The Production and Dissemination of Microdata at Bundesbank

Stefan Bender, Head of Research Data and Service Center (RDSC), Deutsche Bundesbank

Third Statistic Conference
Measuring the Economy in the Digital Age
October 1-2, 2019
Santiago, Chile

The views expressed here do not necessarily reflect the opinion of the Deutsche Bundesbank, the INEXDA network, or the Eurosystem

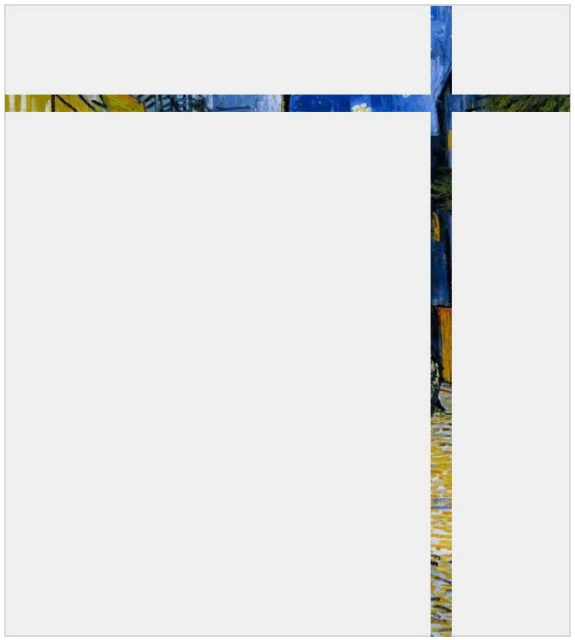
Data in Dutch Painting (1|2)

Credit: Ralph Klüber, p3 Insights

Café Terrace at Night is an 1888 oil painting by the Dutch artist Vincent van Gogh. Van Gogh painted *Café Terrace at Night* in Arles, France.

(a) **Original Painting**

(b) **Bundesbank Data**



(c) **Big Data**

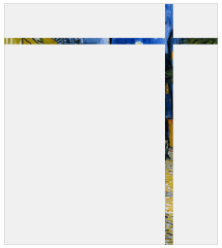
Data in Dutch Painting (2|2)

Credit: Ralph Klüber, p3 Insights

Café Terrace at Night is an 1888 oil painting by the Dutch artist Vincent van Gogh.
Van Gogh painted *Café Terrace at Night* in Arles, France.

(a) **Original Painting**

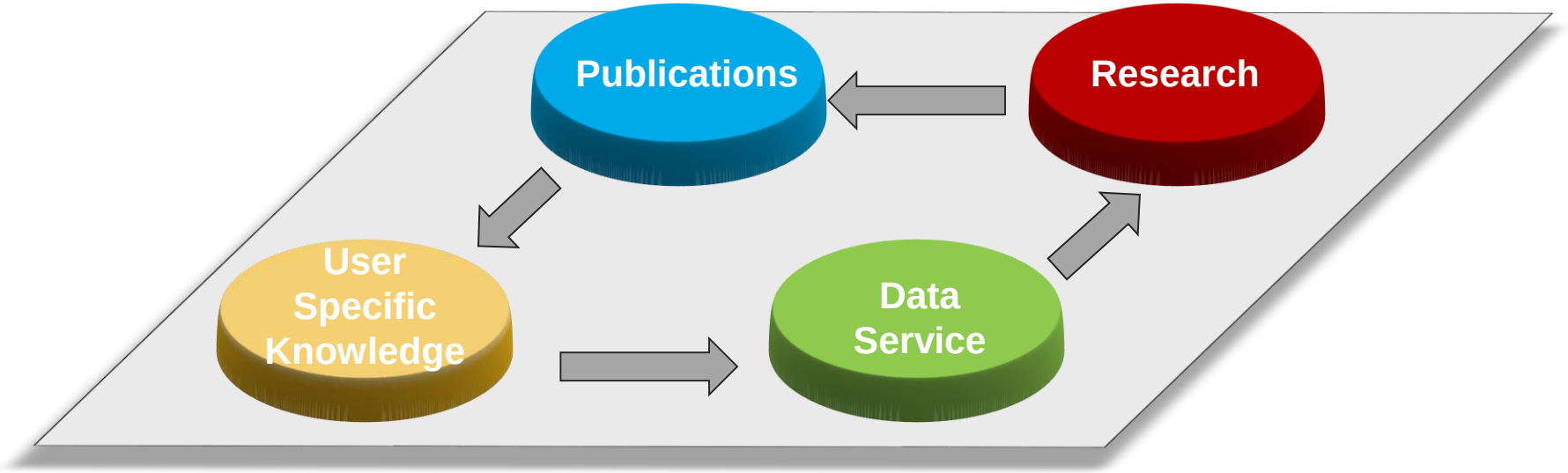
(b) BBk Data



(c) Big data

(d) **Designed Product**

Knowledge Life Cycle

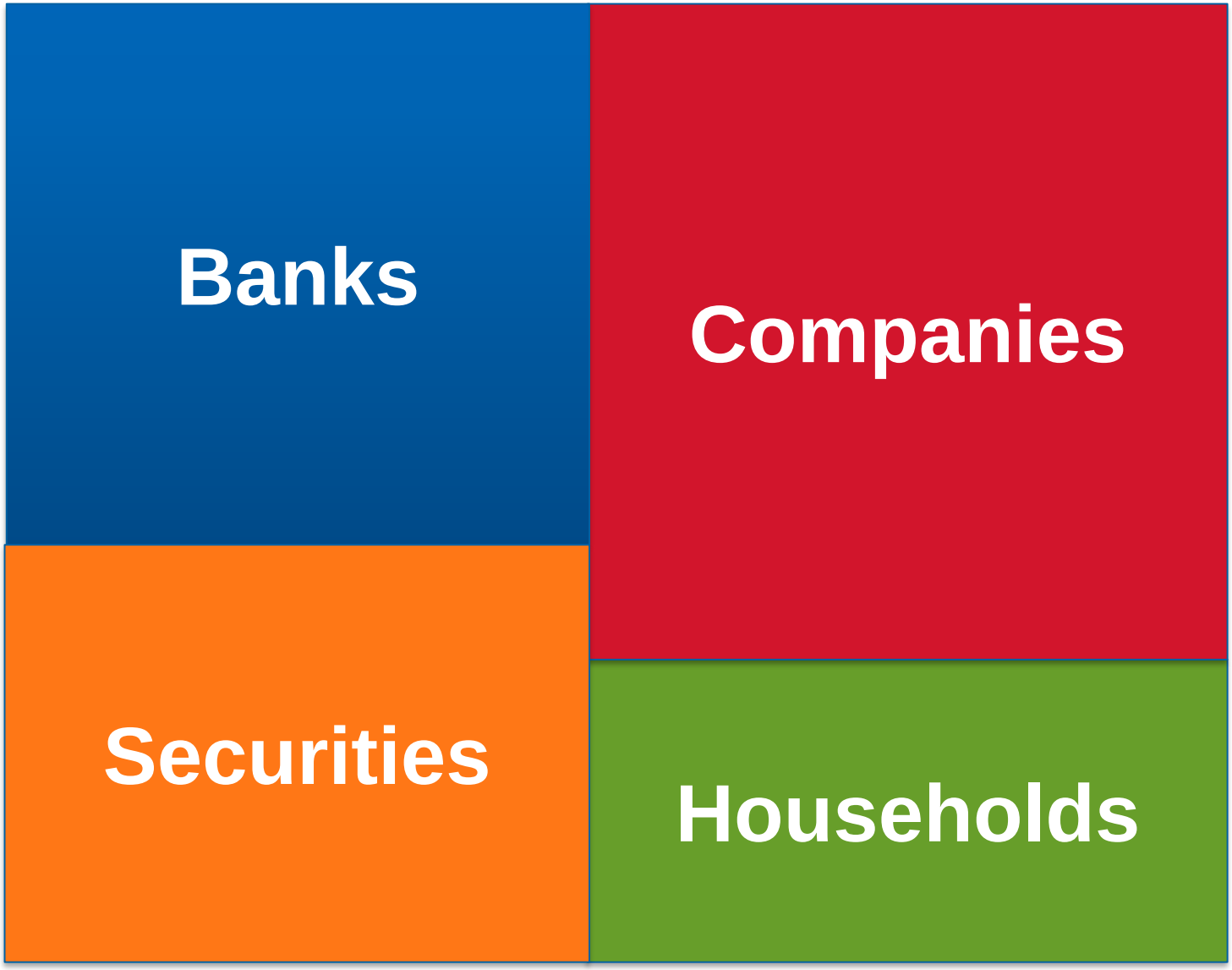


Overview

- (Motivation)
- The Research Data and Service Center of Bundesbank (RDSC)
- Future steps (one example):

  Enhancing data access: Automated tools for users and producers (by Hendrik Doll and Christian Resch)
- (INEXDA: The International Granular Data Network)
- Conclusion

Available microdata at the RDSC



Motivation for establishing the RDSC: IMIDIAS

- The RDSC is part of the internal project **Integrated MicroData-based Information and Analysis System (IMIDIAS)**
- Granular data become more and more important for assessing monetary and regulatory policy as well as for financial stability. Hence the Bundesbank has launched the large-scale initiative IMIDIAS aimed at making better use of existing data both, for policy analysis as well as internal and external researchers.
- **Goals of IMIDIAS:**
  - Support policymaking process
  - Encourage cooperation with (external) researchers
  - Promote evidence-based policy-making

Tasks of the RDSC

The RDSC offers access for non-commercial research to (highly sensitive) micro data of the Bundesbank:

- Generate (linked) micro data
- Offer advisory service on data selection and data access (data handling, research potential, scope and validity of data)
- Provide data access and data protection
- Document data and methodological aspects of the data
- Work on own research projects (in close cooperation with the Bank's business areas and the Research Centre)
- Organize conferences and workshops

Factsheet on the RDSC

- 20 employees
- 12 working places for guest researchers in Frankfurt
- 2 working places in Düsseldorf
- In 2018:
  - Around 130 project applications, 73 were realized
  - Around 280 output controls
  - Average of used data products per research project: 2.68
  - Papers of RDSC users are out
- In 2017: over 300 active projects, over 160 institutions involved (around 90 non-German)

Data Generating Process

Until now (in many cases): ad hoc generation of data for research.

