

# EDaWaX

European Data Watch Extended

Forschungsdaten und Forschungsdateninfrastrukturen in den  
Wirtschaftswissenschaften.

Sven Vlaeminck | Leibniz-Informationszentrum Wirtschaft (ZBW)  
1. Oktober 2013 | Open Access Tage 2013 | Hamburg

Funded by  
**DFG**

**EDaWaX**  
European Data Watch Extended

**RatSWD.**  
Rat für Sozial- und  
WirtschaftsDaten

**ZBW** Leibniz-Informationszentrum  
Wirtschaft  
Leibniz Information Centre  
for Economics

# Inhalte der Präsentation:

- > Wirtschaftswissenschaftliche Forschungsdaten
  - Charakteristika ökonomischer Forschungsdaten
  - Wofür werden Forschungsdaten benutzt?
  - Rechtliche Aspekte verschiedener Arten von Forschungsdaten
- > Data Sharing in den Wirtschaftswissenschaften
- > Forschungsdateninfrastrukturen für die Wirtschaftswissenschaften.
  - Die deutschen Forschungsdatenzentren (FDZ)
  - Forschungsdateninfrastrukturen für kleine und mittlere Projekte
  - Sonderfall: Infrastruktur für publikationsbezogene Forschungsdaten

# Wirtschaftswissenschaftliche Forschungsdaten

Eine Typisierung anhand

- der Analyse der Inhalte von publikationsbezogenen Forschungsdatenarchiven, sowie
- verschiedenen Leitfadeninterviews mit Wirtschaftsforscher/inne/n.

(Dank an Daniel Bahls, Jesper Zedlitz und Guido Scherp - alle ZBW).

Funded by  
**DFG**

**EDaWaX**  
European Data Watch Extended

**RatSWD.**

**ZBW** Leibniz-Informationszentrum Wirtschaft  
Leibniz Information Centre for Economics

# Charakteristika ökonomischer Forschungsdaten

- > Forschungsdaten in der BWL und der VWL unterscheiden sich markant.
  - > BWL: Fokus auf Unternehmensdaten jedweder Couleur
  - > VWL: Fokus auf volkswirtschaftlich relevante Daten
- > Häufig genutzte Software:
  - > Excel
  - > Stata und R (Statistikprogramme)
  - > MatLab / Gauss (numerische Simulation, Datenanalyse)
  - > GAMS (General Algebraic Modeling System)
- > Größe der Datensätze: Von einigen MB bis zu wenigen GB

# Wofür werden Forschungsdaten benutzt? (v.a. VWL)

## > Ökonometrische Studien

- Aufgabe: theoretische ökonomische Modelle mittels statistischer Analysen prüfen.
- Oft Datensätze aus unterschiedlichen Quellen (eigene + externe)

## > Simulationen

- Aufgabe: z.B. Verhalten ökonomischer Märkte und Akteure wird unter sich verändernden Bedingungen simuliert.
- Datensätze aus verschiedenen Quellen + Software (z.B. Fortran)

## > Experimentelle Wirtschaftsforschung

- Aufgabe: Testpersonen werden z.B. mit ökonomischen Fragestellungen konfrontiert. Aus Ergebnissen werden Annahmen über das Verhalten von wirtschaftlichen Akteuren geschlossen.
- Datensätze (teils selbst erstellt), Versuchsskizzen,...

