RatSWD Working Paper Series

RatSWD.

German Data Forum

263

Cross national research data: access, legality, ethics and opportunities

Workshop Report

German Data Forum and UK Data Forum

June 2017

Working Paper Series of the German Data Forum (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the German section lists the complete contents of all issues in the series in chronological order.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD or of the Federal Ministry of Education and Research.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD
(since 2014 Regina T. Riphahn; 2009-2014 Gert G. Wagner; 2007-2008 Heike Solga)

Cross national research data: access, legality, ethics and opportunities

Montag, der 19. Dezember 2016

The Royal Statistical Society
12 Errol St, London EC1Y 8LX

Workshop Bericht

Im Dezember 2016 führten das UK Data Forum und der Rat für Sozial- und Wirtschaftsdaten (RatSWD) gemeinsam einen Workshop zum Thema „Cross national research data: access, legality, ethics and opportunities" durch. Ziel des Workshops war der systematische Austausch über aktuelle Entwicklungen beim Forschungszugang zu administrativen Daten in Großbritannien und Deutschland.

Die einzelnen Vorträge boten einen Überblick über sichere Zugangswege zu sensiblen personenbezogenen Daten, die in den beiden Ländern entwickelt wurden. Darüber hinaus bot der Workshop eine Plattform, um über Veränderungen der rechtlichen Rahmenbedingungen beim Zugang zu personenbezogenen Forschungsdaten zu diskutieren. Auch forschungsethische Aspekte wurden adressiert.

Der Schwerpunkt der Diskussion lag jedoch auf Forschungsdateninfrastrukturen und verbreiteten Herausforderungen beim Datenzugang in Europa. Des Weiteren haben sich die Teilnehmenden über Fragen der Verknüpfbarkeit von personenbezogenen Daten in den Sozial-, Wirtschafts-, Umwelt- und Gesundheitswissenschaften auseinandergesetzt. Es herrschte Einigkeit darüber, dass Zugangswege für die Datenanalyse sichergestellt werden müssen, um die Effektivität und Effizienz politischer Maßnahmen auf nationaler und länderübergreifender Ebene abschätzen zu können.

Der Workshop wurde von Tim Holt gemeinsam mit Peter Elias (UK Data Forum) und Stefan Bender gemeinsam mit Claudia Oellers (RatSWD) organisiert. Finanzielle Unterstützung erhielt die Veranstaltung durch das UK Economic and Social Research Council (ESRC) und das Bundesministerium für Bildung und Forschung (BMBF).

Das Workshop-Programm findet sich im Anhang 1 (Appendix 1), die Liste der Teilnehmenden in Anhang 2 (Appendix 2). Die Vortragsfolien zu den einzelnen Abschnitten sind verlinkt und können heruntergeladen werden.

Cross national research data: access, legality, ethics and opportunities

Monday 19th December 2016

The Royal Statistical Society
12 Errol St, London EC1Y 8LX

Workshop Report

This workshop, convened collaboratively by the German Data Forum and the UK Data forum, was designed to explore recent developments in both countries with respect to improved research access to administrative data. Presentations would give participants an overview of the techniques being developed in Germany and the UK to provide secure access to person-level data. The workshop would also provide an opportunity for discussion about the changing legal environment for research access to personal data and the ethical issues surrounding such uses. The main topic for discussion would be data infrastructure issues and common concerns relating to access, both within Europe and more widely, in terms of access to and linkage between personal data for research into social, economic, environmental and health matters. The ultimate aim is to ensure ways of accessing and analysing data to shed light on policy effectiveness and efficiency at both national and supranational levels.

This event was organised by Tim Holt with Peter Elias (UK Data Forum) and Stefan Bender with Claudia Oellers (German Data Forum). It was jointly funded by the UK Economic and Social Research Council (ESRC) and the German Federal Ministry of Education and Research. The agenda for the Workshop can be found at *Appendix 1*, the list of participants at *Appendix 2*. A link to the relevant presentation is given after each heading, below. Due to ill health, Eckart Hohmann was unable to attend. Consequently, the session presented by Dr. Francis Crawley and the subsequent discussion was extended.

New data access arrangements
Presentations on this topic were given by Professor David Hand, Chair of the Board of the UK Administrative Data Research Network (ADRN) and. Stefan Bender, Vice Chair of the German Data Forum (RatSWD)

(See: David Hand, Chair of the Board of the Administrative Data Research Network (ADRN) (pptx file, 86Kb) and

Stefan Bender, Vice Chair of the German Data Forum (RatSWD) (pptx file, 2.1Mb)

Attention was drawn in the presentation to concerns about linking data without compromising confidentiality, and to issues particularly within UK central government departments regarding security of data, resulting in potential delays in releasing, or non-release, of important datasets.

