Data Observer

Manfred Antoni and Rainer Schnell*

The Past, Present and Future of the German Record Linkage Center (GRLC)

1 Introduction

Linking data on the same units (such as persons, enterprises or patents) is an increasingly popular research strategy, also in the social sciences (Schnell, 2014b). Since in many cases the required information resides in different organizations, for many applications, a central linkage unit is needed as trusted third party. Because of the federal structure of Germany and the strict data protection jurisdiction, the legal requirements for a linkage operation have to be negotiated for each project separately. Due to this obstacle, no general purpose central linking unit has been available in Germany.

To foster the application of record linkage in the social sciences, the last author and Stefan Bender, the former director of the Research Data Centre (FDZ) of the German Federal Employment Agency at the Institute for Employment Research (IAB), applied for a research grant at the German Research Foundation (DFG) in 2010. They were awarded with this research grant in 2011 to establish the first academic data linkage facility in Germany.[1] The focus of the group in Duisburg was research and development of record linkage methods, while the group at the IAB primarily provided services like consulting of and conducting record

---

[1] The DFG grants were awarded to Rainer Schnell (SCHN 586/17-2) and Stefan Bender (BE 3172/1-2). The funding for the two groups ended in 2016 and 2015, respectively.

---

*Corresponding author: Rainer Schnell, Research Methodology Group, University Duisburg-Essen, Duisburg, Germany, E-mail: rainer.schnell@uni-due.de
Manfred Antoni, Research Data Centre (FDZ), Institute for Employment Research (IAB), Regensburger Str. 100, D- 90478 Nuremberg, Germany, E-mail: manfred.antoni@iab.de

linkages commissioned by academic institutions. The main goal of the GRLC was to increase the number of record linkage applications using administrative data, especially within – but not limited to – the social sciences. Furthermore, GRLC aimed for internationally leading research in record-linkage methodology.

2 Research and development

2.1 Research

Research within the GRLC focused on two topics, which are central to most record linkage problems: Computational speed and privacy. Since most linkage projects involving more than one data holding agency have to use encrypted identifiers (names etc.), record linkage in these cases is usually based on encrypted identifiers. This field of research is called privacy preserving record linkage or PPRL (Vatsalan et al., 2013).

2.1.1 High-speed PPRL

We systematically compared different approaches for high-speed PPRL (Schnell, 2013, 2014a; Sehili et al., 2015). In large record-linkage operations not all possible pairs of records are being compared, but only subsets (blocks). The choice of these blocks is an active field of research within record-linkage (Christen, 2012). The current recommendation (Schnell, 2015) for linking with encrypted identifiers are multibit-trees (Kristensen et al., 2010) with additional encrypted identifiers such as year of birth as blocking variable. Combing external blocks such as year of birth with multi-bit trees allows for privacy preserving linkage of two census scale data sets within a few hours (Schnell, 2014a). For most applications, this solution is sufficient with regard to speed, accuracy and privacy (Brown et al., 2017). Therefore, this combination is provided with the record-linkage software of the GRLC (see Section 2.2).

2.1.2 Security against de-anonymization

However, the more important challenge in PPRL is security against de-anonymization attacks and privacy (Vatsalan et al., 2014). Although not one real-world attack on research data bases has been reported in the scientific literature (Emam et al., 2011), academic research focuses on attacks within the linkage unit. Therefore, the resilience of PPRL encodings against all known cryptographic attacks is widely considered as essential for the successful imple-

mentation of PPRL protocols. It has to be noted that the accepted practice of record linkage within cancer registries in Germany would fail these criteria. This is due to the fact that a simple alignment of the most frequent names and the most frequent encoded names would identify at least some records (Domingo-Ferrer and Muralidhar, 2016). Therefore, research on privacy within the GRLC concentrated on technical measures to prevent these kinds of attacks (Niedermeyer et al., 2014; Schnell and Borgs, 2016). The currently recommended parameter settings (Christen et al., 2017) yield encryptions which can not be successfully attacked by any known method. However, in general, there are no absolute guarantees in cryptography:

> "Breakthroughs in cryptanalysis can happen suddenly and unexpectedly (...). We can choose to design cryptographic primitives [functions, MA/RS] conservatively, but we can never guarantee security against an unknown future" (Martin, 2012, p. 69).

