multiple Independent Data Publications	183
8.3	Risk Mitigation Through Randomization	185
8.4	Risk Mitigation Through Generalization	187
8.5	An Experimental Comparison	189
8.5.1	Data and Setting	190
8.5.2	Reduction of Risk of Composition Attacks	190
8.5.3	Comparison of Utility of the Two Methods	192
8.6	Risk Mitigation Through Mixed Publications	193
8.7	Conclusion	196
	References	198

9	Statistical Disclosure Limitation for Health Data: A Statistical Agency Perspective	201
	Natalie Shlomo	
9.1	Introduction	201
9.2	Statistical Disclosure Limitation for Microdata from Social Surveys	203
9.2.1	Disclosure Risk Assessment	204
9.2.2	Statistical Disclosure Limitation Methods	207
9.2.3	Information Loss Measures	211
9.3	Statistical Disclosure Limitation for Frequency Tables	213
9.3.1	Disclosure Risk in Whole Population Tabular Outputs	213
9.3.2	Disclosure Risk and Information Loss Measures Based on Information Theory	214
9.3.3	Statistical Disclosure Limitation Methods	217
9.4	Differential Privacy in Survey Sampling and Perturbation	219
9.5	Future Outlook for Releasing Statistical Data	222
9.5.1	Safe Data Enclaves and Remote Access	223
9.5.2	Web-Based Applications	224
9.5.3	Synthetic Data	226
9.6	Conclusion	228
	References	228

Part II Privacy in Distributed and Dynamic Settings

10	A Review of Privacy Preserving Mechanisms for Record Linkage	233
	Luca Bonomi, Liyue Fan, and Li Xiong	
10.1	Introduction	233
10.2	Overview of Privacy Preserving Record Linkage	236
10.2.1	The PPRL Model	236
10.2.2	Taxonomy of Presented Techniques	238
10.3	Secure Transformations	244
10.3.1	Attribute Suppression and Generalization Methods	245
10.3.2	N-Grams Methods	246
10.3.3	Embedding Methods	248
10.3.4	Phonetic Encoding Methods	250
10.4	Secure Multi-Party Computation	251
10.4.1	Commutative Encryption Based Protocols	251
10.4.2	Homomorphic Encryption Based Protocols	252
10.4.3	Secure Scalar Product Protocols	254
10.5	Hybrid Approaches	256
10.5.1	Standard Blocking	257
10.5.2	Sorted Neighborhood Approach	258
10.5.3	Mapping	259
10.5.4	Clustering	259
10.6	Challenges and Future Research Directions	261

10.7 Conclusion	262
References	262
11 Application of Privacy-Preserving Techniques in Operational Record Linkage Centres	267
James H. Boyd, Sean M. Randall, and Anna M. Ferrante	
11.1 Introduction	267
11.1.1 Record Linkage Research Infrastructure	268
11.1.2 Privacy Challenges in Health Record Linkage	270
11.2 Data Governance	271
11.2.1 Legal Obligations	272
11.2.2 Information Governance	272
11.2.3 Separation of Data and Functions	273
11.2.4 Application and Approval Process	273
11.2.5 Information Security	274
11.3 Operational Models and Data Flows	274
11.3.1 Centralized Model	275
11.3.2 Separated Models	276
11.3.3 A Technique to Avoid Data Collusion	278
11.4 Privacy Preserving Methods	278
11.4.1 Privacy Preserving Models	279
11.4.2 Techniques for Privacy Preserving Linkage	279
11.4.3 Requirements of a Privacy Preserving Linkage Technique for Operational Linkage Centres	282
11.5 Conclusion	285
References	285
12 Privacy Considerations for Health Information Exchanges	289
Dalvin Hill, Joseph Walker, and John Hale	
12.1 Introduction	289
12.2 Health Information Exchanges	290
12.2.1 HIE Actors and Systems	290
12.2.2 HIE Models	293
12.2.3 HIPAA, HITECH and HIE Privacy Governance	294
12.3 Privacy Issues with HIEs	295
12.3.1 Patient Expectations and Concerns	296
12.3.2 Tension Between Functionality, Security and Privacy	297
12.3.3 Data Stewardship and Ownership	297
12.4 Principles and Practice of Privacy for HIEs	298
12.4.1 Guiding Principles	298
12.4.2 HIE Privacy in Practice	300
12.5 Emerging Issues	305
12.5.1 Big Data	305
12.5.2 m-Health and Telemedicine	306
12.5.3 Medical Devices	307