RDSC has started to/with:

- Establishing standardised data products.
- Implementing RDSC data quality procedures.
- Documentation of data.
- Harmonisation of data.
- Register data to get data identifiers (DOIs).

The 5 Safes in the RDSC

- **Safe people**: non-disclosure agreement, contract (with penalty up to 60,000 Euro, publishing the name, exclusion from access up to 2 years).
- **Safe projects**: non-commercial research, project description.
- **Safe environment**: working places without internet connection, (cell) phone, photo, printer and drive.
- **Safe data**: (weakly) anonymized data.
- **Safe results**: output control, papers/presentations are checked.

- **Access to real data**, anonymization is only one dimension, others have more effects on data protection.

Summing Up: Bundesbank

- Accessing data through a RDSC/RDC is the best way for using highly sensitive data like Bundesbank's data.
- The Bundesbank hopes to (further) stimulate international research and policy analysis with these data.
- Accessing the data is comparatively easy and inexpensive.

➢ Come, visit us in Frankfurt!

DEUTSCHE BUNDESBANK EUROSYSTEM

Enhancing data access:

Automated tools for users and producers

Hendrik Doll and Christian Resch (RDSC)

Based on a project with and contributions from:

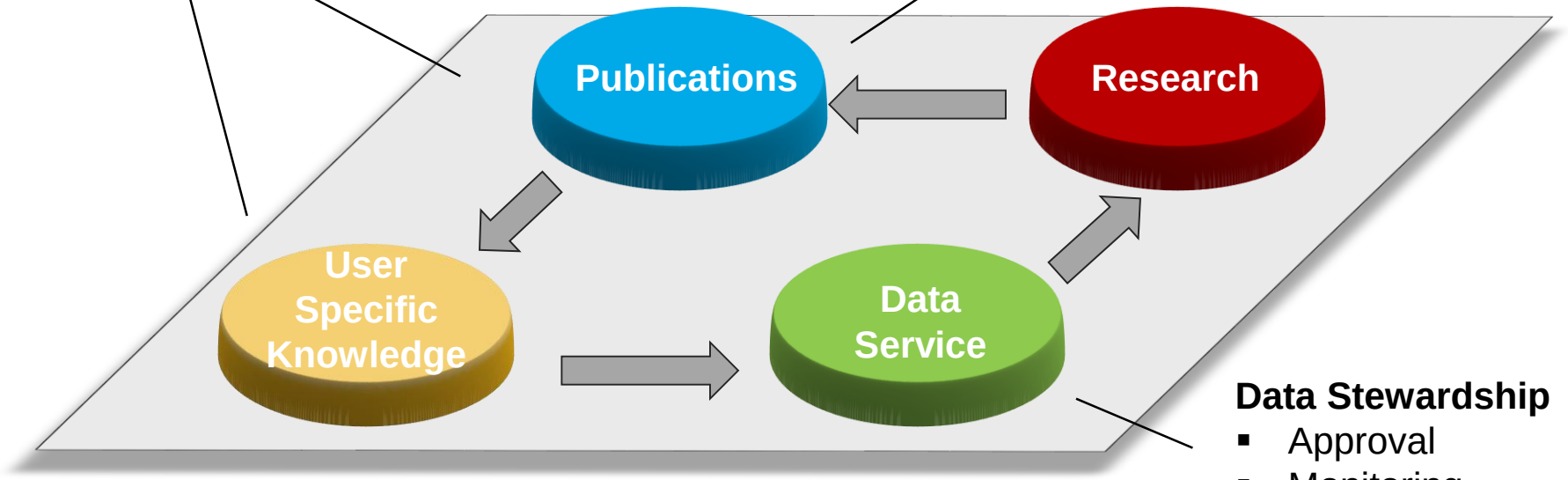
Stefan Bender, Christian Hirsch, Julia-Katharina Ginz, John Chase, Jonathan Morgan, Ian Mulvaney, Andrew Gordon and Julia Lane

Knowledge Life Cycle (with tools)

Collaboration
- Knowledge sharing
- Metadata

Secure workspace
- Services and Tools

Rich Context



Data Stewardship
- Approval
- Monitoring
- Reporting

Motivation

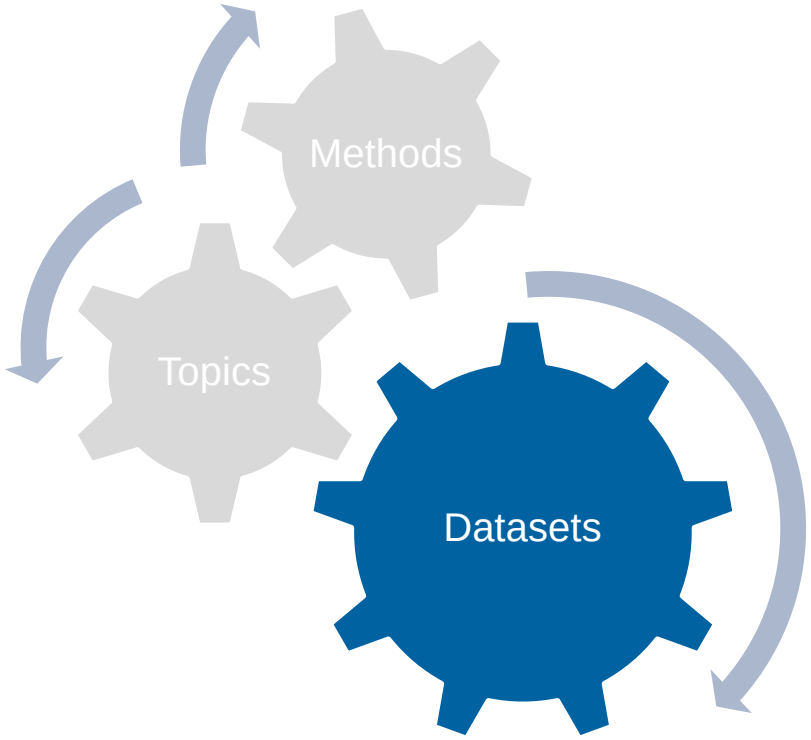
1. Enhance data services for users

- Build Amazon-style recommendation system
- *What have others done with the data?*

2. Justify added value of microdata for research

- Dataset impact factor

How does articles look like?



3 Data Description

The data sources used in our study are (i) Auxmoney for data on P2P lending; (ii) the Deutsche Bundesbank (Interest Rates Statistics) for data on bank lending; (iii) Schufa for data on credit ratings; (iv) the Deutsche Bundesbank (Balance Sheet Statistics) for data on loan loss provisions.

Auxmoney is the oldest and largest P2P lending platform in Germany. According to its website, from the day it began business in 2007 until late 2015, the total volume of credit provided was €219 million in 39,090 projects, with an average nominal interest rate of 9.65%.

Auxmoney provided us with two different datasets. The first includes all loans divided by state between January 2010 and September 2014, with no maturity information. The second includes the average interest rate and the average credit rating represented by the Schufa score for each state per month.[19] [20]

The Deutsche Bundesbank statistics used in this study are provided by two different datasets. The first is the Interest Rates Statistics (MIR, see Bade and Beier (2016) for further information on this data source), which is a stratified sample of the German banking sector used for supervisory activities and gives the amounts and the interest rates per bank and per month applied to noncon-struction consumer credit lines (outstanding and new business) for different maturities (overdraft, up to one year, and more than one year).[21] The statistics are composed of monthly observations between January 2010 and September 2014. The second is the dataset from the Balance Sheet Statistics (BISTA, see Beier, Krueger, and Schaefer (2016) for further information on this data source), which gives information on write-ups and write-downs, from which we derive the banks' loan loss provisions.

Our analysis is at the bank-state level. The regional differentiation of bank loans is possible because of a feature of the German banking system: the presence of Sparkassen (savings banks) and Volksbanken (cooperative banks). Each bank is only present in one German state. Sparkassen are geographically restricted banks with a legal mandate to provide bank services to all creditworthy

[19] Schufa is a German private credit bureau with 479 million records on 66.2 million natural persons. Schufa provides credit ratings for each person requesting a loan and Auxmoney provides the Schufa score of each credit application.
[20] For reasons of data confidentiality, Auxmoney provides its credit intermediation by month and state only if five or more loans were made in that month in that state.
[21] The Interest Rates Statistics (MIR) is the German part of a larger dataset that is used by the ECB for regulatory purposes. It does not cover the whole German banking sector, only a stratified sample. For this reason, our sample does not cover all Sparkassen and Volksbanken in Germany, just the ones present in this data source.