# Rechtliche Aspekte: Grob drei Arten von wirtschaftswissenschaftlichen Forschungsdaten

## > „Freie“ oder „offene“ Daten

- Selbst erstellte Datensätze, Fragebögen und Software (ohne Personenbezug)
- Aggregierte Datensätze der statistischen Ämter (mit Einschränkungen)

## > Mikrodaten

- Personenbezogene Daten (z.B. SOEP)
- Firmendaten (z.B. AFiD)

## > Proprietäre Daten, z.B.

- Firmendaten / Börsendaten (u.a. Thomson Reuters u. Bloomberg)

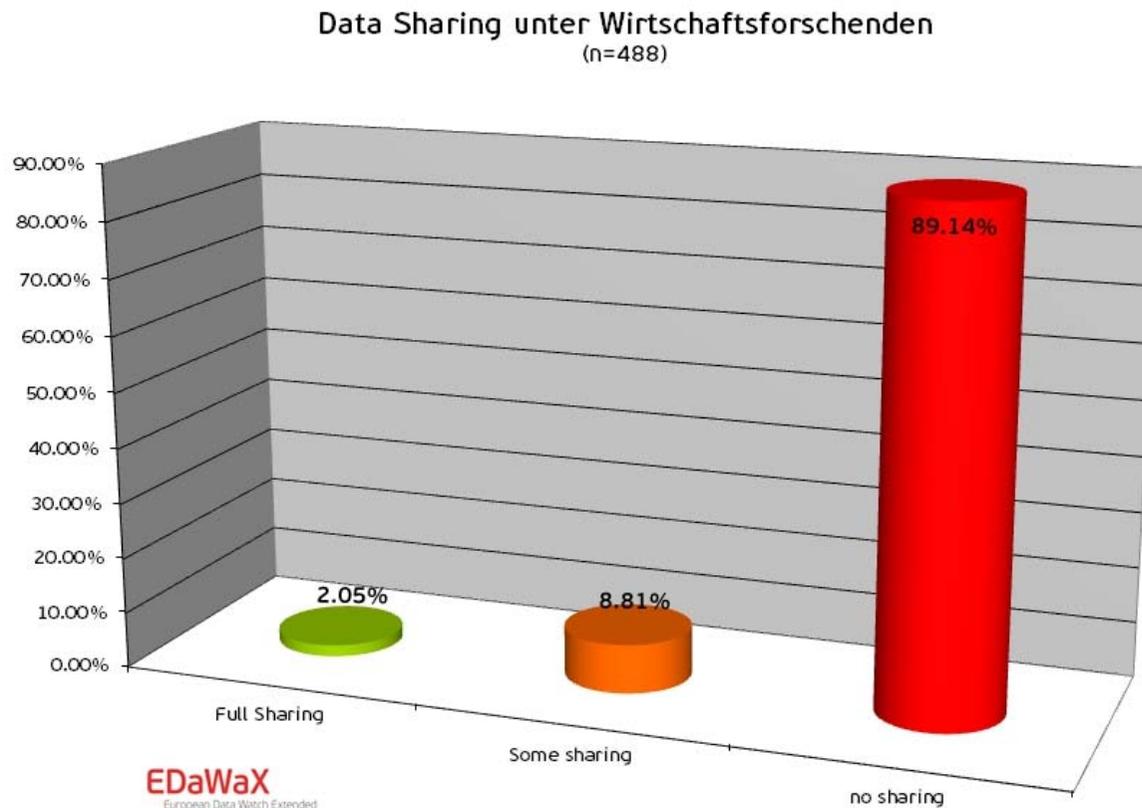
=> Konsequenzen in Hinsicht auf mögliche Bereitstellung d. Daten!

---

# Data Sharing unter Ökonomen

(Dank an Patrick Andreoli-Versbach und Frank Mueller-Langer – MCIER)

# Der Status Quo in der Wirtschaftsforschung besteht darin, Forschungsdaten nicht zu teilen.



Quelle: Andreoli-Versbach/Mueller-Langer (2013): Open Access to Data: An Ideal Professed but not Practised