In the ensuing discussion, a further point was raised regarding research carried out on sensitive policy areas, sometimes creating a tension between the tendency of central government departments to retain a degree of control over the dissemination of findings and the researchers' motivation to publish findings. There may be a role here for a *data negotiator*, based on developing a degree of trust between the interested parties, including researchers and data producers. It was agreed there is evidence of an increased willingness on the part of researchers to share data, but this was sometimes not permitted by data holders. There remain legal difficulties in Germany regarding data linkage between large data producers, but it is happening within certain constraints (no common identifiers; in last Census, no retention of data following data linkage).

A further problem area relates to data linkage across scientific fields - for example, where researchers using data in the field of medical sciences wish to link with social science data where issues of individual consent may be less stringent. This was recognized both within the UK and Germany as a research area of high potential value but where further work was needed to encourage data linkage. The ADRN utilised a trusted third party data linkage system, whereby no single computer has access to the identifiers used to link both sets of data. In Germany, there is a keen desire to develop further the possibilities for data access and linkage between, for example, personal data deriving from medical sources and data about social and economic conditions. At present, all such data linkage in Germany requires full consent.

It was further noted that data sharing between countries within the EU and those outside the EU could be difficult following full compliance with the new Data Protection Regulation, although the US-EU data protection umbrella was adopted in Dec 2016. This issue was discussed further after the following presentation.

The EU Data Protection Regulation (DPR) and cross-national data sharing
Francis Crawley, from the Good Clinical Practice Alliance presented on this topic.

(See: Francis Crawley, Good Clinical Practice Alliance (ppt file, 533Kb)

Following the presentation, a question was posed regarding the risk *vs* benefit of research in the context of consent. The DPR does not address the issue of risk/benefit; it is simply a matter of following established rules. During the drafting process, numerous versions of the Regulation were produced, each of which created difficulties in terms of the implications for scientific research; one version in particular would have made the work of the research community impossible, requiring consent for each specific purpose that could be overridden only in the case of a high level of public interest. But who determines 'high public interest'? The judgement will be different in each country. Derogations eventually incorporated into the Regulation allow the data controller to decide on what is in the public interest. For medical research this decision may be relatively straightforward in terms of the health benefits deriving from the research, but in terms of the representativeness of a social group – a particular population being covered by the research – the decision on public interest may be more difficult.

A question was also raised as to whether or not consent is meaningful in the context of the ways in which data are now collected (*e.g.* video, audio recordings in public areas – 'automated observation'). Informed consent in the context of, for example, Google Maps, is meaningless for most users who do not read the terms and conditions but have to agree to them before they can access an app or website. There is also the issue of contextual consent (*e.g.* Facebook). It was noted that these issues had been explored extensively in a recent publication by Helen Nissenbaum (PE to insert ref.)

In response to questions, Francis Crawley noted that implementation of the DPR will be the task of each EU member state, via national data authorities. Consent is a difficult concept for the collection of new forms of data. The role of ethics will carry more weight than consent, given that an extreme power relationship can nullify consent. But consent will not disappear as an issue to be addressed for some time; it is deeply embedded in society and political structures.

Workshop participants discussed further the issue of consent. Points raised included:

- the development of a social compact by which researchers are held responsible to a higher standard than other data users.

- the possibility of returning control to the individual (*e.g.* through the creation of personal data stores).

- the engagement of a privacy panel, which would be involved with decisions on public benefit.

The ethics of using new forms of data for cross-national research
The presentation on this topic was made by Carthage Smith, Senior Policy Analyst OECD and Coordinator Global Science Forum (GSF), with Peter Elias, Chair, GSF Expert Group on Research Ethics and New Forms of Data.

(See: Carthage Smith, Senior Policy Analyst OECD and Coordinator Global Science Forum (GSF) (pptx file, 1Mb)

Workshop participants were provided with a copy of the Executive Summary of a recent report prepared by an Expert Group working under the auspices of the OECD Global Science Forum. The presentation raised questions around:

(i)   the role of ethics review bodies (ERBs);
(ii) whether there are differences in terms of ethics between social/economic and medical/physical science research;
(iii) academic *vs.* industrial/commercial research;
(iv) transparency (*e.g.* clinical trial registers as a possible model for social/economic research).

In response to these questions, Smith and Elias noted the following:

(i)   The full OECD online report (PE to insert link) provides a more detailed account of the potential role of ethics review bodies, together with a useful glossary defining terminology in this area. It was mentioned also that the UK Economic and Social Research Council has recently published new guidelines on this topic. Given the report's recommendation about the important role to be played by ERBs, more resources will be required to enhance their effectiveness.

(ii)   Essentially there is no difference in terms of ethics between medical and social/economic research. Interaction across these areas will make advances across the next decade.

