

National Research Data Services for Economics in Germany and the Role of the ZBW

*International Conference on Economics
and Business Information*



*Sven Vlaeminck
Publication Services
ZBW – Leibniz Information Centre for Economics
Berlin, 2016.04.20*

Overview:

1. The German national Landscape for RDM-Services

- German Universities and RDM
- Learned Societies in Economic Sciences
- Storage of & Access to Research Data
- Subject-specific DOI-Registration of Research Data in the Social Sciences

2. ZBW's Approach in RDM

- Our Principles in RDM
- Some of ZBW's Activities in RDM
 - da|ra
 - EDaWaX / ZBW journal data archive
 - SowiDataNet
 - Guidebook on research data (management)

3. Summary and Lessons learned

1) THE GERMAN NATIONAL LANDSCAPE FOR RDM SERVICES

1.1 The German national landscape for RDM

On the level of universities (non subject-specific):

- Within the last years, at least seven universities adopted policies on RD. Some also provide assistance in RDM-topics (e.g. competence centers or EU-offices).
- In addition, some implemented university-wide RD-infrastructures to support researchers in RDM:
 - LMU Munich
 - University of Bielefeld
 - University of Heidelberg
 - University of Mannheim
 - TU Berlin
- Also, the German Rectors' Conference (HRK) asks the universities to implement RD-Policies and RDM-support at university level.

HUMBOLDT-UNIVERSITÄT ZU BERLIN



Humboldt-Universität zu Berlin Research Data Management Policy

Preamble

Handling of research data responsibly is critical for confirming research results, for the advancement of science and for the dissemination of knowledge. This policy is addressed to all researchers at Humboldt-Universität zu Berlin both in their role as independent researchers, and as teachers and supervisors. Their obligation includes instructing students and doctoral candidates about handling of research data properly, as well as imparting subject-specific skills and standards.

What is research data?

UNIVERSITÄT HEIDELBERG | ZUKUNFT SEIT 1386

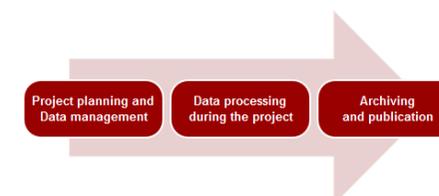
Contact | Deutsch

Heidelberg University > Institutions > Competence Centre for Research Data >

Competence Centre for Research Data

In accordance with the university's [Research Data Policy](#) the Competence Centre for Research Data offers scientists at Heidelberg University central services for archiving and publishing their research data. The Competence Centre is a joint facility of the university's [Computing Centre](#) and the [University Library](#).

Our services:
Support at every stage of the research process



1.2 The German national landscape for RDM

On the level of learned societies:

- The learned societies for economics (*German Economic Association - VfS*) and business studies (*German Association for Business Research - VHB*) name research data management in their code of conduct.
- “Research should be transparent and comprehensible. The underlying assumptions should be clear. For empirical work, the datasets and programs used should be made available to replicate the results, if legally possible.” (VfS)
- The *German Association for Business Research (VHB)* even published a detailed paper on how to document and publish research data.



Themenfeld 5: Umgang mit Daten und Dokumentation des Forschungsprozesses

Die Ergebnisse von Forschungsarbeiten sind nur intersubjektiv überprüfbar, wenn auch ein Zugang zu den Daten gewährt wird bzw. im Nachgang auch deren Überprüfbarkeit sichergestellt wird. Wie kann dieser Zugang sichergestellt werden? Für welchen Zeitraum? Wie kann mit vertraulichen Daten umgegangen werden? Wie werden die Interessen der Autor/innen gewahrt, die die Daten unter Umständen mit viel Aufwand gesammelt haben und nun auch alleine davon profitieren wollen? Wie sollte eine reversionssichere Dokumentation von Forschungsarbeiten aussehen? Welche Sachverhalte insbesondere bei empirischen Untersuchungen (z.B. Projektpläne, Publikationsprojektpläne, Wechsel von Autorenschaften, Anpassung des Rohdatensatzes durch Bereinigungen usw., inhaltliches vs. rechtliches Eigentum an Daten und Erkenntnissen, Art der Mitarbeit usw.) soll in welcher Weise dokumentiert werden, um später zur Klärung von Sachverhalten beitragen zu können? Wie sind diese Unterlagen bzw. auch die verwendeten (Roh-)Daten zu archivieren – in welchem Format, mit welchen Sicherungsinstrumenten und für welchen Zeitraum? Wie kann sichergestellt werden, dass die Arbeitsschritte des Forschers aus den ursprünglichen Daten auch nachvollzogen werden können?