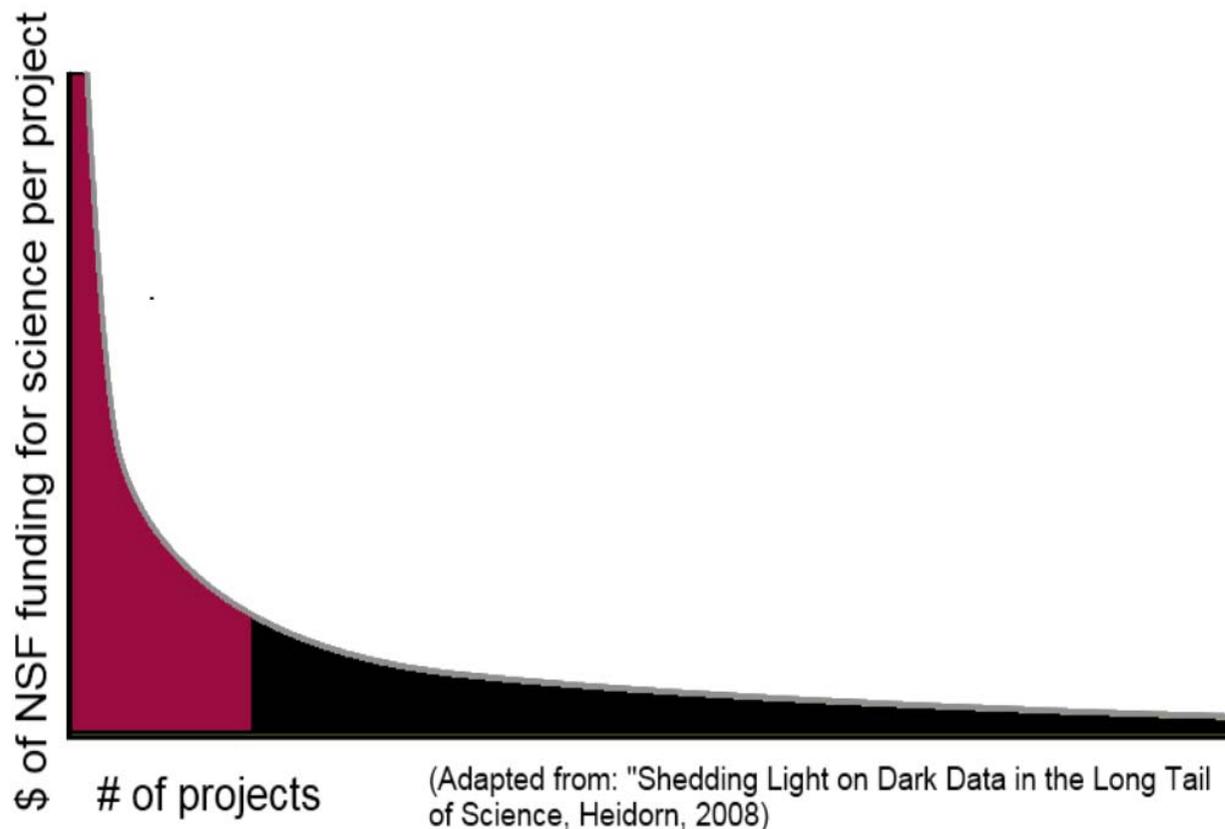
# Forschungsdateninfrastrukturen für die Wirtschaftswissenschaften

# Die deutschen Forschungsdatenzentren (FDZ)

- > ...werden durch den Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditiert. Ziele des RatSWD sind u.a.
  - Verbesserung der Datennutzung und des Datenzugangs
  - Wahrung des Datenschutzes
  - Standardsetzungen (z.B. Dokumentation) und kontinuierliche Evaluation der FDZS
  - Bereitstellung wissenschaftlicher und statistischer Daten (Forschungsdatenportal, Metadaten)
- > Gegenwärtig sind 25 Datenzentren akkreditiert, z.B.
  - Statistisches Bundesamt/Statistische Landesämter, Rentenversicherung, Bundesagentur für Arbeit, ZEW, SOEP, etc.
- => Hochwertige Forschungsdateninfrastruktur und Dokumentation