Therefore, data custodians have to be convinced that the privacy demands of the EU privacy regulation are met. These regulations do not require absolute anonymization, but demands that "(...) costs of and the amount of time required for identification, taking into consideration the available technology" has to be considered in evaluating pseudonymisation techniques (Council of European Union, 2016, art 26).

The effort required to break the currently recommend encryption within a linkage unit is considered by current research as more than sufficient to fulfill this EU criterion. Thus, the remaining problem is the actual implementation within a national framework (see Section 2.3).

2.1.3 Data quality

Finally, a further line of research uses record linkage to study data quality. Since access to administrative data is required for this kind of research, this field is tackled by GRLC staff at the IAB. This research uses linked data to determine and quantify errors or sources of bias in either of the originally distinct data sources. Along these lines, Sakshaug et al. (2017) describe the linkage of paradata on the gross sample of a large household panel study to administrative data of the IAB. Based on these data, their future work will focus on, for instance, the utility of such linked data for nonresponse bias adjustments for survey data. Alternatively, ongoing research focuses on the quality of linked data themselves, which may be affected by linkage consent bias, low quality of linkage identifiers and a resulting imperfect linkage result. The overview by Sakshaug and Antoni (2017) puts such potential linkage errors into the context of a more comprehensive total survey error framework.

2.2 Linkage software

During the last decade, the research group of the second author has developed a Java program for record linkage called Merge Toolbox (*MTB*, Schnell et al., 2004). GRLC has extended the capabilities of the program by including routines for privacy preserving record linkage (Schnell et al., 2009), special routines for self-generated-identification codes (Schnell et al., 2010) and updated the input/output-options so that MTB is able to read and write CSV files and native binary Stata-15 files. MTB consists of different modules such as a data editor for record linkage and the main linkage module. Since 2012, MTB has been downloaded by 1104 researchers. MTB is discussed in the leading textbook on record linkage by Christen (2012). An implementation of multibit-trees for privacy preserving record linkage based on Bloomfilters (Schnell, 2015) is provided as a C++ stand-alone program and as a library for R.

Due to the research on privacy preserving record linkage within the GRLC, many new functions for encrypting linkage keys have been developed (Schnell and Borgs, 2016). These new functions have not been implemented in MTB, but in a new R library *PPRL*. Since record linkage often involves large datasets with millions of records, most functions within PPRL have been been optimized for speed using C++ as the main language. The PPRL library will be released within 2017 as an open-source project.

2.3 The politics of implementing a prototype for population covering privacy preserving record linkage in Germany

Administrative databases covering the whole population are rare within Germany. Accessing those databases for research purposes is a challenging problem. Due to the federal structure, most databases are scattered across the federal states (Bundesländer). Therefore, only the registries of social security administrations have regularly been used as population covering databases.[2] Due to the lack of actual regular data linkages across different federal states or across different organizations, data custodians often reject research proposals since no prototype for such linkages exist in Germany. The Duisburg group of the GRLC therefore tried to implement such a prototype for more than a decade.

---

[2] Of course, even these are incomplete: Persons not covered by the general social security system (children, civil servants and self-employed) are not within these databases. However, in Germany, even medical research databases including cancer registries operate within federal states and not nationwide.

Each year, about 714,000 births occur in about 750 hospitals in Germany. For each birth, about 400 medical variables are collected ('perinatal data'). Since about 110,000 newborns are assigned to further medical treatment, additional data ('neonatal data') is available for these cases. These two datasets are linked nationwide. However, about 20% of the records can not be linked due to the lack of information to discriminate between similar records (Schnell and Borgs, 2015). The organization responsible for this linkage is a federal institution, the Federal Joint Committee ('Gemeinsamer Bundesverband', GBA).[3] Hence it has been suggested by a technical working group consisting of medical, mathematical and record linkage experts to include more information, for example an encrypted version of the name of the mother to the two datasets (Schnell et al., 2015). As researchers of the Federal Institute for Data Security (BSI) have been involved in the proposal, it has been adopted by the GBA (Gemeinsamer Bundesausschuss, 2015). The Duisburg group of the GRLC is involved in the technical implementation of the procedure (Meier et al., 2017). This implementation will be the first population covering record linkage in Germany using modern methods of PPRL.

