

Topic 1: Utility and disclosure risk in anonymised data

Josep Domingo-Ferrer

Universitat Rovira i Virgili, Tarragona, Catalonia



Skopje, Sep. 20, 2017



- 10.05 Krishnamurty Muralidhar (U. Oklahoma) and Josep Domingo-Ferrer (U. Rovira i Virgili). “Mahalanobis distance-based record linkage revisited”. **Summary**: It is shown that, when the adversary knows the anonymization method used, she can use that information to improve record linkage.
- .– Sana Rashid (U. Southampton), Jörg Drechsler (IAB) and Robin Mitra (U. Southampton). **NOT PRESENTED** “Accounting for longitudinal data structures when disseminating synthetic data to the public”. **Summary**: Several strategies for generating longitudinal synthetic data sets are evaluated experimentally; synthesis models that are neither pure random nor fixed effects perform best.

10:20 *Shinsuke Ito (Chuo U.), Naomi Hoshino (Natl. Stats. Center Japan), Fumika Akutsu (Stats. Bureau Japan), Ryo Kikuchi (NTT)*. “Investigating new methods for creating anonymized microdata based on Japanese Census Data”. **Summary:** New methods to create anonymized data from the Japan Population Census are investigated; they are evaluated in terms of disclosure risk and utility.

10:35 Questions to authors

10:50 Coffee break

11:10 *Josep Domingo-Ferrer, Sara Ricci and Jordi Soria-Comas (U. Rovira i Virgili)*. “Empirical comparison of anonymization methods regarding their risk-utility trade-off”. **Summary:** A methodology to compare SDC methods in terms of the risk-utility trade-off is proposed that is based on measures not used so far in similar studies.



- 11:25 Jennifer Taub, Mark Elliot and Joseph Sakshaug (*U. Manchester*). “A study of the impact of synthetic data generation techniques on data utility using the 1991 UK samples of anonymised records”. **Summary:** This paper studies the utility of synthetic data; the approach is experimental and several sets of analyses, tests and models are used.
- 11:40 Mira Todorova (*Stats Macedonia*) “Measures for information loss in protected data”. **Summary:** This paper gives an overview of information loss measures for microdata and tabular data.
- 11:55 Questions to authors
- 12:00 Topic 1 general discussion and conclusions