# Forschungsdateninfrastruktur für kleine und mittlere Projekte

## Dark Data and the Long Tail of Science



# Forschungsdateninfrastrukturen für kleine und mittlere Projekte

- > Zahlreiche Herausforderungen:
  - „stored in desks and drawers“
  - Oft ohne/geringe Dokumentation
  - ABER: erhebliches wissenschaftliches Potenzial!
- > Erster Dienst (Beta-Phase): *datorium* (GESIS)
  - Self-upload und Metadatenerstellung
  - Data Review
  - Rechtevergabe



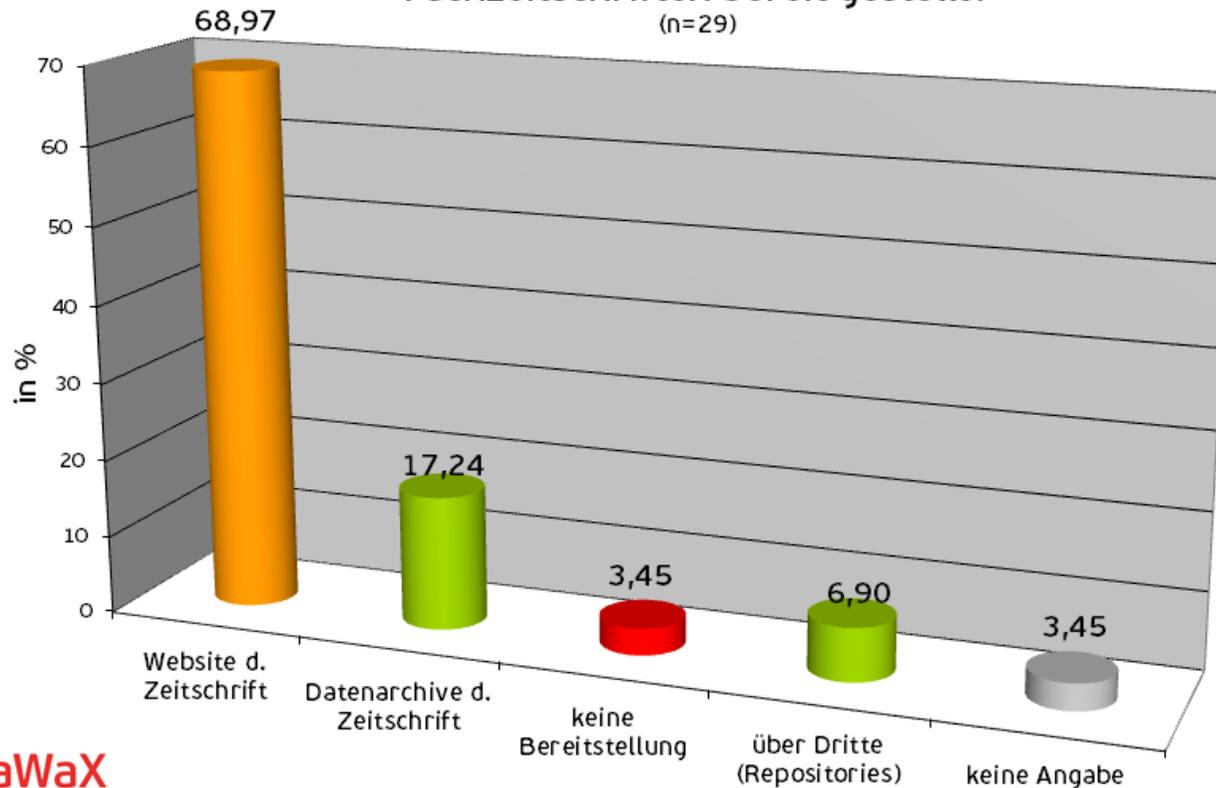
=> Forschungsinfrastrukturen erst in Entwicklung -> „Gap“