Model evaluation with central bank sample

Two algorithms to find data sets in publications:

- Allen AI
- KAIST

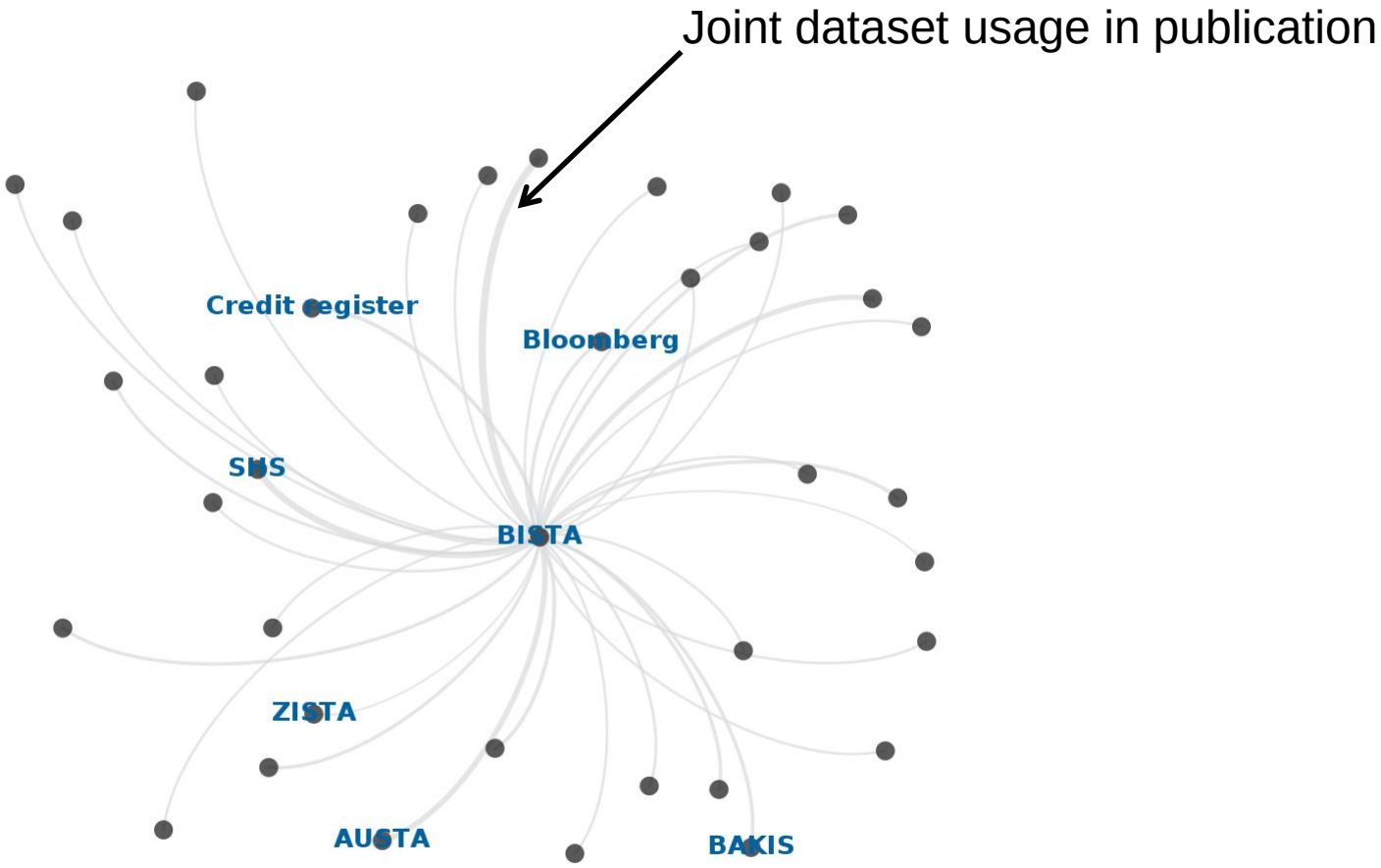
Experiences:

- Allen AI only finds what it is trained on
- KAIST also finds unknown datasets
- Yet, KAIST is very noisy

Model evaluation: Example

Title	KAIST	Allen AI
The PHF: A comprehensive panel survey on household finances and wealth in Germany	PHF	PHF, Panel on Household Finances
Saving and learning: Theory and evidence from saving for child's college	PSID s family data, PSID, Panel Study of Income Dynamics, Data, rst surveyed in the spring of, CDS TA	Panel Study of Income Dynamics (PSID)
Monetary policy and the oil futures market	commodity prices, VAR, JEL Classication	- / -
Is the willingness to take financial risk a sex-linked trait? Evidence from national surveys of household finance	DNB Household Survey, BIS Papers	- / -
China's role in global inflation dynamics	WEO, UN Comtrade database, Journal of Econometrics, OECD s Main Economic Indicators	- / -

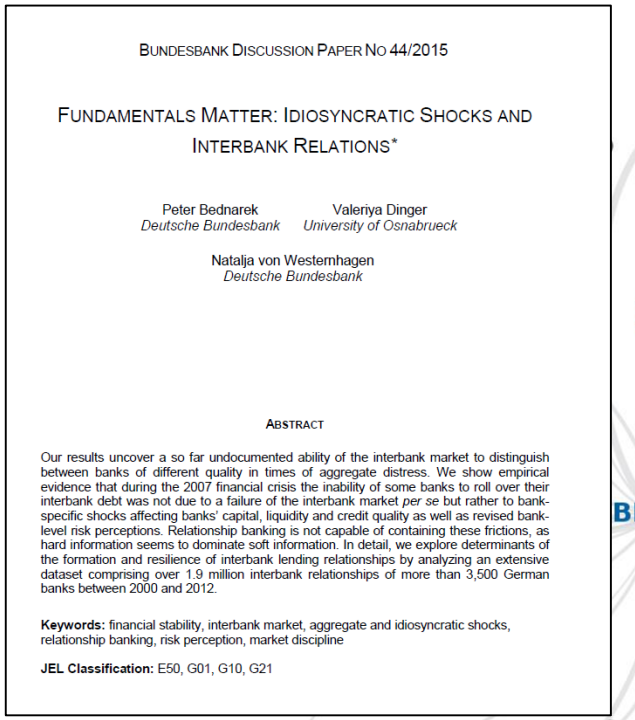
From competition results to dataset network



BISTA ego-network

From competition results to dataset network

BISTA and credit register:



BUNDESBANK DISCUSSION PAPER NO 44/2015

FUNDAMENTALS MATTER: IDIOSYNCRATIC SHOCKS AND INTERBANK RELATIONS*

Peter Bednarek, *Deutsche Bundesbank*
Valeriya Dinger, *University of Osnabrueck*
Natalja von Westernhagen, *Deutsche Bundesbank*

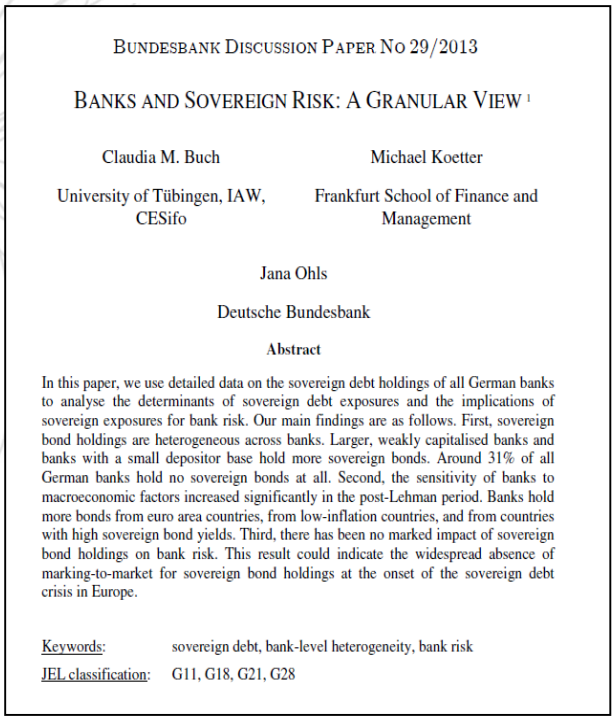
ABSTRACT

Our results uncover a so far undocumented ability of the interbank market to distinguish between banks of different quality in times of aggregate distress. We show empirical evidence that during the 2007 financial crisis the inability of some banks to roll over their interbank debt was not due to a failure of the interbank market *per se* but rather to bank-specific shocks affecting banks' capital, liquidity and credit quality as well as revised bank-level risk perceptions. Relationship banking is not capable of containing these frictions, as hard information seems to dominate soft information. In detail, we explore determinants of the formation and resilience of interbank lending relationships by analyzing an extensive dataset comprising over 1.9 million interbank relationships of more than 3,500 German banks between 2000 and 2012.