This linkage model has also been suggested for the planned German National Mortality Register (Rat für Sozial- und Wirtschaftsdaten, 2011). GRLC has been consulted repeatedly as technical experts by the Working Group for the Mortality Registry of the of German Data Forum (RatSWD) and the commercial advisor (Prognos, 2013) for the Federal Ministry of Health. Although the federal government failed for two legislation periods to pass the bill for the registry, we are confident that it will happen within the next legislation period and that it will be based on methods developed within the GRLC.

2.4 International cooperation

During the last decade, record linkage has received an increasing amount of attention in official statistics, medicine, economics, demography and the social sciences. Researchers have begun to meet on international conferences on record linkage and established a first international organization, the International Population Data Linkage Network (IPDLN). GRLC has been one of the earliest institutional members of this group and attended all conferences organized by IPDLN. In 2017, IPDLN started its own journal, the International Journal of Population Data Science (IJPDS). The journal's first issue contains the conference proceedings of the IPDLN conference in 2016 and contains four extended abstracts of presentations by members of the GRLC. In 2016, a British group initiated a six month

---

3 For details, see www.english.g-ba.de.

workshop at the Isaac Newton Institute for Mathematical Studies in Cambridge, to which international researchers in record linkage methodology were invited. During this workshop, members of GRLC started cooperations with record-linkage research groups in Vancouver, Canberra and Perth. GRLC applied for research grants with all three groups. Due to this collaborative efforts, two joint papers have been published so far (Brown et al., 2017; Christen et al., 2017).

2.5 Future research

Concerning methodological problems, GRLC will focus future research on solutions for record linkage despite (some) missing identifiers (Ong et al., 2014) and cryptographic techniques for making attacks on encrypted identifiers even harder than today (Schnell and Borgs, 2016). In collaboration with the Canadian and Australian partners, GRLC will develop and test PPRL solutions for census scale data bases. Moreover, the research collaboration of the IAB group with the Leibniz Institute for Educational Trajectories (LIfBi) and the University of Manchester on issues of linkage errors and data quality will be continued.

3 Services

3.1 Completed linkage projects

During the period funded by the German Research Foundation, both groups of the GRLC have conducted numerous linkages on behalf of third parties. The linked data originated from a variety of sources such as individual, household or establishment surveys, administrative data (mainly from the IAB), commercial company data and publicly available data (e.g., German higher education institutions). Table 1 lists some examples of completed linkage projects.

3.2 Future linkage projects

While the GRLC was funded by the DFG, all linkage projects could be performed free of charge. Since the DFG funding ended, any future linkage project done by either group of the GRLC have to be financed either directly by funds of a client or by third-party funding acquired in collaboration with the GRLC.

   The effort necessary for and the feasibility of a linkage project strongly depends on the number and size of the databases to be linked, on the availability

Table 1: Examples of completed linkage projects.