12.6	Conclusion	308
	References	308
13	Managing Access Control in Collaborative Processes for Healthcare Applications	313
	Xuan Hung Le and Dongwen Wang	
13.1	Introduction	314
13.2	Related Works	314
13.3	An Illustrative Example: New York State HIV Clinical Education Initiative	316
13.4	Development of the Enhanced RBAC Model	318
13.4.1	Overview of the Enhanced RBAC Model	319
13.4.2	Support Team Collaboration: Bridging Entities and Contributing Attributes	320
13.4.3	Extending Access Permissions to Include Workflow Contexts	322
13.4.4	Role-Based Access Delegation Targeting on Specific Objects: Providing Flexibility for Access Control in Collaborative Processes	322
13.4.5	Integration of Multiple Representation Elements for Definition of Universal Constraints	324
13.4.6	Case Studies to Encode Access Policies for CEI	326
13.5	System Framework for Implementation of Enhanced RBAC	329
13.5.1	System Architecture	330
13.5.2	Encoding of Access Policies	331
13.5.3	Interpretation of Access Control Policies	333
13.5.4	Application Layer	334
13.5.5	Demonstration Tool	334
13.6	Evaluation of the Enhanced RBAC Model	335
13.6.1	Selection of Study Cases	336
13.6.2	Access Permissions Computed with the Enhanced RBAC Model and the CEIAdmin System	339
13.6.3	Comparison Between the Enhanced RBAC Model and the CEIAdmin System	340
13.6.4	Development of the Gold-Standard	340
13.6.5	Measuring Effectiveness Based on Gold-Standard	342
13.6.6	Results	344
13.7	Discussion	345
13.7.1	Features of the Enhanced RBAC Model	345
13.7.2	System Framework for Implementation	349
13.7.3	Evaluation	350
13.7.4	Limitations	353
13.8	Conclusion	354
	References	355

14 Automating Consent Management Lifecycle for Electronic Healthcare Systems	361
Muhammad Rizwan Asghar and Giovanni Russello	
14.1 Introduction.....	361
14.2 Legal Background.....	363
14.2.1 Legal Framework for Consent.....	363
14.2.2 Consent in Healthcare Systems.....	365
14.2.3 Consent Limitations.....	366
14.3 A Case Study.....	368
14.4 Overview of Teleo-Reactive Policies.....	369
14.4.1 TR Policy Representation.....	369
14.4.2 TR Policy Evaluation.....	370
14.5 The ACTORS Approach.....	371
14.5.1 Authorisation Policies.....	373
14.5.2 Policy Templates.....	374
14.5.3 TR Policies.....	375
14.6 Managing Consent in Healthcare Scenarios.....	376
14.7 Related Work.....	382
14.8 Conclusion and Future Work.....	384
References.....	385
15 e-Health Cloud: Privacy Concerns and Mitigation Strategies	389
Assad Abbas and Samee U. Khan	
15.1 Introduction.....	389
15.2 An Overview of the e-Health Cloud.....	391
15.2.1 e-Health Cloud Benefits and Opportunities.....	391
15.2.2 Deployment Models for Cloud Based e-Health Systems.....	393
15.2.3 Threats to Health Data Privacy in the Cloud.....	394
15.2.4 Essential Requirements for Privacy Protection.....	397
15.2.5 User/Patient Driven Privacy Protection Requirements.....	399
15.2.6 Adversarial Models in the e-Health Cloud.....	399
15.3 Privacy Protection Strategies Employed in e-Health Cloud.....	400
15.3.1 Approaches to Protect Confidentiality in the e-Health Cloud.....	400
15.3.2 Approaches to Maintain Data Integrity in the e-Health Cloud.....	402
15.3.3 Approaches to Offer Collusion Resistance in the e-Health Cloud.....	406
15.3.4 Approaches to Maintain Anonymity in the e-Health Cloud.....	407
15.3.5 Approaches to Offer Authenticity in the e-Health Cloud.....	410
15.3.6 Approaches to Maintain Unlinkability in the e-Health Cloud.....	412
15.4 Discussion and Open Research Issues.....	416