Keywords: financial stability, interbank market, aggregate and idiosyncratic shocks, relationship banking, risk perception, market discipline

JEL Classification: E50, G01, G10, G21

BISTA and SHS:



BUNDESBANK DISCUSSION PAPER NO 29/2013

BANKS AND SOVEREIGN RISK: A GRANULAR VIEW [1]

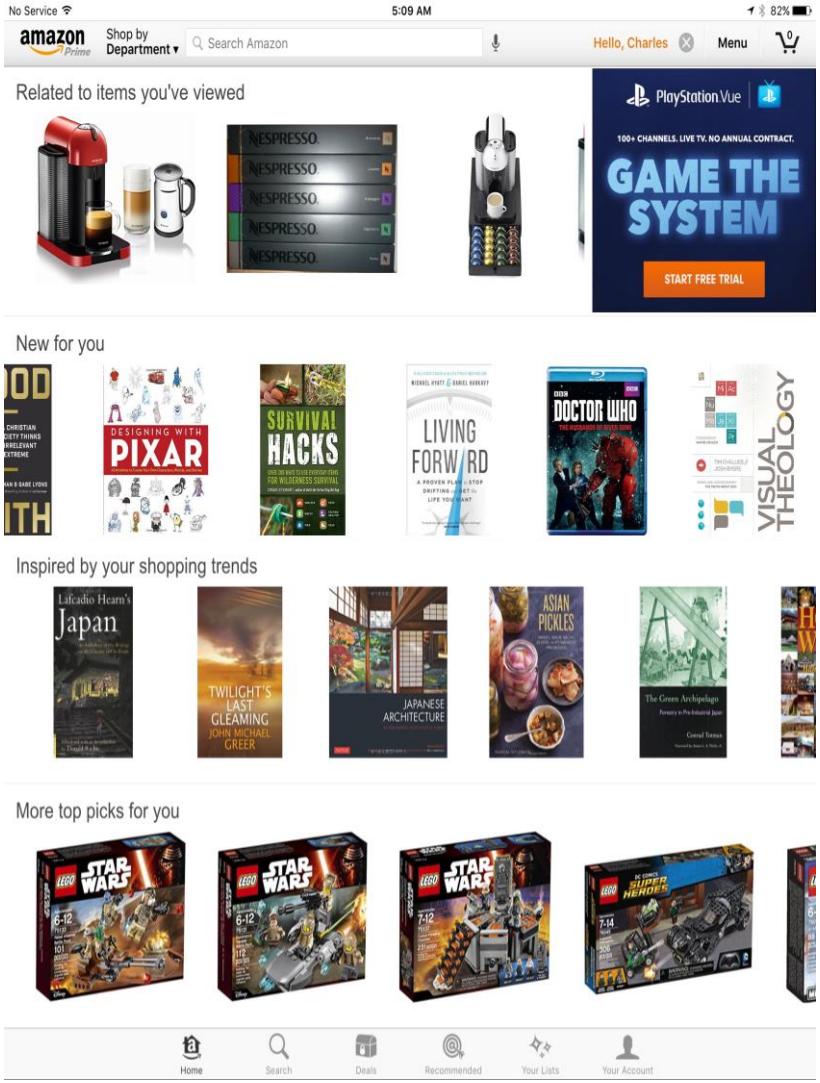
Claudia M. Buch, University of Tübingen, IAW, CESifo
Michael Koetter, Frankfurt School of Finance and Management
Jana Ohls, Deutsche Bundesbank

Abstract

In this paper, we use detailed data on the sovereign debt holdings of all German banks to analyse the determinants of sovereign debt exposures and the implications of sovereign exposures for bank risk. Our main findings are as follows. First, sovereign bond holdings are heterogeneous across banks. Larger, weakly capitalised banks and banks with a small depositor base hold more sovereign bonds. Around 31% of all German banks hold no sovereign bonds at all. Second, the sensitivity of banks to macroeconomic factors increased significantly in the post-Lehman period. Banks hold more bonds from euro area countries, from low-inflation countries, and from countries with high sovereign bond yields. Third, there has been no marked impact of sovereign bond holdings on bank risk. This result could indicate the widespread absence of marking-to-market for sovereign bond holdings at the onset of the sovereign debt crisis in Europe.

Keywords: sovereign debt, bank-level heterogeneity, bank risk
JEL classification: G11, G18, G21, G28

From dataset network to recommendations

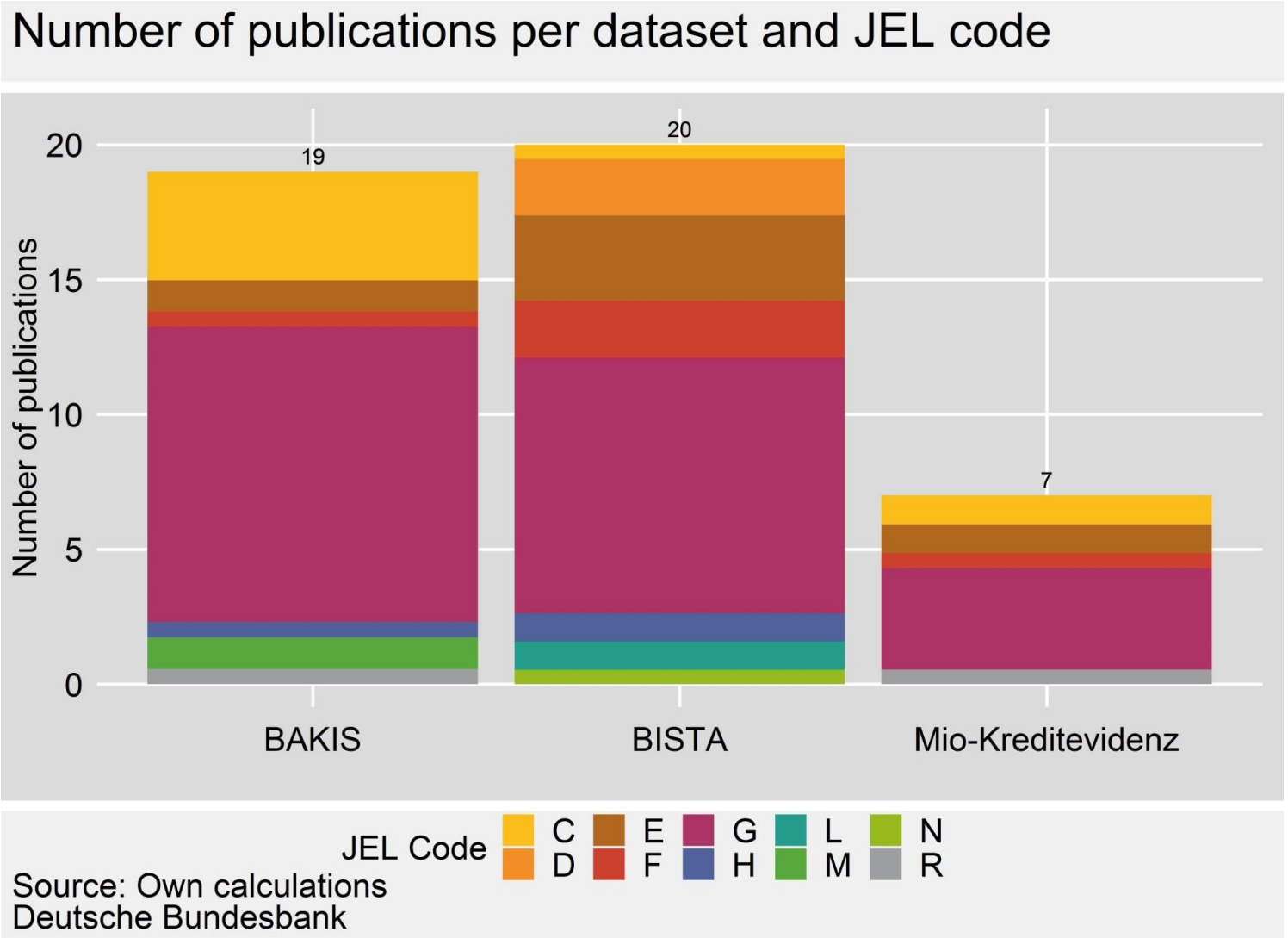


Related to data you've viewed:

1. Jointly used datasets
2. Publications
3. Similar data
4. …

Thanks to Julia Lane

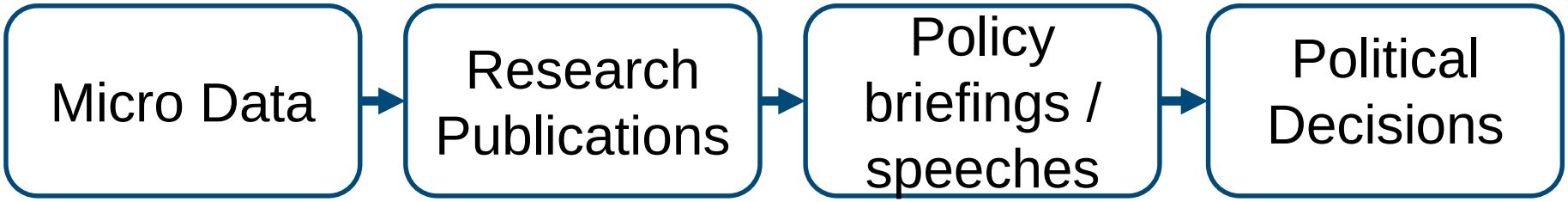
From dataset network to dataset impact factor



Dataset impact factor: Our axioms

- ✓ More publications using dataset are better
- ✓ Broader usage is better
- More links to other datasets are better
- More recent usage is better
- Higher impact factor of publications is better
- Dataset is better when it sparks new research

