Lecture Notes in Computer Science **11126**

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

More information about this series at http://www.springer.com/series/7409

Josep Domingo-Ferrer · Francisco Montes (Eds.)

Privacy in Statistical Databases

UNESCO Chair in Data Privacy
International Conference, PSD 2018
Valencia, Spain, September 26–28, 2018
Proceedings

Springer

Editors
Josep Domingo-Ferrer
Universitat Rovira i Virgili
Tarragona
Spain

Francisco Montes
Universitat de València
Burjassot
Spain

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-99770-4 ISBN 978-3-319-99771-1 (eBook)
https://doi.org/10.1007/978-3-319-99771-1

Library of Congress Control Number: 2018952341

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Privacy in statistical databases is a discipline whose purpose is to provide solutions to the tension between the social, political, economic, and corporate demand of accurate information, and the legal and ethical obligation to protect the privacy of the various parties involved. In particular, the need to enforce of the EU General Data Protection Regulation (GDPR) in our world of big data has made this tension all the more pressing. Stakeholders include the subjects, sometimes also known as the 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 past 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 2018 (PSD 2018) was held under the sponsorship of the UNESCO Chair in Data Privacy, which has been providing a stable umbrella for the PSD biennial conference series since 2008. Previous PSD conferences were held in various locations around the Mediterranean, and had their proceedings published by Springer in the LNCS series: PSD 2016, Dubrovnik, LNCS 9867; PSD 2014, Eivissa, LNCS 8744; PSD 2012, Palermo, LNCS 7556; PSD 2010, Corfu, LNCS 6344; PSD 2008, Istanbul, LNCS 5262; PSD 2006, the final conference of the Eurostat-funded CENEX-SDC project, held in Rome, LNCS 4302; and PSD 2004, the final conference of the European FP5 CASC project, held in Barcelona, LNCS 3050. The eight PSD conferences held so far are a follow-up of a series of high-quality technical conferences on SDC that started 18 years ago with Statistical Data Protection (SDP) 1998, held in

Lisbon in 1998 and with proceedings published by OPOCE, and continued with the AMRADS project SDC Workshop, held in Luxemburg in 2001 and with proceedings published by Springer in LNCS 2316.

The PSD 2018 Program Committee accepted for publication in this volume 23 papers out of 42 submissions. Furthermore, 11 of these submissions were reviewed for short oral presentation at the conference. Papers came from 15 different countries in four different continents. Each submitted paper received at least two reviews. The revised versions of the 23 accepted papers in this volume are a fine blend of contributions from official statistics and computer science. Topics covered include tabular data protection, microdata and big data masking, synthetic data, record linkage, and spatial and mobility data.

We are indebted to many people. First, to the Organizing Committee for making the conference possible and especially to Jesús Manjón, who helped prepare these proceedings. In evaluating the papers, we were assisted by the Program Committee and by Daniel Baena, Dimitrios Karapiperis, and José Antonio González Alastrué as external reviewers. 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 Prof. Stephen Fienberg, who was a Program Committee member of all past editions of the PSD conference.

July 2018

Josep Domingo-Ferrer
Francisco Montes

Organization

Privacy in Statistical Databases, PSD 2018

Program Committee

Jane Bambauer	University of Arizona, USA
Bettina Berendt	Katholieke Universiteit Leuven, Belgium
Elisa Bertino	CERIAS, Purdue University, USA
Aleksandra Bujnowska	EUROSTAT, European Union
Jordi Castro	Polytechnical University of Catalonia, Spain
Josep Domingo-Ferrer	Universitat Rovira i Virgili, Spain
Jörg Drechsler	IAB, Germany
Khaled El Emam	University of Ottawa, Canada
Mark Elliot	Manchester University, UK
Sébastien Gambs	Université du Québec à Montréal, Canada
Sarah Giessing	Destatis, Germany
Sara Hajian	Eurecat Technology Center, Spain
Alan Karr	CoDA, RTI, USA
Julia Lane	New York University, USA
Bradley Malin	Vanderbilt University, USA
Laura McKenna	Census Bureau, USA
Gerome Miklau	University of Massachusetts-Amherst, USA
Krishnamurty Muralidhar	University of Oklahoma, USA
Anna Oganyan	National Center for Health Statistics, USA
Christine O'Keefe	CSIRO, Australia
David Rebollo-Monedero	Polytechnical University of Catalonia, Spain
Jerome Reiter	Duke University, USA
Yosef Rinott	Hebrew University, Israel
Pierangela Samarati	University of Milan, Italy
David Sánchez	Universitat Rovira i Virgili, Spain
Eric Schulte-Nordholt	Statistics Netherlands, The Netherlands
Natalie Shlomo	Manchester University, UK
Aleksandra Slavković	Penn State University, UK
Jordi Soria-Comas	Universitat Rovira i Virgili, Spain
Tamir Tassa	The Open University, Israel
Vicenç Torra	University of Skövde, Sweden
Vassilios Verykios	Hellenic Open University, Greece
William E. Winkler	Census Bureau, USA
Peter-Paul de Wolf	Statistics Netherlands, The Netherlands

Program Chair

Josep Domingo-Ferrer	UNESCO Chair in Data Privacy, Universitat Rovira i Virgili, Spain

General Chair

Francisco Montes	Universitat de València, Spain

Organizing Committee

Joaquín García-Alfaro	Télécom SudParis, France
Jesús Manjón	Universitat Rovira i Virgili, Spain
Romina Russo	Universitat Rovira i Virgili, Spain

Contents

Microdata and Big Data Masking

Record Linkage

Spatial and Mobility Data