Ein notwendiges Kriterium für Wissenschaft besteht in der **Nachvollziehbarkeit** der erzielten Ergebnisse. Zeitschriften haben deshalb von je her darauf geachtet, dass der Forschungsprozess zur Erzielung der in einem Aufsatz berichteten Ergebnisse angemessen dokumentiert ist. Anders als in den Naturwissenschaften hat sich allerdings keine Kultur der Replikation von Ergebnissen herausgebildet, weshalb auch die Angemessenheit bisheriger Dokumentationen in der Regel nicht geprüft wurde. Ausgelöst durch verschiedene

1.3 The German national landscape for RDM RatSWD. German Data Forum

On the level of research data storage & access:

Currently, the landscape subdivides in two parts:

- For large surveys and data provided by federal statistical offices almost 30 specialised research data centres exist.
- For the ‘long tail’ of research data generated by projects and non-university research institutes, currently almost no storing opportunities for RD are available.

Research Data Centers – a manifold supply of data

Today’s researchers have an advanced and sustainable research infrastructure at their disposal which is the result of independent initiatives from within the scientific community. A decentralized network of almost 30 data centers provides easy and low-cost access to a wide range of research-relevant data.

All data centers are accredited by the German Data Forum according to uniform and transparent standards.



1.4. The German national landscape for RDM



DOI-Registration:

- In Germany, the DOI-registration of research data is organised by subject.
- For social sciences' research data the DOI-registration agency da|ra (www.da-ra.de/en) takes over the task to mint PIDs for these resources.
- da|ra is a member of the DataCite consortium.
- Joint service of GESIS (Leibniz Institute for the Social Sciences) & ZBW.
- Currently 65 data centres and other customers.
- To date, ~433.000 DOIs have been registered by da|ra.



2) ZBW'S APPROACH IN RDM

2.1 Our principles in RDM...

Against the backdrop that ZBW is a specialist library with a national duty, our **general approach** is characterised by a few principles:

- With our projects and services, we focus on **organisations / institutions** rather than on single researchers,
 - We act **in close cooperation** with the economic research,
 - We orient ourselves to the **needs of the community** and seek **constant feedback** from the target group,
 - We work **with** rather than (only) **for** researchers,
 - We act with a **strong subject-specific** point of view, without neglecting overall developments and discussions,
 - We think it us useful to have a **division of responsibilities** in RDM based on complementary skills.
-

2.2 Some of ZBW's activities in RDM

AUFFINDEN
FORSCHUNGSDATEN IN DEN
ZITIERN
SOZIAL- UND WIRTSCHAFTSWISSENSCHAFTEN
DOKUMENTIEREN

ZBW Leibniz-Informationszentrum
Wirtschaft
Leibniz Information Centre
for Economics

gesis Leibniz-Institut
für Sozialwissenschaften

RatSWD Rat für Sozial- und
Wirtschaftsdaten

EDaWaX

European Data Watch Extended

[About](#) [Partners](#) [Events](#) [Downloads](#) [Contact](#) [Imprint](#)

Action Poll- tools for working with research data

Posted: June 16th, 2015 | Author: [Sven](#) | Filed under: [Data Sharing](#), [Research Data](#) | Tags: [data management](#), [poll](#), [RDA](#), [tools](#) | [No Comments](#) »



You are a researcher and you are working with data? Great!

This short, anonymous poll- targeted at researchers- aims to characterise the range of research data management tools that are used across domains and regions. It is being conducted by [Research Data Alliance \(RDA\) Long Tail of Research Data Interest Group](#).

The results will be reported at the [next RDA meeting in Paris](#) in September 2015 and also made available through the Long Tail Interest Group [webpage](#).

The poll will be open from June 15 to July 24, 2015 and is available on <https://www.surveymonkey.com/r/TVQ7RY6>.

It is only one-page long and will take less that 5 minutes to complete.

For more information, please contact: [Kathleen Shearer](#), Executive Director, [COAR](#) [Read the rest of this entry](#) »

[Empfehlen](#) [Tweet](#)

[Subscribe to RSS](#)

Search for:

June 2015

M	T	W	T	F	S	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

« [Apr](#)

Project Partners:

ZBW Leibniz-Informationszentrum
Wirtschaft
Leibniz Information Centre
for Economics

DIW SOEP

Max Planck Institute for
Innovation and Competition

RatSWD
Rat für Sozial- und
Wirtschaftsdaten



da|ra

Registration agency for
social and economic data



- da|ra is the German DOI-registration agency for social and economic data. da|ra is part of DataCite.
- **Goals:** Enable a stable and reliable citability of research data (most often for large datasets in RDCs). Facilitate reuse and discovery of datasets.
- **Target group:** Mainly economic research data centres & OA journals.
- **Partners:** Jointly run by GESIS (social sciences) & ZBW (economics)