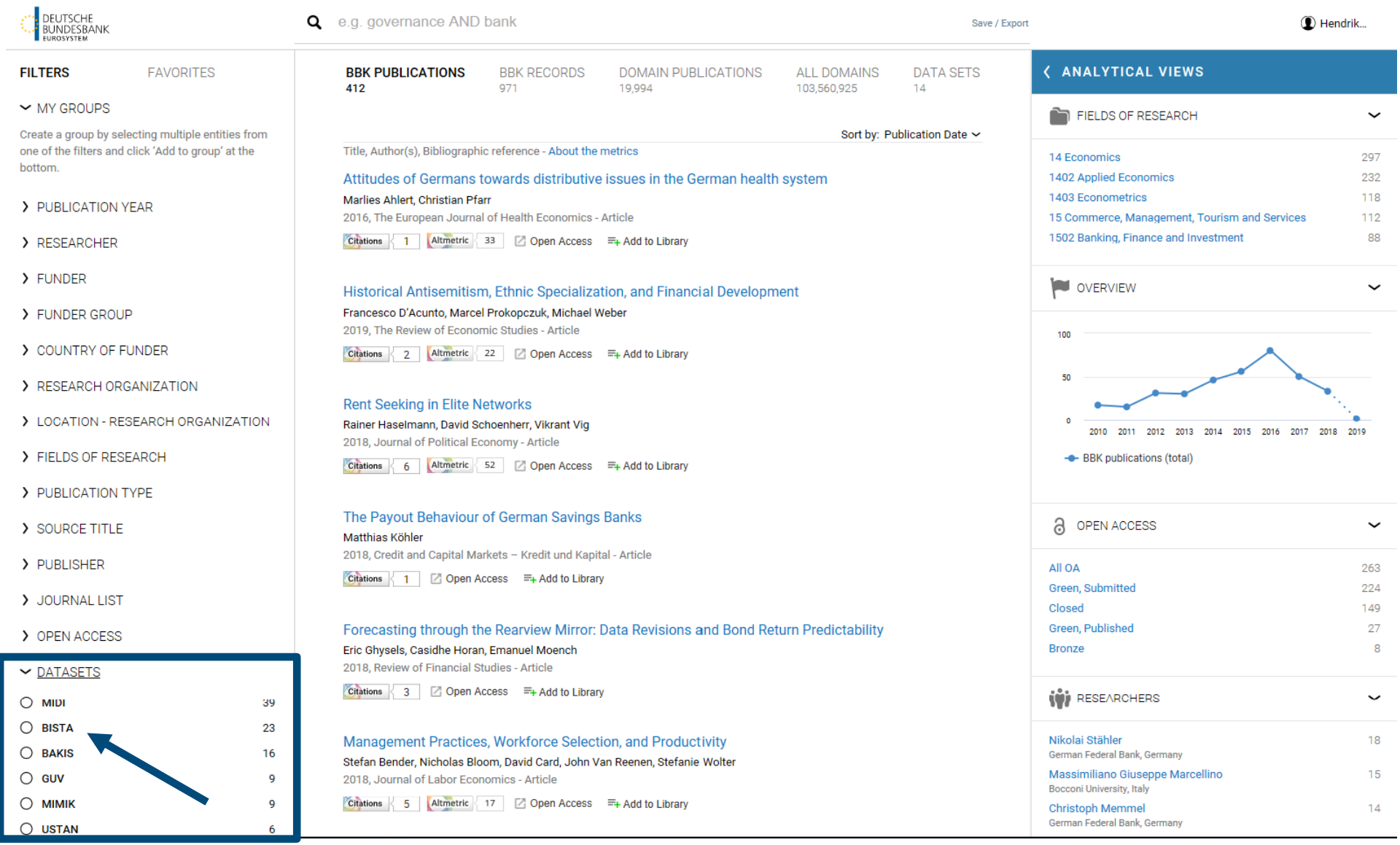
Evidence-based policy: information flow



New possibilities:

- Measure dataset impact on political decisions
- Optimizally allocate resources to improve societally important data

User interface to explore micro data usage



Find publications using your dataset

DEUTSCHE BUNDESBANK EUROSYSTEM

BISTA OR SHS-BASE
Datasets

Hendrik...

BBK Publication - Article

Cyclicality of SME lending and government involvement in banks

Journal of Banking & Finance, 77, 64-77, April 2017
https://doi.org/10.1016/j.jbankfin.2017.01.010

Authors

Patrick Behr - Fundação Getulio Vargas
Daniel Foos - German Federal Bank
Lars Norden - Fundação Getulio Vargas
Corresponding Author

Abstract

Recent regulatory efforts aim at lowering the cyclicality of bank lending because of its potentially detrimental effects on financial stability and the real economy. We investigate the cyclicality of SME lending of local banks with versus without a public mandate, controlling for location, size, loan maturity, capitalization, funding structure, liquidity, profitability, and credit demand-side factors. The public mandate is set by local governments and stipulates a sustainable provision of financial services to local customers and a deviation from strict profit maximization. We find that banks with a public mandate are 25% less cyclical than other local banks. The result is credit supply-side driven and especially strong for public mandate banks with high liquidity and stable deposit funding. Our findings have implications for the bank structure, financial stability and the finance-growth nexus in a local context.

less

Dataset references - 2

Sorted by: Date

Gewinn- und Verlustrechnung der Banken (GUV)

Monatliche Bilanzstatistik (BISTA)

Publication references - 25

Sorted by: Date

Social Capital and the Viability of Stakeholder-Oriented Firms: Evidence from Savings Banks*
Charlotte Ostergaard, Ibolya Schindele, Bent Vale
2016, Review of Finance - Article
Citations 6 Altmetric 2 Open Access Add to Library

Small Banks and Local Economic Development*

Open Access

Add to Library

Export citation

Publication metrics

Dimensions Badge

15 Total citations
11 Recent citations
14 Field Citation Ratio
n/a Relative Citation Ratio

Altmetric

Twitter (1)
Mendeley (83)

Research Categories

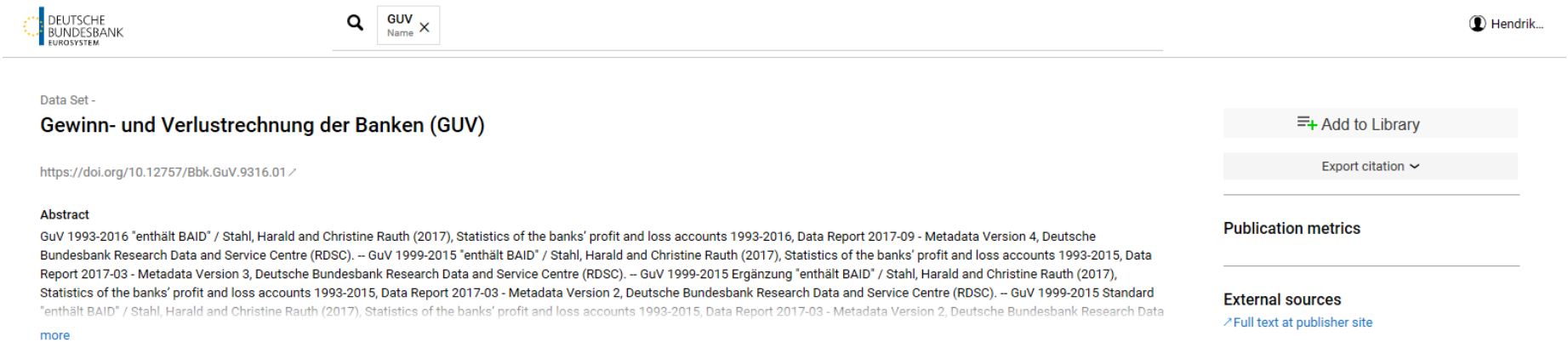
Fields of Research

14 Economics
15 Commerce, Management, Tourism and Services
1402 Applied Economics
1403 Econometrics
1502 Banking, Finance and Investment

External sources

Full text at publisher site

Get dataset information and inspiration



What we did

- Close the gap between publications and data services
- Build the first recommendation system, dataset impact factor and user interface

What‘s next

- Formalise into integrated system around data-services
- Rinse and repeat

INEXDA

The Granular Data Network

The views expressed here do not necessarily reflect the opinion of the Deutsche Bundesbank, the INEXDA network, or the Eurosystem.

INEXDA: in a nutshell

- INEXDA provides a platform for exchanging experiences on statistical handling of granular data for central banks, national statistical institutes and international organizations.
- INEXDA supports the G20 process, especially the Data Gaps Initiative 2 recommendation aiming to promote the exchange of (granular) data as well as metadata.
- Focus of the INEXDA Data Access Working Group is to define common access procedures and use these for the development of an open software solution.

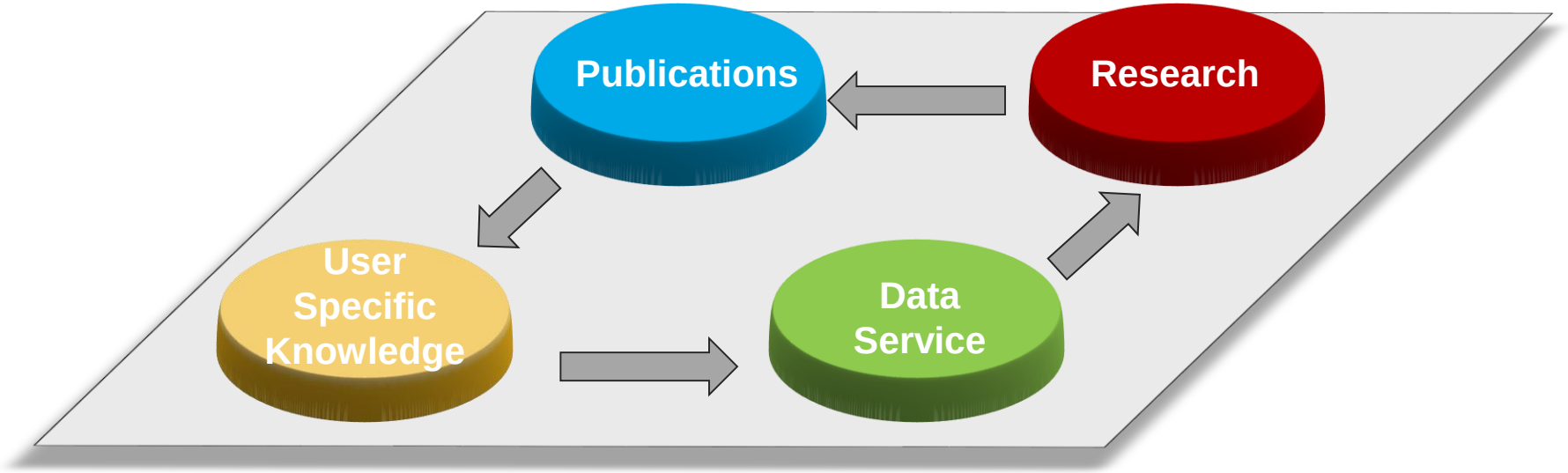
Next steps

- **End of August 2019**
  - Collection of feedback on proposed new metadata items
- **October 2019**
  - Meeting of the INEXDA Data Access Working Group
- **End of 2019**
  - Proof of concept for open source data stewardship module software in collaboration with NYU's ADRF