# Sonderfall: Infrastrukturen für publikationsbezogene Forschungsdaten (1)

- > keine eigene Kategorie von Forschungsdaten, aber:
  - > besondere Herausforderungen (vor allem Dokumentation), da Replizierbarkeit von Forschungsergebnissen zentrales Ziel ist
  - > Infrastrukturkomponenten dafür kaum verbreitet, und
  - > Lösungen der Fachzeitschriften sind wenig nachhaltig und kaum funktional:

# Bereitstellung von Forschungsdaten

Wie werden Forschungsdaten in wirtschaftswissenschaftlichen Fachzeitschriften bereit gestellt?  
(n=29)



**EDaWaX**  
European Data Watch Extended

> Forschungsdaten werden zumeist nur als ‚Supplementary Material‘ auf den Webseiten der Zeitschriften bereit gestellt.

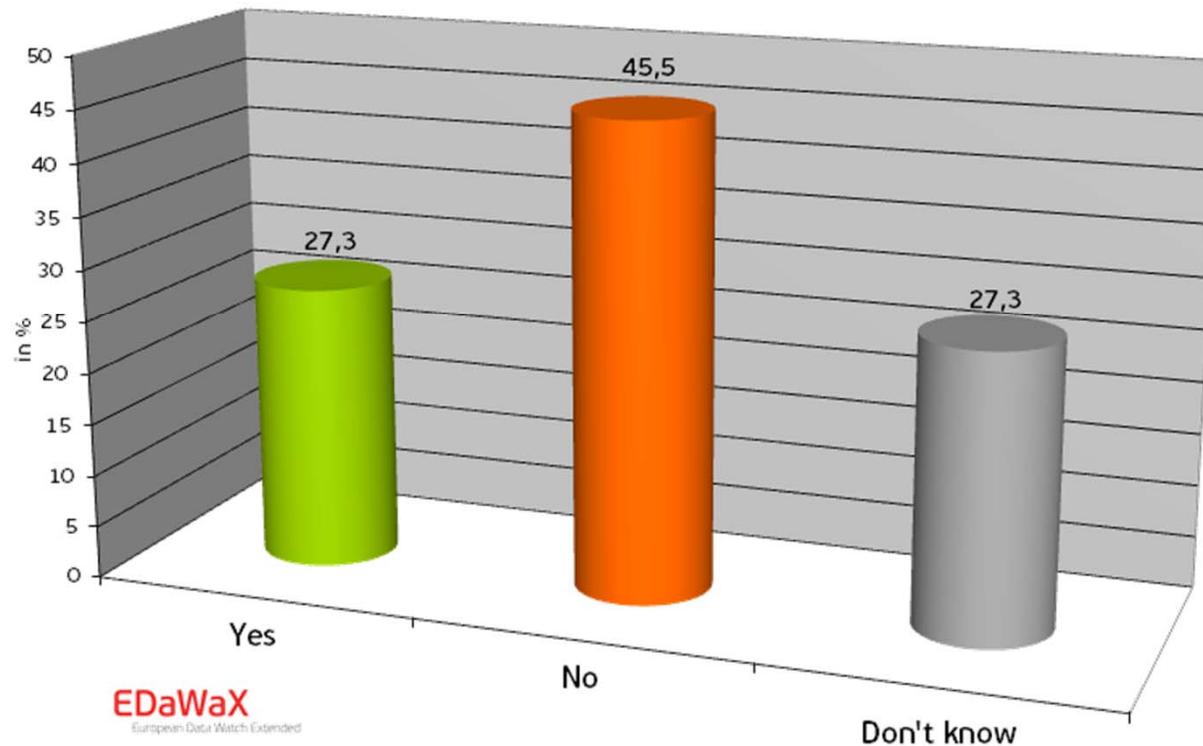
**EDaWaX**  
European Data Watch Extended

RatSWD.