Project partner	Description	Reference
BMAS, LIfBi	Linkage of waves 4 and 5 of Starting Cohort 6 of *National Educational Panel Study* (NEPS)	[a]
Bureau van Dijk, IAB	Linkage of Bureau van Dijk's company database *Orbis* to administrative establishment data of the IAB	Schild (2016)
IAB	Linkage of waves 1 and 5-8 of Panel Study *Labour Market and Social Security* (PASS) to administrative data of the IAB	Antoni and Bethmann (2014)
IAB, Leibniz-Institute for Economic Research (RWI), SOEP	Geocoding of administrative data of the IAB, creation of cluster data to measure neighbourhood effects	Bügelmeyer et al. (2015), Scholz et al. (2012)
Federal Joint Committee (GBA)	Consultation and support for the establishment of a national perinatal register	Schnell et al. (2015)
German Institute for Economic Research (DIW)	Linkage of establishment survey *SOEP-LEE* to administrative establishment data of the IAB	Eberle and Weinhardt (2016), Weinhardt et al. (2017)
German Institute for Economic Research (DIW), IAB, Socio-Economic Panel (SOEP)	Deriving areas with specific shares of migrant population using geocoding as a basis for the *IAB-SOEP migrant sample*	Kroh et al. (2015)
Gesis	Linkage of *PIAAC-L* survey data with administrative employment data of the IAB	Perry and Rammstedt (2016)[b]
Max Planck Institute for Research on Collective Goods	Linkage of data on state exams in law from the Hamm Court of Appeals and the Faculty of Law of the University of Münster	Towfigh et al. (2014)
Munich Institute for the Economics of Aging (MEA)	Linkage of *Saving and Old-Age Provision in Germany* (SAVE) survey data to administrative employment data of the IAB	Coppola and Lamla (2013)
Office for quality assurance in Hesse (GQH)	Record linkage of distinct databases on three levels of stroke therapy	Gramlich (2014)
University of Bielefeld, Collaborative Research Center (SFB) 882	Linkage of survey data of respondents and their partners to administrative employment data of the IAB	Schild and Antoni (2014)
University of Mannheim	Linkage of company data on private equity transactions to administrative establishment data of the IAB	Antoni et al. (2017)

[a] The linkage of previous waves way done by the NEPS working group at the IAB. The resulting data set NEPS-SC6-ADIAB will be made available by the research data centers of IAB and LIfBi in 2018. See Fuß et al. (2016) for more details on the NEPS.

[b] The linked data set called PIAAC-L-ADIAB is available to members of the PIAAC-Leibniz-Network (PIAAC-LN) through the FDZ at the IAB.

of and legal restrictions regarding access to linkage identifiers and on regulations regarding subsequent access to the linked data. The required personnel resources and therefore funding for a linkage project as well as its very feasibility will therefore have to be determined on a project-by-project basis. We thus encourage potential scientific clients or cooperation partners to contact the GRLC to discuss the feasibility of any linkage ideas.[4]

Due to administrative constraints of the GRLC hosting institutions, any linkage project will have to cover personnel and overhead costs of GRLC staff for at least three months full-time equivalent. Even if these financial requirements are fulfilled, a requested linkage project has to be declined by the GRLC if not all legal requirements for a linkage are met. Finally, since both GRLC hosting institutions are public organizations, GRLC might have to decline a linkage project due to other duties. However, both GRLC units welcome suggestions for collaborations on linkage projects.

3.3 Data access

In general, all clients or project partners of the GRLC retain access to their respective linked data. Depending on the legal or proprietary restrictions governing the original databases, storage of and access to the linked data are implemented in different ways. If clients own or are allowed to store all elements of the linked data at their own institution, no restrictions arise from the linkage by the GRLC. However, depending on the project, access to some or all components of the linked data may be restricted, for instance by legal issues. In these cases, the resulting linked data cannot be shared with researchers outside of the original project context. Since all GRLC linkages of data covered by privacy regulations are done within the Research Data Centre (FDZ) of the Federal Employment Agency, the legal rules of the FDZ may apply to these datasets.

The FDZ provides cost-free access to several linked data sets to the general scientific community (see, for instance, Antoni and Seth, 2012; Antoni and Bethmann, 2014; Trübswetter and Fendel, 2016). Some of the underlying linkage operations were done by members of the GRLC, some by staff of the FDZ after consulting the GRLC. Access to these and other linked standard data sets of the FDZ is provided through the FDZ's secure infrastructure. As the administrative data of the IAB are subject to strict social security data protection (see Hochfellner et al., 2014), access to such linked data is limited to on-site use or remote access.[5]

---

4 Any request should be addressed to contact@record-linkage.de.

5 See Eberle et al. (2017) or `http://fdz.iab.de/en/FDZ_Data_Access.aspx` for more details.

While FDZ's standard data sets can be accessed by academic researchers without any fee, this does not apply to custom-built data sets linked by the GRLC. As described above, data access of researchers that own all elements of the linked data is not affected by data linkage. In such cases, linked data can be stored and accessed within any environment permissible for the underlying data. However, as soon as external data are linked with data of the IAB, the combined data may not leave the secure computing environment of the FDZ. As FDZ's potential access modes for linked data, on-site use and remote data access, both require extensive output checking to comply with data protection requirements, access to any customized linked data will be restricted to the project group of the respective client and will be subject to an additional fee.