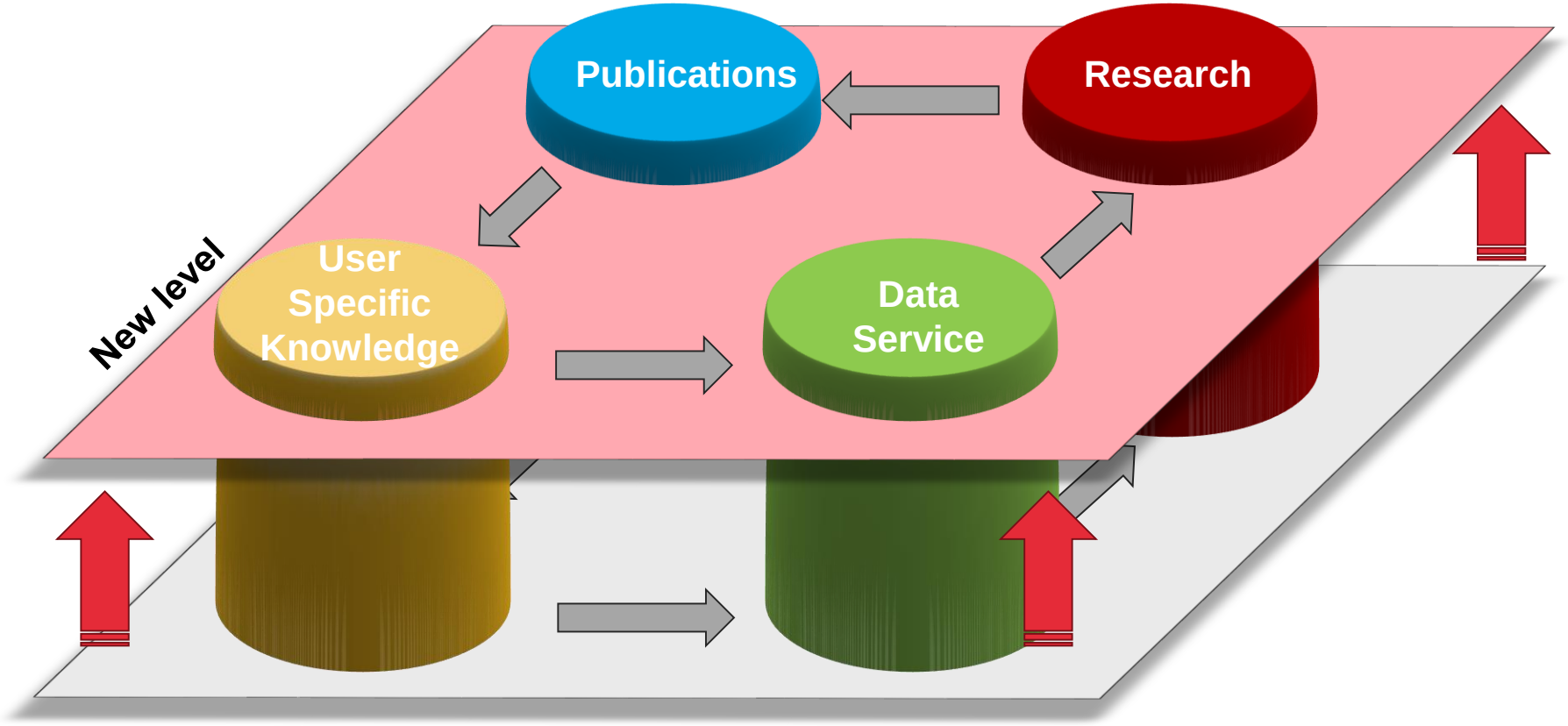
Summing Up I: General

- **Developments** for RDSC(s) and INEXDA was/are fast, but **incremental**: trust building, growing data complexity, learning process …
- (New) **skills** for researchers / data producers.
- **Engagement** of researchers (value of data work, data impact factor).
- **Efficiency**: access system in a RDC, metadata/recommendation system project management in a RDC, …
- **Harmonization/Internationalization**: G20 initiative on data sharing and data access of central banks, INEXDA.

Summing Up II: Getting the Knowledge Life Cycle …



Summing Up III: … into the next level



Thank you !

- **Website**: www.bundesbank.de\fdsz
- **Contact**: fdsz@bundesbank.de

Motivation

- **Aggregate datasets** are important for **monitoring macroeconomic developments** and **macroeconomic policy**
- **Granular data** is necessary to understand **global developments** and in particular **differences across countries**
- Combining datasets and looking beyond aggregate statistics into heterogeneous developments require the **transformation** of **“data“** into **“knowledge“**
- **Local constraints** make it difficult, or often impossible, to link micro datasets from different jurisdictions, even for research and financial stability analysis
- **Better accessibility** and **sharing of granular data** would open up **new possibilities** for analysis by providing new **insights into the effect of policies**

What can **we do** from the **statistical side** to support this process?

What's new in (central bank) statistics?

- Micro data overhaul the traditional value-added chain in central banking statistics.
  - Traditional central banking statistics are collected for a **specific purpose**.
  - Micro data are collected only once and can be used for **multiple purposes**: The statistical reporting burden declines.
  - **Data protection** becomes more challenging.
- Technological innovations have revolutionized the infrastructure for collecting, storing, and using micro-data.
  - Advanced knowledge in storage and organization of large (integrated) micro-data.
  - Improved tools for analyzing and processing microdata.
  - Cheaper storage technologies.
  - Standardization.
- Official statistics has lost the monopoly in providing information to society.

How do we do this?

Step 1

Extract used datasets from publications

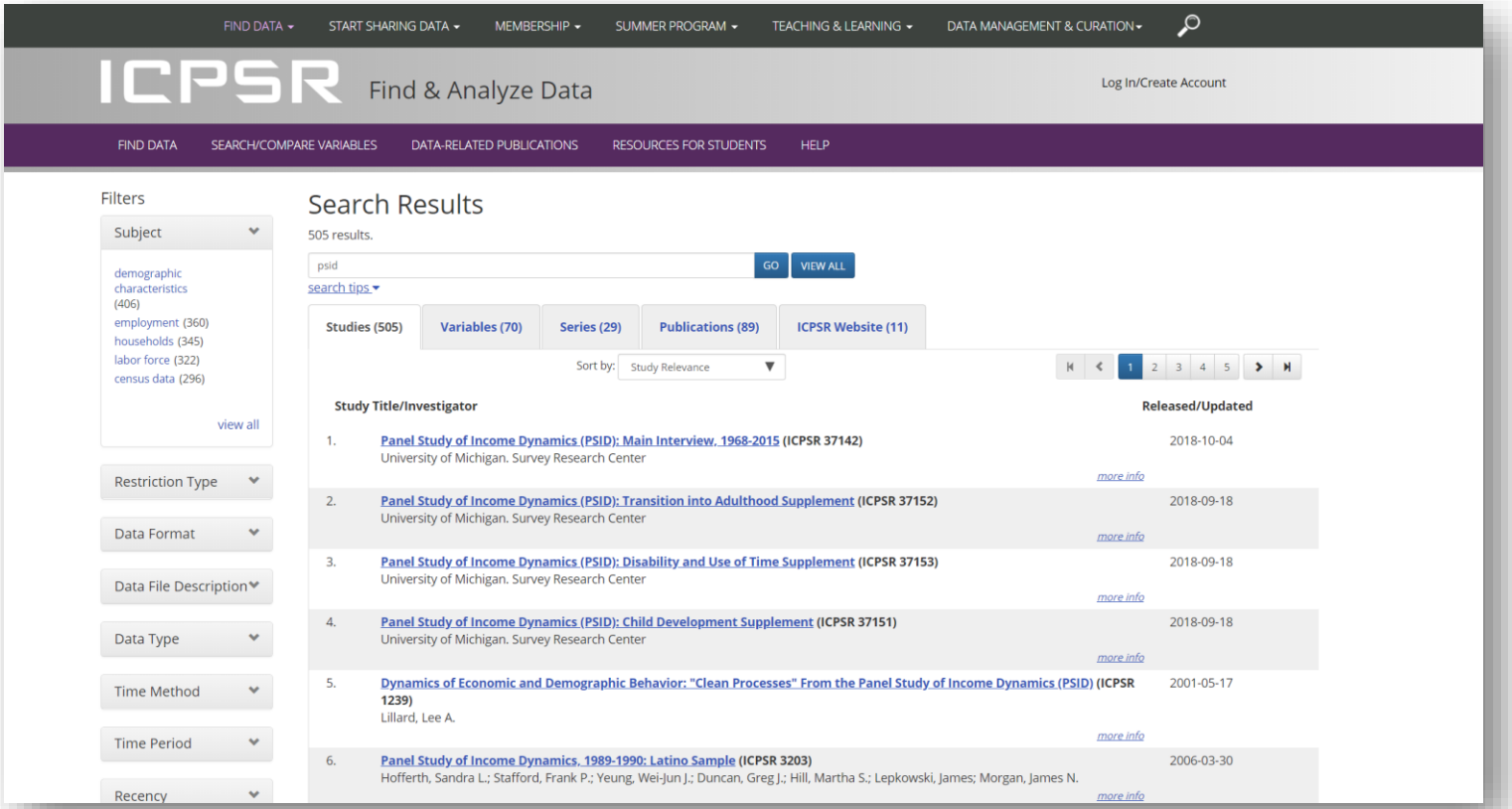
Step 2

Learn from extracted dataset usage and build tools – Recommendations and dataset impact