**ZBW** Leibniz-Informationszentrum Wirtschaft  
Leibniz Information Centre for Economics

# Erstellen von Metadaten & PIs durch Journals:

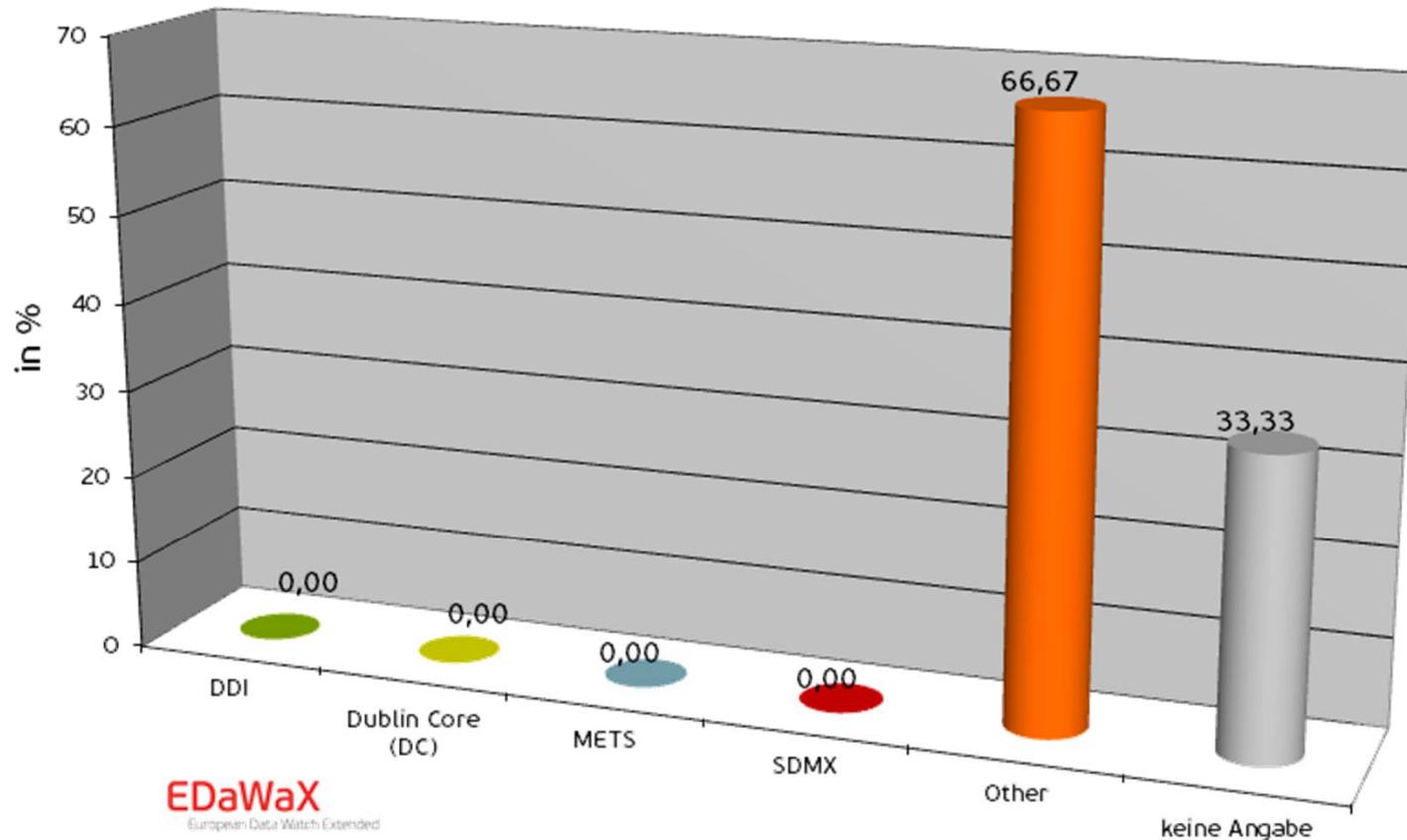
Erstellung von weiteren Metadaten oder von Persistent Identifiern durch  
Fachzeitschriften mit Data Policy  
(n=11)



> Weniger als  $\frac{1}{3}$  der Befragten erstellen weitere Metadaten

# Welche Metadaten­schemata werden genutzt?

Welche Metadaten­schemata werden von Fachzeitschriften mit Data Policies genutzt ?  
(n=3)



➤ ...und wenn doch, werden keine Standards verwendet.