3.4 Consulting and training

Members of the GRLC have conducted several training workshops on methods of record linkage, both in Germany and different international contexts. These workshops covered a broad range of topics like the prototypical record linkage process, preprocessing, blocking, comparison, classification, privacy preserving record linkage and software options. The training format ranged from university classes over short workshops of up to two days (e.g., within the Joint Program in Survey Methodology, JPSM) to online workshops with video-recorded lectures and live online sessions.[6]

Within the limits of professional courtesy, the GRLC will continue to consult academic researchers on topics such as record linkage techniques, the wording of linkage consent questions, the general feasibility of linkage projects or on suitable software solutions. While such a limited consulting will remain free of charge, any more extensive form of knowledge transfer will be fee-based.[7]

3.5 Online information portal

The GRLC hosts the website www.record-linkage.de. The site contains about 50 pages of information (in English) on record-linkage methodology in general, on privacy preserving record-linkage, record-linkage projects in Germany and record-linkage software. All software and papers (co-)authored by members of the GRLC are available for download.

---

6 See, for instance, https://survey-data-science.net/courses/all/introduction-record-linkage-big-data-applications.

7 Institutions interested in receiving or hosting a training workshop by members of the GRLC are encouraged to address requests to contact@record-linkage.de.

4 Future prospects

The advantages of linking multiple surveys, of linking surveys with existing databases or of linking multiple existing databases containing information on the same units are obvious. Since the amount of information on persons or institutions is increasing, the demand for record linkage will increase as well. The new data protection regulation of the EU (Council of European Union, 2016) allows such linkages for research purposes and asks for the use of pseudonymization techniques. Therefore, research on these techniques will also increase (Schnell et al., 2017). The recently founded International Journal of Population Data Science (IJPDS) will facilitate the publication of research on data linkage. This ever more favorable research climate for linkage will yield more and more requests for linkage solutions. By developing and implementing such solutions, the GRLC will continually contribute to improvements of the international research data infrastructure.

References

Antoni, M., A. Bethmann (2014), PASS-Befragungsdaten verknüpft mit administrativen Daten des IAB : (PASS-ADIAB) 1975-2011. FDZ-Datenreport 03/2014.

Antoni, M., E. G. Maug, S. Obernberger (2017), Private equity and human capital risk. Available at SSRN: https://ssrn.com/abstract=2602771.

Antoni, M., S. Seth (2012), ALWA-ADIAB – Linked Individual Survey and Administrative Data for Substantive and Methodological Research. Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften 132(1): 141–146.

Brown, A.P., C. Borgs, S. M. Randall, R. Schnell (2017), Evaluating Privacy-Preserving Record Linkage Using Cryptographic Long-Term Keys and Multibit Trees on Large Medical Datasets. BMC Medical Informatics and Decision Making 17(1): 83.

Bügelmeyer, E., S. Schaffner, N. Schanne, T. Scholz (2015), Das DIW-IAB-RWI-Nachbarschaftspanel: Ein Scientific-Use-File mit lokalen Aggregatdaten und dessen Verknüpfung mit dem deutschen Sozio-ökonomischen Panel. RWI Materialien 97.

Christen, P. (2012), Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Springer, Berlin.

Christen, P., R. Schnell, D. Vatsalan, T. Ranbadudge (2017), Efficient Cryptanalysis of Bloom Filters for Privacy-Preserving Record Linkage. 628–640 in: J. Kim (ed), The Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham. Springer.

Coppola, M., B. Lamla (2013), Saving and Old Age Provision in Germany (Save): Design and Enhancements. Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften 133(1): 109–116.

Council of European Union (2016), Council regulation (EU) no 679/2016.

Domingo-Ferrer, J., K. Muralidhar (2016), New Directions in Anonymization: Permutation Paradigm, Verifiability by Subjects and Intruders, Transparency to Users. Information Sciences 337–338: 11–24.

Eberle, J., Müller, D., Heining, J. (2017), A modern job submission application to access IAB's confidential administrative and survey research data. FDZ Methodenreport 01/2017 (en).