(iii)   The topics and issues addressed in the OECD report apply equally to the private sector, but whether notice will be taken is another question. The use of new forms of data by private sector organisations is aimed at increasing profitability, with ethical issues taking a more subsidiary role. This was discussed by the Expert Group, but the question was beyond its remit for their report. In principle, a process of engagement is desirable. The report did recommend that If data are bought from the private sector for publically-funded research, researchers must be very clear that this has been the case.

(iv)   Transparency via a public register is slowly being addressed – *e.g.* data management plans are required to be included in proposals; the use of Digital Object Identifiers on data sets; use of common meta-language for data description. The point was also made that data librarians have an important role to play in t.

Participants noted that the ethical principle of 'do no harm' is highly relevant in the context of research ethics, but any policy based on research will re-allocate resources in some way and will thus 'harm' some group, albeit a group that has previously had relative advantage over another.

A city-based focus on data access – legal, ethical and technical aspects
This presentation was given by Professor Julia Lane, Center for Urban Science and Progress, New York University.

(See: Julia Lane, Center for Urban Science and Progress, New York University (pptx file, 13.1Mb)

Technologies used in the collection of new forms of data present a broader set of challenges. The example from New Zealand of data coordinated across a variety of agencies to inform and impact upon policy illustrates that those challenges can be addressed.

In the US, the scale of operations in the automated collection of data at the city level makes the notion of informed consent inoperable. Data scientists rather than statisticians are linking data at this level. Beyond the system itself, engaging people is important. An important point here relates to the notion of creating champions within agencies.

Following this presentation, a number of points were raised by workshop participants:

It would be interesting to follow up on the activities of the UK Administrative Data Taskforce: where things have gone wrong, where they went well; data linkage on ex-offenders – how they have survived economically after a period of time. The integrated approach outlined in the presentation looks very interesting indeed.

It was note that the UK Data Services is promoting transparency via initiatives such as trial registries in the social sciences, and via the Open Science Framework.

The scholarships made available within the project outlined by Lane provide a good opportunity for up-skilling. Agencies were charged to place students on these courses. Potential students have to write a proposal, state what data they will bring with them, and need the support of their head of department. Those being trained include both data scientists and those engaged with policy development and evaluation. For both groups an understanding of data is essential.

A proto-type was illustrated in the presentation that needs further development (and co-funding) to promote its credibility and to create a sustainable business model.

Next steps – improving cross-national data access – what needs to be done, how and when?

All participants were requested to produce a short paragraph on what they view as a potentially useful initiative in improving cross-national data access, identifying gaps etc.

More emphasis might be placed on the benefits of research in terms of public engagement, as outlined in Lane's presentation. This had been discussed by the OECD Expert Group who concluded that the benefits of a piece of research are best judged by an ERB, rather than the researchers themselves who will naturally highlight benefits and minimize risks. This balanced judgement could be part of the approvals process, at which stage any beneficial research impact can be promoted. Whilst ADRN needs to focus on problem areas, its role could be enhanced in terms of external profiling.

Caution was advised in being over-optimistic about the ability of Big Data to solve all problems. Computer scientists will be inclined to this view; statisticians are likely to be more conservative.

A degree of risk has to be accepted, reducing risk to zero is impossible. A useful form of words is given in the House of Commons Digital Economy Bill: any agency should 'take reasonable steps to minimise the risk…'

Elias outlined a possible list for future steps, welcoming the comments and contributions of all participants in the Workshop:

1. Research data centres with 'safe settings' (secure access facilities) now exist in the UK and Germany. Good progress has been made in providing research access to a

variety of administrative datasets in each country, but the full potential for collaboration has yet to be realised. What are the obstacles to further data linkage and cooperation?

2. Legal issues in both countries are likely to focus upon the notion of 'consent'. There is a need for monitoring of the impact of the new EU Regulation which will come into force in 2018. The Wellcome Trust has taken a lead in this area, particularly with respect to data from the medical sector. In terms of social and economic data, further continued monitoring is needed.

3. Develop capacity to undertake the types of research discussed in the workshop – NB JL's presentation – that have created much interest from data holders and the research community. Learn from what is happening in the US, take its best elements and act upon these. Capacity is a major concern. Graduates from data science courses and degrees are being taken up immediately by the private sector; we need to attract them to academia = developing capacity for public research.

4. Ethics Review Bodies – How do they operate at the detailed level? What do they do? What are the key skills required? How effective are they? Confidence and faith in ERBs by researchers etc. is essential.

Other ideas invited:

5. (Smith) OECD doing work on data skills for science. There is a need to determine the skills required? Where are the gaps? *E.g.* data stewards, data librarians. Is there a strategy for building skills in data sciences? The UK Office for National Statistics is opening a data science campus at Newport. They have identified gaps, but it is a small part of a bigger issue – 'The need for Data Science'.