# Sonderfall: Infrastrukturen für publikationsbezogene Forschungsdaten (2)

- > keine eigene Kategorie von Forschungsdaten, aber:
  - > besondere Herausforderungen (vor allem Dokumentation), da Replizierbarkeit von Forschungsergebnissen zentrales Ziel ist
  - > Infrastrukturkomponenten dafür kaum verbreitet, und
  - > Lösungen der Fachzeitschriften sind wenig nachhaltig und kaum funktional.
- > Funktionale Lösungen existieren beispielsweise mit
  - > Dryad (<http://datadryad.org>) für die Lebenswissenschaften
  - > Pangea (<http://www.pangaea.de>) für Erd- und Umweltwissenschaften
  - > Dataverse (<http://thedata.org>) v.a. für die Sozialwissenschaften

## Economics: The Open-Access, Open-Assessment E-Journal Dataverse

## AGE-SPECIFIC RISE OF INCOME AND CONSUMPTION INEQUALITY [DATASET]

[< View Previous Study Listing](#)

hdl:1902.1/20442

Version: 1– Released: Mon Mar 04 03:51:47 EST 2013

## CATALOGING INFORMATION

[Data & Analysis](#)[Comments \(0\)](#)[Versions](#)**i** If you use these data, please add the following citation to your scholarly references. [Why cite?](#)

## Data Citation

Zhu, Guozhong, 2013, "Age-specific Rise of Income and Consumption Inequality [Dataset]",  
<http://hdl.handle.net/1902.1/20442> Economics: The Open-Access, Open-Assessment E-Journal [Distributor] V1  
 [Version]

Citation Format 

## Publications

Guozhong Zhu (2013). Age-specific Rise of Income and Consumption Inequality. Economics Discussion Papers, No 2013-21, Kiel Institute for the World Economy.  
<http://www.economics-ejournal.org/economics/discussionpapers/2013-21/>

## Data Citation Details ▾

Title	Age-specific Rise of Income and Consumption Inequality [Dataset]
Study Global ID	hdl:1902.1/20442
Authors	Zhu, Guozhong (Guanghua School of Management, Peking University, China)
Production Date	2013
Distributor	Economics: The Open-Access, Open-Assessment E-Journal 
Contact	Korinna Werner-Schwarz (IfW), <a href="mailto:korinna.werner-schwarz@economics-ejournal.org">korinna.werner-schwarz@economics-ejournal.org</a>
Distribution Date	2013
Deposit Date	March 04, 2013
Provenance	<a href="#">Economics: The Open-Access, Open-Assessment E-Journal Dataverse</a>

## Description and Scope ▾

## Description

Based on Panel Study of Income Dynamics (PSID) and Consumer Expenditure Survey (CEX), the author presents evidence that the rise of income/consumption inequality over the past decades is more significant among younger households. This is consistent with the theory that the secular rise of inequality is due to increasing heterogeneity in earning ability. The author further shows that such age-specificity implies significant changes to the previously documented life-cycle profiles of inequality which are the basis of many important economic inferences.

## Economics: The Open-Access, Open-Assessment E-Journal Dataverse

### TAKING A DSGE MODEL TO THE DATA MEANINGFULLY [DATASET]

[< View Previous Study Listing](#)

hdl:1902.1/13661

Version: 1 – Released: Thu Nov 26 05:21:18 EST 2009

[Cataloging Information](#)

**DATA & ANALYSIS**

[Comments \(0\)](#)

[Versions](#)

**i** Use the check boxes next to the file name to download multiple files. Data files will be downloaded in their default format. You can also download all the files in a category by checking the box next to the category name. You will be prompted to save a single archive file. Study files that have restricted access will not be downloaded.