15.5 Conclusion	417
References	418
Part III Privacy for Emerging Applications	
16 Preserving Genome Privacy in Research Studies	425
Shuang Wang, Xiaoqian Jiang, Dov Fox, and Lucila Ohno-Machado	
16.1 Introduction	426
16.2 Policies, Legal Regulation and Ethical Principles of Genome Privacy	427
16.2.1 NIH Policies for Genomic Data Sharing	427
16.2.2 U.S. Legal Regulations for Genomic Data	430
16.2.3 Ethical Principles for Genome Privacy	432
16.2.4 Summary	433
16.3 Information Technology for Genome Privacy	433
16.3.1 Genome Privacy Risks	434
16.3.2 Genome Privacy Protection Technologies	434
16.3.3 Community Efforts on Genome Privacy Protection	436
16.4 Conclusion	437
References	438
17 Private Genome Data Dissemination	443
Noman Mohammed, Shuang Wang, Rui Chen, and Xiaoqian Jiang	
17.1 Introduction	443
17.2 Literature Review	445
17.2.1 Privacy Attacks and Current Practices	445
17.2.2 Privacy Preserving Techniques	446
17.3 Problem Statement	447
17.3.1 Privacy Protection Model	448
17.3.2 Privacy Attack Model	448
17.3.3 Utility Criteria	449
17.4 Genomic Data Anonymization	449
17.4.1 Anonymization Algorithm	449
17.4.2 Privacy Analysis	453
17.4.3 Computational Complexity	453
17.5 Experimental Results	454
17.6 Conclusion	458
References	459
18 Threats and Solutions for Genomic Data Privacy	463
Erman Ayday and Jean-Pierre Hubaux	
18.1 Threats for Genomic Privacy	463
18.1.1 Kin Genomic Privacy	465

18.2	Solutions for Genomic Privacy	470
18.2.1	Privacy-Preserving Management of Raw Genomic Data.....	470
18.2.2	Private Use of Genomic Data in Personalized Medicine	472
18.2.3	Private Use of Genomic Data in Research	477
18.2.4	Coping with Weak Passwords for the Protection of Genomic Data.....	481
18.2.5	Protecting Kin Genomic Privacy	484
18.3	Future Research Directions	487
18.4	Conclusion	490
	References.....	490
19	Encryption and Watermarking for medical Image Protection	493
	Dalel Bouslimi and Gouenou Coatrieux	
19.1	Introduction.....	493
19.2	Security Needs for Medical Data	495
19.2.1	General Framework.....	495
19.2.2	Refining Security Needs in an Applicative Context: Telemedicine Applications as Illustrative Example	497
19.3	Encryption Mechanisms: An <i>A Priori</i> Protection.....	498
19.3.1	Symmetric/Asymmetric Cryptosystems & DICOM	498
19.3.2	Block Cipher/Stream Cipher Algorithms.....	499
19.4	Watermarking: An <i>A Posteriori</i> Protection Mechanism	503
19.4.1	Principles, Properties and Applications	503
19.4.2	Watermarking Medical Images	506
19.5	Combining Encryption with Watermarking.....	512
19.5.1	Continuous Protection with Various Security Objectives: A State of the Art.....	512
19.5.2	A Joint Watermarking-Encryption (JWE) Approach	516
19.6	Conclusion	521
	References.....	521
20	Privacy Considerations and Techniques for Neuroimages	527
	Nakeisha Schimke and John Hale	
20.1	Introduction.....	527
20.2	Neuroimage Data	529
20.3	Privacy Risks with Medical Images	530
20.3.1	Neuroimage Privacy Threat Scenarios.....	530
20.3.2	Volume Rendering and Facial Recognition.....	532
20.3.3	Re-identification Using Structural MRI	534
20.4	Privacy Preservation Techniques for Medical Images	535
20.4.1	De-Identification Techniques	535
20.4.2	Privacy in Neuroimage Archives and Collaboration Initiatives	543