Step 3

Rinse and repeat - Get feedback and improve

Identify tagged corpus



Step 1 result: Dataset predictions

				varianc
Changing Perceptions of Mental Health in Canada PDF	- 1 0 +1 Senior Health Scholar withAlberta Innovates , Health Solutions (AIHS) National Population Health Survey (NPHS) and the Canadian Community Health Survey (CCHS) cycles Survey CCHS NPHS National Population Health Survey	- 1 0 +1 Canadian Community Health Survey Canadian Community Health Survey	- 1 0 +1 National Population Health Survey	- 1 0 Statistic
The Influence of Flooring on Environmental Stressors PDF	- 1 0 +1 ALOS	- 1 0 +1 Patient Hospital Consumer Assessment	- 1 0 +1 healthcare worker survey	- 1 0 Hospita Assessr Healthc and Sys Respon
Where Broken Windows Should Be Fixed PDF	- 1 0 +1 ICC Study of Crime and Law Enforcement (NSCR)	- 1 0 +1 Study of Crime and Law Enforcement	- 1 0 +1 Study of Crime and Law Enforcement (NSCR)	- 1 0 Netherl Organiz Scientif Crime a 52 4 Innovat Incentiv 516 Jou SSO dat Amster SSO
The not-so-happy ever after PDF	- 1 0 +1 — Empty	- 1 0 +1 — Empty	- 1 0 +1 — Empty	- 1 0 Sansan Indian S electron 200910

Road forward

1. Get central bank publications
2. Evaluate models with central bank publications
3. Build tools: Recommendations and dataset impact factor

Getting the data: Scraping central bank publications

RePEc

EDIRC IDEAS Papers Articles Authors Institutions Data (FRED®)

Economics Departments, Institutes and Research Centers in the World

Central Banks, Monetary Authorities

Associations and Societies
- Association of African Central Banks - AACB
- Central Bank Research Association - CEBRA

Afghanistan
- Da Afghanistan Bank, Kabul

Albania
- Banka e Shqiperise (Bank of Albania), Tirana

Algeria
- Banque d'Algérie (Bank of Algeria), Alger

Angola
- Banco Nacional de Angola (National Bank of Angola), Luanda

Argentina
- Banco Central de la República Argentina (Central Bank of Argentina), Buenos Aires

Armenia
- Central Bank of Armenia, Yerevan

Aruba
- Centrale Bank van Aruba (Central Bank of Aruba)

Australia
- Reserve Bank of Australia, Sydney

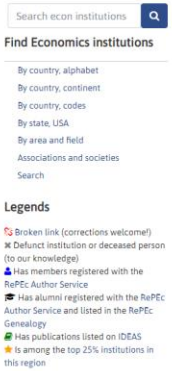
Austria
- Oesterreichische Nationalbank (National Bank of Austria), Wien

Azerbaijan
- Azərbaycan Respublikasının Mərkəzi Bankı (Central Bank of Azerbaijan Republic), Baku

Bahamas
- Bahamas Central Bank, Nassau

Bahrain
- Central Bank of Bahrain, Manama

Bangladesh
- Bangladesh Bank, Dhaka

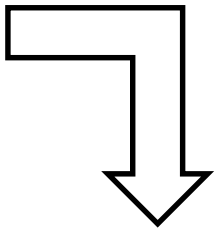

Search econ institutions

Find Economics institutions

By country, alphabet
By country, continent
By country, codes
By state, USA
By area and field
Associations and societies
Search

Legends

Broken link (corrections welcome!)
Defunct institution or deceased person (to our knowledge)
Has members registered with the RePEc Author Service
Has alumni registered with the RePEc Author Service and listed in the RePEc Genealogy
Has publications listed on IDEAS
Is among the top 25% institutions in this region



Retrieve information from HTML with web scraper

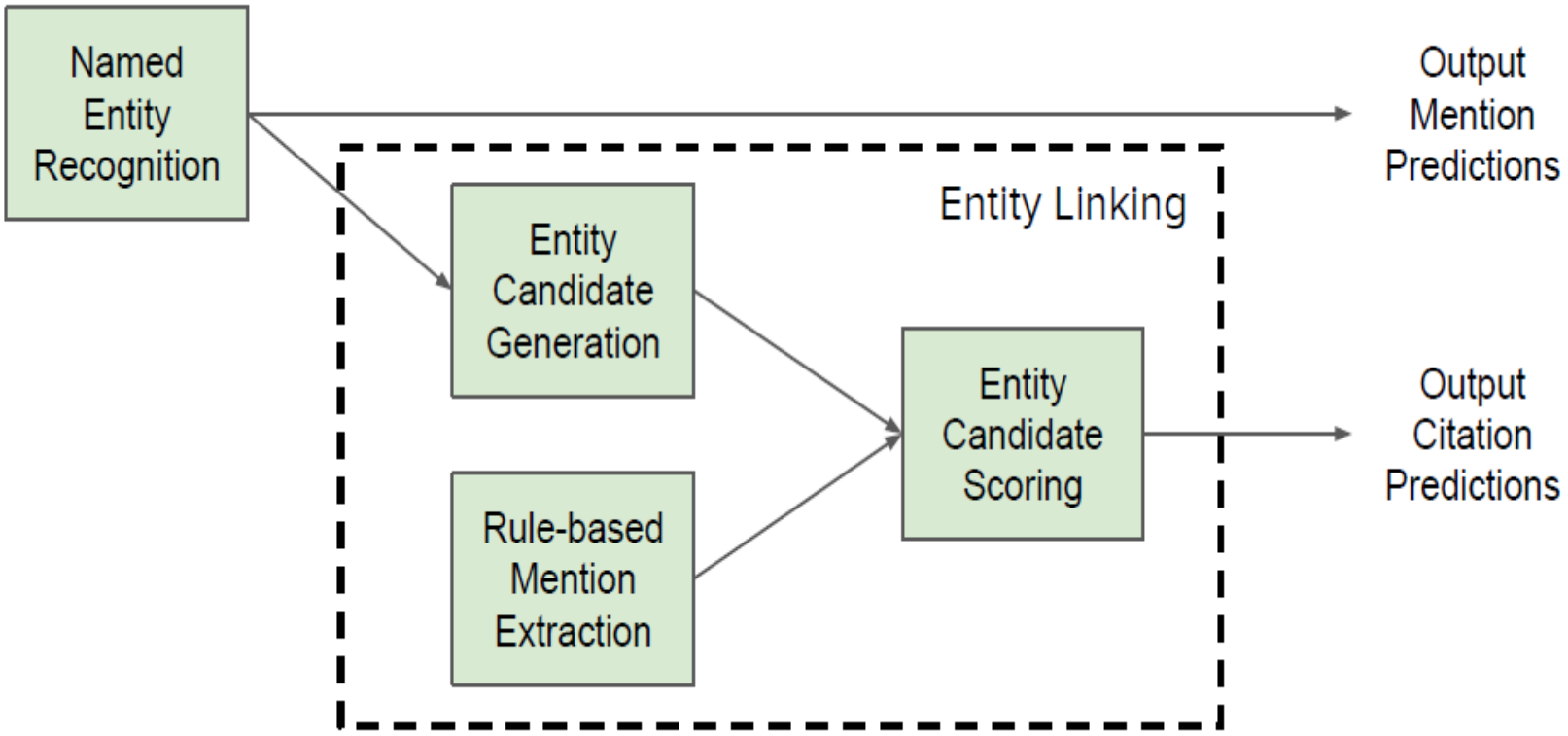
Source: https://edirc.repec.org/central.html

➢Over 57,000 central bank publications downloaded

Thanks to GALILEO-Team

Winning approach: Dataset extraction

1. String extraction
2. Entity resolution



Thanks to Daniel King

INEXDA: The Granular Data Network

On 6th January 2017,

BANCA D'ITALIA · BANQUE DE FRANCE EUROSYSTÈME · BANK OF ENGLAND · BANCO DE PORTUGAL EUROSISTEMA · DEUTSCHE BUNDESBANK EUROSYSTEM

have launched the **I**nternational **N**etwork of **Ex**changing Experiences on Statistical Handling of Granular **Da**ta (INEXDA), an international cooperative project to declare their willingness to further strengthen their cooperation.