6. (Hand) In any plans for further work there is a need to recognise the fast-changing data environment.

7. (Bender) We should look more at the quality of data, issues of trust, privacy, public engagement. Overarching, cross-disciplinary discussion about new forms of data and how data are used is urgently needed.

8. (Alexander) A useful seminar took place last year using a mock ethics panel and a fictional scenario. This experience could be built on.

9. (Proctor) ESRC Centre for Doctoral Training in 'Big Data' may be established in Oct 2017. The British Library and Alan Turing Institute have plans to set up ethics review board.

10. (Fitzgerald) EU to devote resources to Open Science Data Plan?

11. (Elliott) Industrial Strategy Challenge Fund – a opportunity for social scientists and others for more investment in the area of data science?

12. (Oellers) Could there be a permanent role for one organization to coordinate disparate local ethical bodies?

13. (Crawley) How is consent used and viewed by various organisations? Also, re-examine EU guidelines for vulnerable groups. Develop a typology for an ethics committee.

In concluding the workshop, the chair of the UK Data Forum expressed thanks to the UK Economic and Social Research Council and the German Federal Ministry of Education and Research for funding, to the Royal Statistical Society for providing the venue, to all presenters and chairs, and to organisers (especially to Claudia Oellers, Lynne Marston and Margaret Birch).

Cross national research data: access, legality, ethics and opportunities

Monday 19th December, 2016
The Royal Statistical Society
12 Errol Street, London EC1Y 8LX

9:30   Registration

9:50   Introduction: Tim Holt – chair of the UK Data Forum

10:00 – 11:00 New data access arrangements (chair: Tim Holt)
This session will outline recent developments that aim to make administrative data more available for research. How do these arrangements operate? How successful are they? What obstacles to improved access remain and how are these being addressed? What are the opportunities for cross national research, particularly using data such as social security records, tax records, etc.

Speakers: David Hand, Chair of the Board of the Administrative Data Research Network (ADRN)
Stefan Bender, Vice Chair of the German Data Forum (RatSWD)

11:00 – 11:15 Morning coffee/tea

11:15 – 12:30 The EU Data Protection Regulation (DPR) and cross-national data sharing (chair: Stefan Bender)
This session will address issues relating to the implementation of the recently adopted EU DPR and its potential impact on cross national data sharing in the social and medical sciences. What will be the immediate and the longer-term consequences for researchers?

Speakers: Francis Crawley, Good Clinical Practice Alliance

12:30 – 13:15 Lunch

13:15 – 14:00 The ethics of using new forms of data for cross-national research (chair: Stefan Bender)
This session will be based around the recently published OECD report, placing particular emphasis on ethical issues that arise with cross-national research teams and/or cross-national data sharing.

Speakers: Carthage Smith, Senior Policy Analyst OECD and Coordinator Global Science Forum (GSF)
Peter Elias, Chair, GSF Expert Group on Research Ethics and New Forms of Data

**14:00 – 15:00 A city-based focus on data access – legal, ethical and technical aspects
                (chair: Peter Elias)**

There is a growing interest in the 'data-connected city' – linking a wide variety of data about city workers/dwellers to facilitate research designed to improve the welfare of the city. This raises issues about data access, ethics and the legality of such research. How are these issues being overcome? What can we learn from these efforts?

Speaker: Julia Lane, Center for Urban Science and Progress, New York University

15:00 – 15:15 Break

**15:15 – 16:30 Next steps - improving cross-national data access - what needs to be done,
                how and when?**

16:30 – 17:30 Drinks reception

Speakers

David Hand	Imperial College, London
Peter Elias	University of Warwick
Francis Crawley	Good Clinical Practice Alliance
Julia Lane	Center for Urban Science and Progress
Stefan Bender	RatSWD
Carthage Smith	OECD
Tim Holt	UK Data Forum

Admin

Lynne Marston	University of Warwick
Margaret Birch	University of Warwick

Participants

Anne Alexander	University of Cambridge
Natalie Banner	Wellcome Trust
Dan Edwards	Government Office for Science
Joseph Ellery	ESRC
Jane Elliott	ESRC
Rory Fitzgerald	City University
Sharon Heys	Swansea University
Gundi Knies	Essex
Albert King	UKDF
Rebecca Leithall	ESRC
Joanna Littlechild	Department for Work and Pensions
Katherine McNeill	Essex University
Lucy Martin	ESRC
Claudia Oellers	German Data Forum
Rob Proctor	University of Warwick
Barbara Sanchez Solis	Austrian Social Science Data Archive
Anna Schneider	Napier University