Dr. Peter Christen  
Australian National University

Privacy-preserving record linkage:  
Overview, a taxonomy, and scalable techniques.

Record linkage is the process of identifying which records in two or more databases correspond to the same real-world entity. Three major challenges of this process are (1) achieving high linkage quality, (2) scalability to linking very large databases, and (3) protecting the privacy and confidentiality of personal identifying or confidential data that are used in the linkage process.

This presentation consists of two parts.

In the first part I will provide an overview of record linkage, its applications and challenges, and present a taxonomy we have developed which characterises privacy-preserving record linkage techniques along fifteen dimensions. This taxonomy allows us to identify shortcomings of current techniques, and corresponding directions for future research.

In the second part I will present three specific approaches for scalable privacy-preserving record linkage developed by my PhD student Ms Dinusha Vatsalan. These three protocols are based on public reference values, Bloom filters, and sorted neighbourhood clustering. Experiments conducted on real-world databases with millions of records validate that these approaches are practical in real-world applications in terms of privacy, linkage quality, and scalability.

**Biography:** Peter Christen is an Associate Professor at the Research School of Computer Science at the Australian National University. He received his Diploma in Computer Science Engineering from ETH Zurich in 1995 and his PhD in Computer Science from the University of Basel in 1999 (both in Switzerland).

His research interests are in data mining and data matching (record linkage). He has published over 70 articles in these areas, including in 2012 the book 'Data Matching' published by Springer. He is the principle developer of the Febrl (Freely Extensible Biomedical Record Linkage) open source data cleaning, deduplication and record linkage system.

He has served on the program committees of various data mining related workshops and conferences, and has been on the organization committee for the Australasian Data Mining (AusDM) conference since 2006. He has also served as reviewer for a variety of books and top-tier international journals, and as assessor for the Australian Research Council.
Entering/Exiting the Distinguished Seminar:
☐ Please arrive on time for the seminar and stay for the entire one-hour period.
☐ Late arrivals and early departures are not permitted due to the disruptions these cause.
☐ You are required to stay for the entire one-hour period.

Noise:
☐ When the seminar begins, please stop your conversations.
☐ Wait until the seminar is completely over before putting your materials away in your backpack, standing up, or talking to friends.

Electronic Devices:
☐ Cell phones should be turned off. No talking on cell phones, text messaging, or emailing on laptops during the talk.
☐ No listening to iPods or other electronic recording devices during class.

Common Courtesy:
☐ Food and drink are available at the seminar. Please bring all disposable plates, cups, etc. to the wastebasket before leaving the room.