20.5	Conclusion	544
	References	544
21	Data Privacy Issues with RFID in Healthcare	549
	Peter J. Hawrylak and John Hale	
21.1	Introduction	549
21.1.1	RFID as a Technology	550
21.2	Dimensions of Privacy in Medicine	553
21.3	RFID in Medicine	556
21.3.1	Inventory Tracking	556
21.3.2	Tracking People	556
21.3.3	Device Management	557
21.4	Issues and Risks	558
21.5	Solutions	562
21.6	Conclusion	563
	References	564
22	Privacy Preserving Classification of ECG Signals in Mobile e-Health Applications	569
	Riccardo Lazzeretti and Mauro Barni	
22.1	Introduction	569
22.2	Plain Protocol	572
22.2.1	Classification Results	575
22.3	Cryptographic Primitives	575
22.3.1	Homomorphic Encryption	576
22.3.2	Oblivious Transfer	577
22.3.3	Garbled Circuits	578
22.3.4	Hybrid Protocols	579
22.4	Privacy Preserving Linear Branching Program	580
22.4.1	Linear Branching Programs (LBP)	580
22.4.2	ECG Classification Through LBP and Quadratic Discriminant Functions	584
22.4.3	ECG Classification Through LBP and Linear Discriminant Functions	586
22.4.4	Complexity Analysis	587
22.5	Privacy Preserving Classification by Using Neural Network	590
22.5.1	Neural Network Design	590
22.5.2	Quantized Neural Network Classifier	593
22.5.3	Privacy-Preserving GC-Based NN Classifier	595
22.5.4	Privacy-Preserving Hybrid NN Classifier	597
22.5.5	Comparison with the LBP Solution	598
22.6	Privacy Preserving Quality Evaluation	599
22.6.1	SNR Evaluation in the Encrypted Domain	599
22.6.2	SNR-Based Quality Evaluation	603
22.7	Conclusion	608
	References	609

23	Strengthening Privacy in Healthcare Social Networks	613
	Maria Bertsima, Iraklis Varlamis, and Panagiotis Rizomiliotis	
23.1	Introduction	613
23.2	Social Networks	615
23.2.1	On-line Social Networks	615
23.2.2	Healthcare Social Networks	616
23.3	Privacy	618
23.3.1	Background	618
23.3.2	Personal and Sensitive Data	619
23.3.3	Privacy Principles	621
23.3.4	Privacy Threats	622
23.4	Privacy Requirements for HSNs	627
23.4.1	Privacy as System Requirement	627
23.5	Enhancing Privacy in OSNs and HSNs	628
23.6	On-line Social Networks in the Healthcare Domain	631
23.6.1	Advice Seeking Networks	632
23.6.2	Patient Communities	632
23.6.3	Professional Networks	633
23.7	Conclusion	633
	References	634

Part IV Privacy Through Policy, Data De-identification, and Data Governance

24	Privacy Law, Data Sharing Policies, and Medical Data: A Comparative Perspective	639
	Edward S. Dove and Mark Phillips	
24.1	Introduction	639
24.2	Overview of Data Privacy Legal Frameworks	642
24.3	Data Privacy Laws and Guidelines	648
24.3.1	The OECD Privacy Guidelines	648
24.3.2	The Council of Europe Convention 108	650
24.3.3	The European Union Data Protection Directive 95/46	652
24.3.4	UK Data Protection Act 1998	656
24.3.5	Canadian Privacy Legislation	658
24.3.6	The HIPAA Privacy Rule	659
24.4	Data Sharing Policies	664
24.4.1	US National Institutes of Health	665
24.4.2	Canadian Data Sharing Policies	666
24.4.3	Wellcome Trust (UK)	669
24.5	Towards Better Calibration of Biomedical Research, Health Service Delivery, and Privacy Protection	671
24.6	Conclusion	674
	References	674

25	HIPAA and Human Error: The Role of Enhanced Situation Awareness in Protecting Health Information	679
	Divakaran Liginlal	
25.1	Introduction	679
25.2	HIPAA, Privacy Breaches, and Related Costs	682
25.3	Situation Awareness and Privacy Protection	685
25.3.1	Definition of Situation Awareness	685
25.3.2	Linking Situation Awareness to Privacy Breaches	686
25.3.3	SA and HIPAA Privacy Breaches	688
25.4	Discussion and Conclusion	693
	References	695
26	De-identification of Unstructured Clinical Data for Patient Privacy Protection	697
	Stephane M. Meystre	
26.1	Introduction	697
26.2	Origins and Definition of Text De-identification	698
26.3	Methods Applied for Text De-identification	701
26.4	Clinical Text De-identification Application Examples	704
26.4.1	Physionet Deid	704
26.4.2	MIST (MITRE Identification Scrubber Toolkit)	705
26.4.3	VHA Best-of-Breed Clinical Text De-identification System	706
26.5	Why Not Anonymize Clinical Text?	708
26.6	U.S. Veterans Health Administration Clinical Text De-identification Efforts	709
26.7	Conclusion	713
	References	714
27	Challenges in Synthesizing Surrogate PHI in Narrative EMRs	717
	Amber Stubbs, Özlem Uzuner, Christopher Kotfila, Ira Goldstein, and Peter Szolovits	
27.1	Introduction	717
27.2	Related Work	719
27.3	PHI Categories	722
27.4	Data	724
27.5	Strategies and Difficulties in Surrogate PHI Generation	725
27.5.1	HIPAA Category 1: Names	726
27.5.2	HIPAA Category 2: Locations	728
27.5.3	HIPAA Category 3: Dates and Ages	729
27.5.4	HIPAA Category 18: Other Potential Identifiers	731
27.6	Errors Introduced by Surrogate PHI	732
27.7	Relationship Between De-identification and Surrogate Generation	732
27.8	Conclusion	733
	References	734