Eberle, J. and Weinhardt, M. (2016), Record Linkage of the Linked Employer-Employee Survey of the Socio-Economic Panel Study (SOEP-LEE) and the Establishment History Panel (BHP). German RLC Working Paper No. wp-grlc-2016-01.

Emam, K.E., E. Jonker, L. Arbuckle, B. Malin (2011), A Systematic Review of Re-Identification Attacks on Health Data. PLoS One 6(12): e28071.

Fuß, D., J. von Maurice, H.-G. Roßbach (2016), A Unique Research Data Infrastructure for Educational Research and Beyond: The National Educational Panel Study. Jahrbücher für Nationalökonomie und Statistik 236(4): 517–528.

Gemeinsamer Bundesausschuss (2015), Alternative Verfahren zur pseudonymisierten Verknüfung der Daten: Hier für die Leistungsbereiche Geburtshilfe und Neonatologie. techreport, Berlin.

Gramlich, T. (2014), 'STROKES' – Record Linkage der Schlaganfälle in Hessen 2007-2010. German RLC Working Paper No. wp-grlc-2014-03.

Hochfellner, D., D. Müller, A. Schmucker (2014), Privacy in Confidential Administrative Micro Data. Journal of Empirical Research on Human Research Ethics 9(5): 8–15.

Kristensen, T. G., J. Nielsen, C.N.S. Pedersen (2010), A Tree-Based Method for the Rapid Screening of Chemical Fingerprints. Algorithms for Molecular Biology 5.

Kroh, M., S. Kühne, J. Goebel, F. Preu (2015), The 2013 IAB-SOEP migration sample (m1): Sampling design and weighting adjustment. SOEP Survey Papers 271, Berlin.

Martin, K. M. (2012), Everyday Cryptography. Fundamental Principles and Applications. Oxford University Press, Oxford.

Meier, J., R. Schnell, G. Heller (2017), Verknüfung der QS-Verfahren Geburtshilfe und Neonatologie. Technische Dokumentation zur Umsetzung der Verknüfung der Leistungsbereiche Geburtshilfe und Neonatologie. techreport, IQTIG/GBA, Berlin.

Niedermeyer, F., S. Steinmetzer, M. Kroll, R. Schnell (2014), Cryptanalysis of Basic Bloom Filters Used for Privacy Preserving Record Linkage. Journal of Privacy and Confidentiality 6(2): 59–79.

Ong, T.C., M.V. Mannino, L.M. Schilling, Kahn M.G. (2014), Improving Record Linkage Performance in the Presence Of Missing Linkage Data. Journal of Biomedical Informatics 52: 43–54.

Perry, A., B. Rammstedt (2016), The Research Data Center PIAAC at GESIS. Jahrbücher für Nationalökonomie und Statistik 236(5): 581–593.

Prognos (2013), Aufwand-Nutzen-Abschätzung zum Aufbau und Betrieb eines nationalen Mortalitätsregisters: Endbericht.

Rat für Sozial- und Wirtschaftsdaten (2011), Ein Nationales Mortalitätsregister für Deutschland: Bericht der Arbeitsgruppe und Empfehlung des Rates für Sozial- und Wirtschaftsdaten (RatSWD).

Sakshaug, J., Antoni, M., R. Sauckel (2017), The Quality and Selectivity of Linking Federal Administrative Records to Respondents and Nonrespondents in a General Population Sample Survey of Germany. Survey Research Methods 11(1): 63–80.

Sakshaug, J.W., M. Antoni (2017), Errors in Linking Survey and Administrative Data. 557–573 in: P.P. Biemer, E.D.D. Leeuw, S. Eckman, B. Edwards, F. Kreuter, L.E. Lyberg, C. Tucker, B.T. West, (eds.) Total Survey Error in Practice. Wiley, Hoboken.

Schild, C. (2016), Linking "Orbis" company data with establishment data from the German Federal Employment Agency. German RLC Working Paper No. wp-grlc-2016-02.

Schild, C.-J., M. Antoni (2014), Linking Survey Data with Administrative Social Security Data - the Project "Interactions Between Capabilities in Work and Private Life". German RLC Working Paper No. wp-grlc-2014-02.