Some of our customers:

- **Goals:**
 - Development and implementation of a software to support the management of publication-related research data in economics journals (= software for a data archive; similar to Dataverse).
 - Analyses of the current state of data sharing in economic sciences.
 - Analyses on data management practises of journals in economics and business research.
 - Advise of journal editors (data policies and data archives)
 - **Target group:** Editors of scholarly journals, publishing houses.
 - **Partners:** German Data Forum (RatSWD), Research Data Centre (RDC-SOEP), Max Planck Institute for Competition and Innovation (MCIER)
- www.edawax.de (en) and www.journaldata.zbw.eu (en)

Funded by :



SowiDataNet

- **Goals:**

- Development of a technical infrastructure (data repository) to discover and to access institutionally created datasets (for quantitative data).
- Constitution of a research data association for economics and the social sciences in Germany.
- Simple and low-threshold infrastructure component to store research data and for long-term access to data.

- **Target group:** Institutional projects/research groups without personal and/or financial resources for the archiving and documentation of self-compiled datasets.

- **Partners:** DIW, GESIS, WZB, ZBW.

➤ www.sowidatanet.de (de)



„Guidebook on Research Data“

- **Goal:** Give information and advice in RDM topics that are not part of universities' curricula.
- **Target group:** young academics & lecturers
- **Content (27 pages A4):**
 - What are useful sources to procure data?
 - How to cite research data?
 - How to document self-compiled datasets?
 - How and where to store these datasets?
- **Partners:** ZBW, GESIS, German Data Forum
- Version 2.0 released in Spring 2015.
- www.auffinden-zitieren-dokumentieren.de (de)

Wo finde ich relevante Forschungsdaten für meine wissenschaftliche Arbeit?



Möchte man für eine wissenschaftliche (Qualifikations-)Arbeit eine empirische Auswertung durchführen, steht man vor der Frage, wo sich passende Daten für statistische Auswertungen finden lassen. Zwar liefern häufig der Betreiber einer solchen Arbeit, Kollegen und nicht zuletzt die einschlägige Literatur Hinweise auf mögliche Datenquellen, jedoch ist das Datenerhalt in den Sozial- und Wirtschaftswissenschaften nicht nur auf dem besten Stand.

Die folgende Tabelle verdeutlicht beispielhaft die unterschiedliche Herkunft von quantitativen Forschungsdaten. Quantitative Forschungsdaten werden in diesem Wegweiser weitgehend angeklammert. Wichtig ist jedoch, in nachfolgenden Verweisen dieser Broschüre auch qualitative Forschungsdaten stärker zu thematisieren.

TABELLE 1
Beispiele für Forschungsdaten in den Sozial- und Wirtschaftswissenschaften

Statistische Daten	Survey-Daten	Prozessproduzierte Mikrodaten
Statistische Ämter des Bundes und der Länder: - Bildung - Präzision - Gesundheit - Landwirtschaft - Umwelt/Energie - VGR (VGR) - Zahl der Beschäftigten - Zahl der Erwerbslosen/Arbeitslosen	- ALLER - Community Innovation Survey - EU IPS - EU SIC - GESIS - Gruppenspezifische Mitarbeiterbefragungen - Ifo-Geschäftsklimaindizes - Innovationsmonitor des IWI - Mannheim Innovation Panel - NIPS - PESTIM - PAAC - PIA - SIAKE - SOEP - StatistikPlus - ZEW HighTech Gründungspanel	Öffentliche Verwaltung: - Bundesinnen- und Vermögenskataster - Kraftfahrzeugkennzeichen - Meldestellen - Steuerdaten (Einkommensteuer und Umsatzsteuer) Wirtschaftsdaten: - Aktienkurse - Bilanzdaten - Strukturmerkmale Reisungsdaten für Arbeit: - Daten zur Arbeitsmarkt- und Berufsinformation - Bilanzdaten - Daten zu Arbeitgeberleistungen und Arbeitsmarktsegmenten - Sozialversicherungsdaten

Die vorangestellten Datenbanken sind im Regelfall kostenpflichtig.