28	Building on Principles: The Case for Comprehensive, Proportionate Governance of Data Access	737
	Kimberlyn M. McGrail, Kaitlyn Gutteridge, and Nancy L. Meagher	
28.1	Introduction.....	737
28.2	Current Approaches to Data Access Governance.....	739
28.2.1	Existing Norms for Data Access Governance.....	739
28.2.2	The Preeminence of “Consent or Anonymize” as Approaches to Data Access Governance.....	740
28.2.3	Existing Data Access Governance in Practice.....	743
28.3	The Evolution of Data and Implications for Data Access Governance.....	744
28.3.1	Big Data.....	744
28.3.2	Open Data.....	745
28.3.3	The Ubiquity of Collection of Personal Information.....	745
28.3.4	The Limits of Existing Approaches to Data Access Governance.....	746
28.4	A Comprehensive Model for Governance: Proportionate and Principled.....	747
28.4.1	Proportionality.....	747
28.4.2	Principle-Based Regulation.....	748
28.4.3	Case Studies Using Proportionate and Principled Access.....	749
28.5	Building on the Present: A Flexible, Governance Framework.....	752
28.5.1	Science.....	754
28.5.2	Approach.....	754
28.5.3	Data.....	755
28.5.4	People.....	755
28.5.5	Environment.....	755
28.5.6	Interest.....	756
28.5.7	Translating Risk Assessment to Review Requirements.....	756
28.5.8	Adjudication Scenarios.....	757
28.6	Conclusion.....	759
	References.....	760
29	Epilogue	765
	Aris Gkoulalas-Divanis and Grigorios Loukides	
29.1	Introduction.....	765
29.2	Topics and Directions in Privacy Preserving Data Sharing.....	766
29.3	Topics and Directions in Privacy Preservation for Distributed and Dynamic Settings.....	768
29.4	Topics and Directions in Privacy Preservation for Emerging Applications.....	769
29.5	Topics and Directions in Privacy Preservation Through Policy, Data De-identification, and Data Governance.....	771

29.6 Conclusion	772
References	772
About the Authors	775
Glossary	815
Index	827

List of Figures

Fig. 3.1	Example: released cell histogram (<i>left</i>) and subcube histogram (<i>right</i>), and N_i is a random Laplace noise (see Sect. 3.2 for Laplace mechanism)	40
Fig. 3.2	Generate synthetic data via parametric methods	41
Fig. 3.3	Generate synthetic data via non-parametric methods	41
Fig. 3.4	Generate synthetic data via semi-parametric methods	41
Fig. 3.5	DExample of private quadtree: noisy counts (<i>inside boxes</i>) are released; actual counts, although depicted, are not released. Query Q (<i>dotted red rectangle</i>) could be answered by adding noisy counts of marked nodes (Color figure online) [6]	45
Fig. 3.6	Taxonomy tree of attributes [29]	48
Fig. 3.7	Tree for partitioning records [29]	48
Fig. 3.8	A context-free taxonomy tree of the sample data in Table 3.1 [5]	49
Fig. 3.9	The partitioning process of Fig. 3.1 [5]	50
Fig. 3.10	The FAST framework [16]	53
Fig. 4.1	Excerpt from doctor's notes	60
Fig. 4.2	Experiment flow chart	67
Fig. 4.3	Histogram of ages from original data (<i>left</i>) and using k-d tree algorithm with $\epsilon = 2.0$, ET = 0.677 (<i>right</i>)	72
Fig. 4.4	Histogram of genders from original data (<i>left</i>) and using cell-based algorithm with $\epsilon = 2.0$ (<i>right</i>)	72
Fig. 4.5	Proportion of variable counts vs. ϵ for all algorithms (for the first reduced dataset)	73
Fig. 4.6	Proportion of variable counts vs. ϵ for all algorithms (for the second reduced dataset)	73
Fig. 4.7	Proportion of variable counts preserved vs. ϵ for k-d tree (for MART_rs1)	74