Schnell, R. (2013), Privacy-Preserving Record Linkage and Privacy-Preserving Blocking for Large Files with Cryptographic Keys using Multibit Trees. 187–194 In: American Statistical Association, editor, Proceedings of the American Statistical Association. American Statistical Association, Alexandria.

Schnell, R. (2014a). An efficient Privacy-Preserving Record Linkage Technique for Administrative Data and Censuses. Journal of the International Association for Official Statistics 30: 263–270.

Schnell, R. (2014b). Linking Surveys and Administrative Data Pp. 273–287. In: U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, S. P. (eds.), Improving Surveys Methods: Lessons from Recent Research. Routledge, Taylor & Francis Group, New York.

Schnell, R. (2015), Privacy preserving record linkage. 201–225 In: K. Harron, H. Goldstein, C. Dibben (eds.), Methodological Developments in Data Linkage. Wiley, Chichester.

Schnell, R., T. Bachteler, S. Bender (2004), A Toolbox for Record Linkage. Austrian Journal of Statistics 33: 125–133.

Schnell, R., T. Bachteler, J. Reiher (2009), Entwicklung einer neuen fehlertoleranten Methode bei der Verknüpfung von personenbezogenen Datenbanken unter Gewährleistung des Datenschutzes. MDA – Methoden, Daten, Analysen 3: 203–217.

Schnell, R., T. Bachteler, J. Reiher (2010), Improving the Use of Self-Generated Identification Codes. Evaluation Review 34: 391–418.

Schnell, R., C. Borgs (2015), Building a National Perinatal Database Without the Use of Unique Personal Identifiers. 232–239 In: 2015 IEEE 15th International Conference on Data Mining Workshops (ICDM 2015), Atlantic City, NJ, USA. IEEE Publishing.

Schnell, R., C. Borgs (2016), Randomized Response and Balanced Bloom Filters for Privacy Preserving Record Linkage. in: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDM 2016), Barcelona, December 12th 2016 – Dec 15th 2016. IEEE Publishing.

Schnell, R., C. Borgs, G. Heller, F. Niedermeyer (2015), Pseudonymisierte Verknüfung der Perinatal- und Neonatalerhebung mit Bloom-Filtern. techreport, Gemeinsamer Bundesausschuss, Berlin.

Schnell, R., A. Richter, C. Borgs (2017), A Comparison of Statistical Linkage Keys with Bloom Filter-Based Encryptions for Privacy-Preserving Record Linkage Using Real-World Mammography Data. 276–283 in: Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017), Setúbal. Scitepress.

Scholz, T., C. Rauscher, J. Reiher, T. Bachteler (2012), Geocoding of German Administrative Data: The Case of the Institute for Employment Research. FDZ Methodenreport 09/2012 (en).

Sehili, Z., L. Kolb, C. Borgs, R. Schnell, E. Rahm (2015), Privacy Preserving Record Linkage with PPJoin. In: Proceedings of the 16th GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web.

Towfigh, E., C. Traxler, A. Glöckner (2014), Zur Benotung in der Examensvorbereitung und im Ersten Examen. ZDRW Zeitschrift für Didaktik der Rechtswissenschaft 1(1): 8–27.

Trübswetter, P., T. Fendel (2016), IAB-SOEP Migrationsstichprobe verknüft mit administrativen Daten des IAB: Version 1 (IAB-SOEP-MIG-ADIAB 7514, Version 1). FDZ Methodenreport 11/2016.

Vatsalan, D., P. Christen, C. O'Keefe, V.S. Verykios (2014), An Evaluation Framework for Privacy-Preserving Record Linkage. Journal of Privacy and Confidentiality 6(1): 35–75.

Vatsalan, D., P. Christen, V.S. Verykios (2013), A Taxonomy of Privacy-Preserving Record Linkage Techniques. Information Systems 38: 946–969.

Weinhardt, M., A. Meyermann, S. Liebig, J. Schupp (2017), The Linked Employer–Employee Study of the Socio-Economic Panel (SOEP-LEE): Content, Design and Research Potential. Jahrbücher für Nationalökonomie und Statistik, forthcoming.