Zugangsweg
Aufgrund der Datenschutz-Notwendigkeiten und anderen gesetzlichen Bestimmungen unterscheiden sich die Zugriffsmöglichkeiten auf sozial- und wirtschaftswissenschaftliche Forschungsdaten je nach Datentyp zum Teil erheblich. Darüber hinaus ist es für den Zugriff relevant, ob es sich um öffentlich verfügbare und kostenlos bzw. gegen eine geringe Gebühr bereit gestellte Daten handelt oder um Daten, die durch spezifische Lizenzvereinbarungen und/oder aufbereitet wurden und als Produkt verkauft werden. Für sämtliche Datenbanken können die Gebühren, die statistische Ämter für Mikrodaten verlangen müssen, je nachdem wie groß der für die Datenabfrage zu machende Bereich ist.

Aufgrund dieser Differenzierungen lassen sich die Zugangswegtypen in drei Typen einteilen:

- **Aggregierte Statistische Daten (= Mikrodaten)**
Diese Daten sind meist bereinigt und bieten eine nach Kategorisierung online direkt zu beschaffen. Gemeinsame Indikatoren ermöglichen die Suche dem jeweiligen Datentyp. Im Falle von disaggregierten Daten können darüber hinaus besondere Anforderungen an den Zugang bestehen.
- **Statistik-Mikrodaten**
Für statistische Mikrodaten gelten besondere Schutzbestimmungen durch die Bundes- und Landesdatenschutzgesetze sowie das Bundesdatenschutzgesetz. So muss die Erhebung von personenbezogenen Daten immer an ein bestimmtes Forschungsziel geknüpft sein, welches als so relevant erachtet wird, dass die Erhebung personenbezogener Daten zu rechtfertigen ist.
Der Schutz von personenbezogenen Daten ist zudem ein gesetzlicher Auftrag, der durch eine Anonymisierung der Daten gewährleistet wird, wobei der Grad der Anonymisierung von der Art und der Kontrolle der Nutzung abhängt. Aus diesem Grund können statistische Mikrodaten nicht ohne Weiteres von dem Webserver der Forschungsstelle heruntergeladen werden. Dies gilt in ähnlicher Form auch für Firmendaten, wie beispielsweise die Mikrodaten der Projekte AfD- oder Konfigo.
Grundsätzlich finden sich bei Organisationen, die solche Daten anbieten, verschiedene Stufen der Anonymisierung, die beispielhaft anhand einer Grafik der statistischen Ämter dargestellt werden (siehe Abbildung 1).
• Vollständig anonymisierte Mikrodatensätze werden auch für die Öffentlichkeit angeboten, es handelt sich hier um so genannte **Public Use Files (PUF)** oder um **CAMPUS-Files**. Diese Datensätze unterscheiden sich allerdings hinsichtlich ihrer Informationsumfang, der von Datenherkunft und/oder Herkunft ist, deutlich von den Statistiken-Dateien.

ABILDUNG 1
Nutzungswege für Mikrodaten der Statistischen Ämter
Quelle: eigene Darstellung auf Basis von Kapitel 1.2 des Leitfadens Forschungsdaten, Statistisches Bundesamt, Ausgabe März 2015, S. 16.



On-Site-Nutzung
Datenabfrage innerhalb des IZ/ Statistischen Amtes
Kontrollierte Datenverarbeitung (KDF)

Off-Site-Nutzung
Datenabfrage außerhalb des IZ/ Statistischen Amtes
Campus-File vs. Public Use File (PUF)
Statistiken-Datei (SUF)

Legende: hoch (blau), niedrig (gelb)

3) SUMMARY AND LESSONS LEARNED

3. Summary and lessons learned

- In Germany, we observe a coexistence of activities by universities, non-university research institutes and learned societies in RDM.
- No coordinated approach yet, but many institutions are in a state of flux.
- Fragmented RD-infrastructure: Good infrastructure for relevant studies (-> RDCs), but poor support for RD created by projects and smaller institutions.
- ZBW intends to support economists in RDM, especially in areas where currently no services exist (triggered by expressed demands from the community).

(Some) lessons learned:

- For developing suitable RDM-services, it is crucial to work close with the researchers!
 - Services and infrastructures must be low-threshold and must not interfere with the researcher's work.
 - The benefits of the support offered must be clearly visible and tangible.
-

Panel discussion on data sharing at the annual meeting of the learned society 'German Economic Association' (VfS)



EDaWaX-workshop with editors of journals in economics and business studies.

Thank you for your attention!

➤ Questions?



Credit: ['Questions'](#) by Oberazzi on flickr.com. [License](#): CC-BY-SA-NC 2.0

Sven Vlaeminck

ZBW – Leibniz-Information Centre for Economics

Hamburg / Germany

Tel.: +49 40 – 42834 419

Email: s.vlaeminck@zbw.eu