

Preface

Privacy in statistical databases is a discipline whose purpose it is to provide solutions to the tension between the social, political, economic, and corporate demand for accurate information, and the legal and ethical obligation to protect the privacy of the various parties involved. Those parties are the subjects, sometimes also known as respondents (the individuals and enterprises to which the data refer), the data controllers (those organizations collecting, curating, and to some extent sharing or releasing the data), and the users (the ones querying the database or the search engine, who would like their queries to stay confidential). Beyond law and ethics, there are also practical reasons for data controllers to invest in subject privacy: if individual subjects feel their privacy is guaranteed, they are likely to provide more accurate responses. Data controller privacy is primarily motivated by practical considerations: if an enterprise collects data at its own expense and responsibility, it may wish to minimize leakage of those data to other enterprises (even to those with whom joint data exploitation is planned). Finally, user privacy results in increased user satisfaction, even if it may curtail the ability of the data controller to profile users.

There are at least two traditions in statistical database privacy, both of which started in the 1970s: the first one stems from official statistics, where the discipline is also known as statistical disclosure control (SDC) or statistical disclosure limitation (SDL), and the second one originates from computer science and database technology. In official statistics, the basic concern is subject privacy. In computer science, the initial motivation was also subject privacy but, from 2000 onwards, growing attention has been devoted to controller privacy (privacy-preserving data mining) and user privacy (private information retrieval). In the last few years, the interest and the achievements of computer scientists in the topic have substantially increased, as reflected in the contents of this volume. At the same time, the generalization of big data is challenging privacy technologies in many ways: this volume also contains recent research aimed at tackling some of these challenges.

“Privacy in Statistical Databases 2016” (PSD 2016) was held under the sponsorship of the UNESCO Chair in Data Privacy, which has provided a stable umbrella for the PSD biennial conference series since 2008. Previous PSD conferences were PSD 2014, held in Eivissa; PSD 2012, held in Palermo; PSD 2010, held in Corfu; PSD 2008, held in Istanbul; PSD 2006, the final conference of the Eurostat-funded CENEX-SDC project, held in Rome; and PSD 2004, the final conference of the European FP5 CASC project, held in Barcelona.

Proceedings of PSD 2014, PSD 2012, PSD 2010, PSD 2008, PSD 2006, and PSD 2004 were published by Springer in LNCS 8744, LNCS 7556, LNCS 6344, LNCS 5262, LNCS 4302, and LNCS 3050, respectively.

The seven PSD conferences held so far are a follow-up of a series of high-quality technical conferences on SDC that started eighteen years ago with “Statistical Data Protection-SDP’98”, held in Lisbon in 1998 and with proceedings published by

OPOCE, and continued with the AMRADS project SDC Workshop, held in Luxembourg in 2001 and with proceedings published by Springer in LNCS 2316.

The PSD 2016 Program Committee accepted for publication in this volume 19 papers out of 35 submissions. Furthermore, 5 of the above submissions were reviewed for short presentation at the conference and inclusion in the companion CD proceedings. Papers came from 14 different countries and four different continents. Each submitted paper received at least two reviews. The revised versions of the 19 accepted papers in this volume are a fine blend of contributions from official statistics and computer science.

Covered topics include tabular data protection, microdata and big data masking, protection using privacy models, synthetic data, disclosure risk assessment, remote and cloud access, and co-utile anonymization.

We are indebted to many people. First, to the Organization Committee for making the conference possible and especially to Jesús A. Manjón, who helped prepare these proceedings, and Goran Lesaja, who helped in the local arrangements. In evaluating the papers we were assisted by the Program Committee and by Yu-Xiang Wang as an external reviewer.

We also wish to thank all the authors of submitted papers and we apologize for possible omissions.

Finally, we dedicate this volume to the memory of Dr Lawrence Cox, who was a Program Committee member of all past editions of the PSD conference.

July 2016

Josep Domingo-Ferrer
Mirjana Pejić-Bach

Privacy in Statistical Databases

UNESCO Chair in Data Privacy, International

Conference, PSD 2016, Dubrovnik, Croatia, September

14-16, 2016, Proceedings

Domingo-Ferrer, J.; Pejic-Bach, M. (Eds.)

2016, X, 273 p. 45 illus., Softcover

ISBN: 978-3-319-45380-4