Select all files

Total Downloads: **282**

**2. Data**

**irelanddata.xls**

MS Excel - 175 KB - 108 downloads

 [Download](#)

data in xls file: quarterly

**irelandoutput.txt**

Plain Text - 22 KB - 90 downloads

 [Download](#)

output file in txt format showing the computer output

**ireland.prg**

Plain Text - 1 KB - 84 downloads

 [Download](#)

program file used in CATS in RATS

# Sonderfall: Infrastrukturen für publikationsbezogene Forschungsdaten (3)

- > keine eigene Kategorie von Forschungsdaten, aber:
  - > besondere Herausforderungen (vor allem Dokumentation), da Replizierbarkeit von Forschungsergebnissen zentrales Ziel ist
  - > Infrastrukturkomponenten dafür kaum verbreitet, und
  - > Lösungen der Fachzeitschriften sind wenig nachhaltig und kaum funktional.
- > Funktionale Lösungen existieren beispielsweise mit
  - > Dryad (<http://datadryad.org>) für die Lebenswissenschaften
  - > Pangea (<http://www.pangaea.de>) für Erd- und Umweltwissenschaften
  - > Dataverse (<http://thedata.org>) v.a. für die Sozialwissenschaften
- > Darüber hinaus entwickelt das DFG-Projekt „*European Data Watch Extended*“ ([www.edawax.de](http://www.edawax.de)) ein generisch nachnutzbares Forschungsdatenarchiv für wirtschaftswissenschaftliche Journals.

## Data Management in Economics Journals (Replication Data)

Followers

1

Unfollow

Social

Google+

Twitter

Facebook

Dataset

Activity Stream

Related

Edit

## Data Management in Economics Journals (Replication Data)

In our paper "Data Management in Scholarly Journals and Possible Roles for Libraries – Some Insights from EDaWaX!" for LIBER Quarterly (22),1 we summarize the findings of an empirical study conducted by the EDaWaX-Project. 141 economics journals were examined regarding the quality and extent of data availability policies that should support replications of published empirical results in economics.

The paper is available at URN:NBN:NL:UI:10-1-114595

### Data and Resources

- |   |  |                         |
|---|--|-------------------------|
|    | <b>Full dataset 154 journals</b><br>This is the population ("universe") of our sample. Of the 154 journals...                                  | <a href="#">Explore</a> |
|    | <b>Dataset of all journals with data policies</b><br>This is a more detailed dataset we created for examining the 29 data...                   | <a href="#">Explore</a> |
|   | <b>Data Dictionary for the full sample</b><br>In this data dictionary I described every variable and value of the full...                      | <a href="#">Explore</a> |
|  | <b>Data Dictionary of the dataset for journals with ...</b><br>This data dictionary describes every variable and value of the small dataset... | <a href="#">Explore</a> |
|  | <b>Syntax of the detailed analyses (40 cases)</b><br>The code of computation to generate some of the results claimed in the LIBER...           | <a href="#">Explore</a> |
|  | <b>Syntax 154 full dataset</b><br>Syntax of the analyses of the 154 economics journals in our sample   | <a href="#">Explore</a> |

# Vielen Dank für Ihr Interesse!

Gibt es Anregungen, Fragen oder Kommentare ?

Kontakt:

Sven Vlaeminck | [s.vlaeminck@zbw.eu](mailto:s.vlaeminck@zbw.eu)  
ZBW – Leibniz-Informationszentrum Wirtschaft  
Neuer Jungfernstieg 21  
20354 Hamburg

Alle Inhalte dieser  
Präsentation sind mit  
der [CC BY-NC-SA 3.0](https://creativecommons.org/licenses/by-nc-sa/3.0/de/)  
DE Lizenz versehen.