Since its foundation, the following institutions have joined INEXDA as members:

BANCO DE ESPAÑA Eurosistema · EUROPEAN CENTRAL BANK EUROSYSTEM · TÜRKİYE CUMHURİYET MERKEZ BANKASI · BANCO CENTRAL DE CHILE · БАНК РОССИИ · eurostat

INEXDA‘s General Mission

- General mission is to **promote data sharing and data access**
- Promoting the **G20 Data Gaps Initiative II**, in particular recommendation 20, addressing the accessibility of granular data. INEXDA is mentioned in a G20 paper
- Acknowledging and supporting the work on data sharing of the **Irving Fisher Committee** on Central Bank Statistics
- INEXDA is governed by an **MoU**, that every member has to sign
- Sharing of granular data between INEXDA members not part of this MoU

Work program for the first two years

1 Comprehensive inventory of **data** in all member institutions

- Agreement on unified metadata schema
- Setup of a platform to collect and exchange metadata
- Start harmonizing metadata across INEXDA member countries

2 Comprehensive inventory of existing **data access procedures**

- ECB pilot collection of information on access for researchers
- ADRF for INEXDA proposed by Julia Lane (NYU)
- Setup of working groups

3 Dissemination of INEXDA **results**

- Prototype of INEXDA webpage by Banque de France.

Working Group on Data Access

1 INEXDA workshop on data access procedures

- Serve as a kick-off event for the INEXDA work stream on data access
- Utilise current momentum to improve data access procedures for data producers and researchers.

2 Define common access procedures

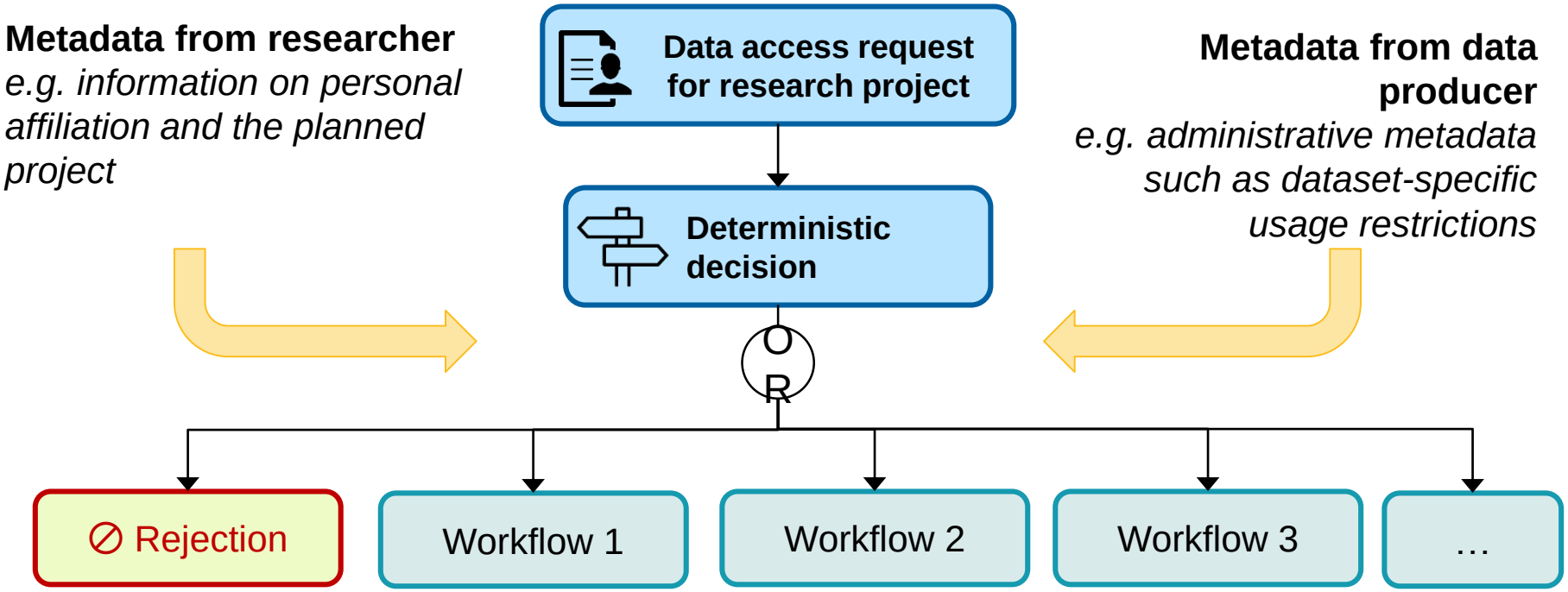
- Identify similar workflows used by data providers.
- From there define a standard
- Need to be in compliance with internal data policies and external government regulations.

3 Development of open software solution

- Set of common access procedures will define software requirements
- Needs to be flexible enough to accommodate different legal frameworks

Ideal state: Automation of decisions in workflows

- ✓ **Metadata** describes how data can be managed and used from a legal, organisational or technical point of view
- ✓ **Workflows follow deterministically** from such administrative dataset metadata



How to become an INEXDA member

Other central banks, national statistical institutes and international organizations are **encouraged to join INEXDA**.

*The following **procedure** has been established to admit new members:*

	A representative of the applying institution has to attend at least one INEXDA meeting in person before the formal application process is initiated
	Send an application letter signed by the head of the statistical department of the respective institution (or, in case of national statistical institutes, by the head of the responsible statistical department) to the chair of INEXDA.
	All INEXDA members must agree to any application by a new institution.
	Sign the MoU*. Congratulations, you are now a member of INEXDA.

* The INEXDA Memorandum of Understanding (MoU) can be accessed here.

INEXDA's metadata schema

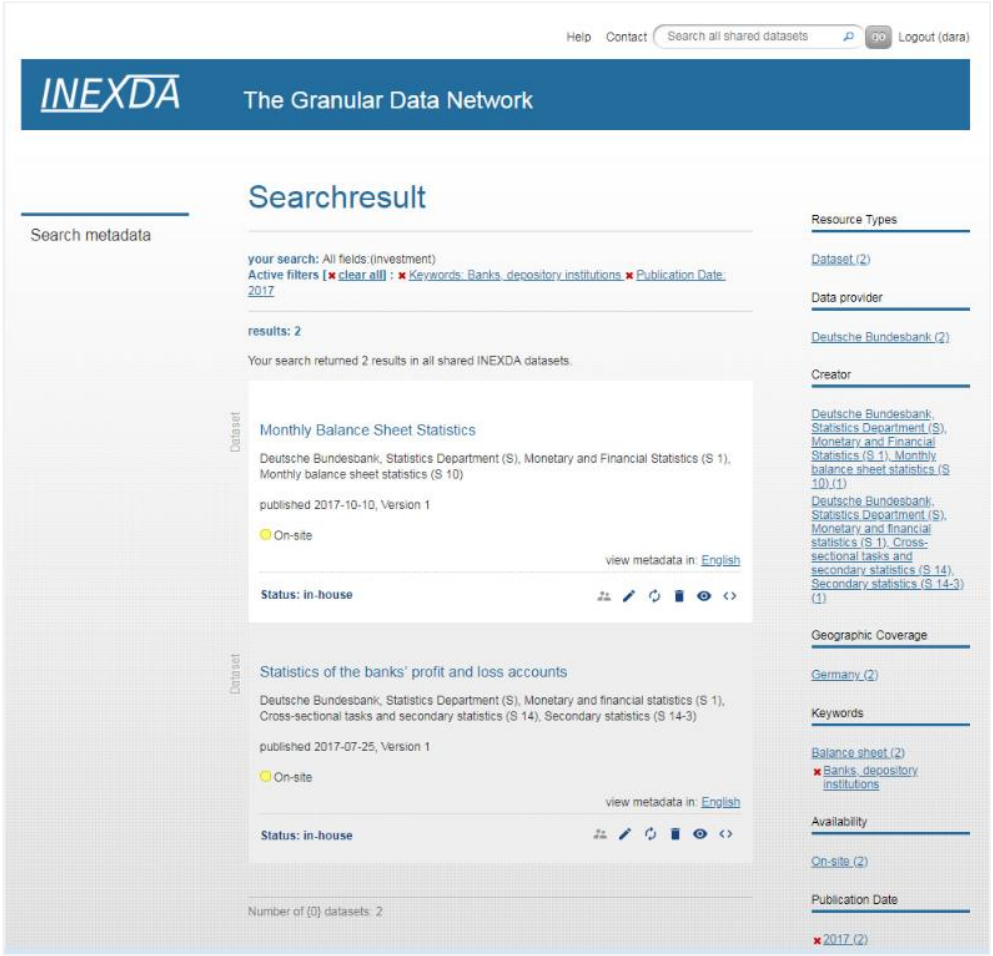
1	Resource Type
2	Resource Identifier
3	Name of Dataset
4	Creator
5	DOI Proposal
6	URL
7	Language of Resource
8	Publication Date
9	Availability
10	Sampled Universe
11	Sampling
12	Temporal Coverage
13	Time Dimension
14	Collection Mode
15	Unit Descriptions
16	Descriptions
17	Geographical Coverage
18	Keywords
19	Alternative Identifiers
20	Relations
21	Publications

- Based on the GESIS DOI registration service da|ra (GESIS is cooperating with DataCite):

  https://www.da-ra.de/en/home

- Name of metadata items closely follows da|ra conventions to enable seamless DOI registration, if desired later in the project.

- Is basis for INEXDA metadata database that was established to store and view metadata from INEXDA members.



The database serves as …

1. An information tool for INEXDA members, researchers and analysts
2. The basis for the harmonization activities of INEXDA (e.g. item relation)

Current contributions to the database come from

- Deutsche Bundesbank
- European Central Bank
- Banca d'Italia
- Banco de España
- Banco de Portugal