

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Enhancing user knowledge by providing (richer) context to microdata¹

Christian Hirsch,

Deutsche Bundesbank

¹ This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Enhancing user knowledge by providing (richer) context to microdata

Christian Hirsch, Deutsche Bundesbank

**IFC-BNM-ECB Satellite Seminar on
"Post-crisis data landscape: micro data for the macro world"**
August 2019

Based on a project with and contributions from:

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

Motivation

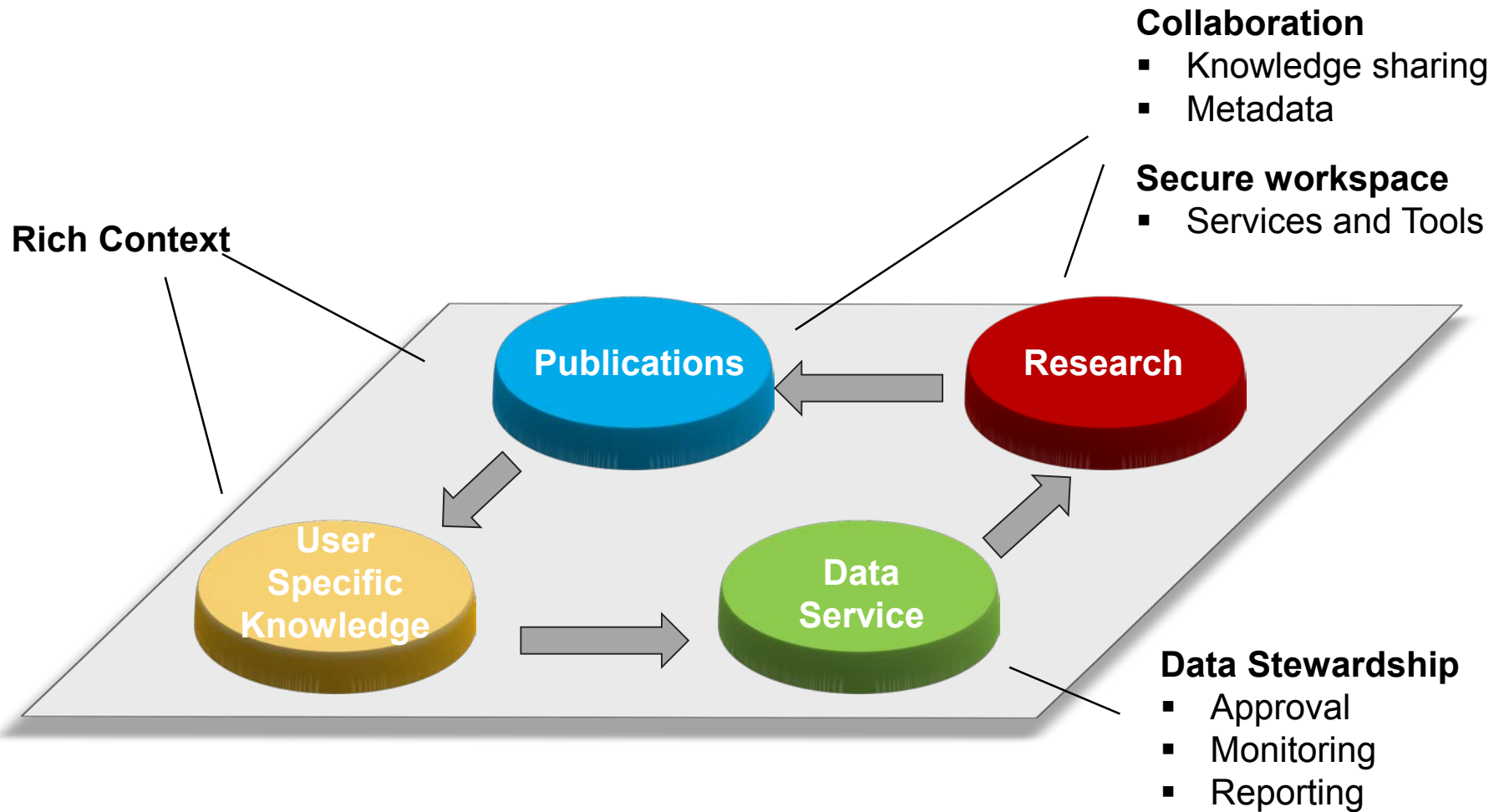
Status Quo: One-directional information flow

- Researchers use microdata and metadata to produce outcome (publications)
- No structured knowledge flows back to data producers

Goal: Enable feedback loops

- Make knowledge generated in the research process usable
- Two sources of information to extract:
 1. Automatic: Text mining to find information in research publications
 2. Tacit knowledge: Enable and incentivize data users to feedback information

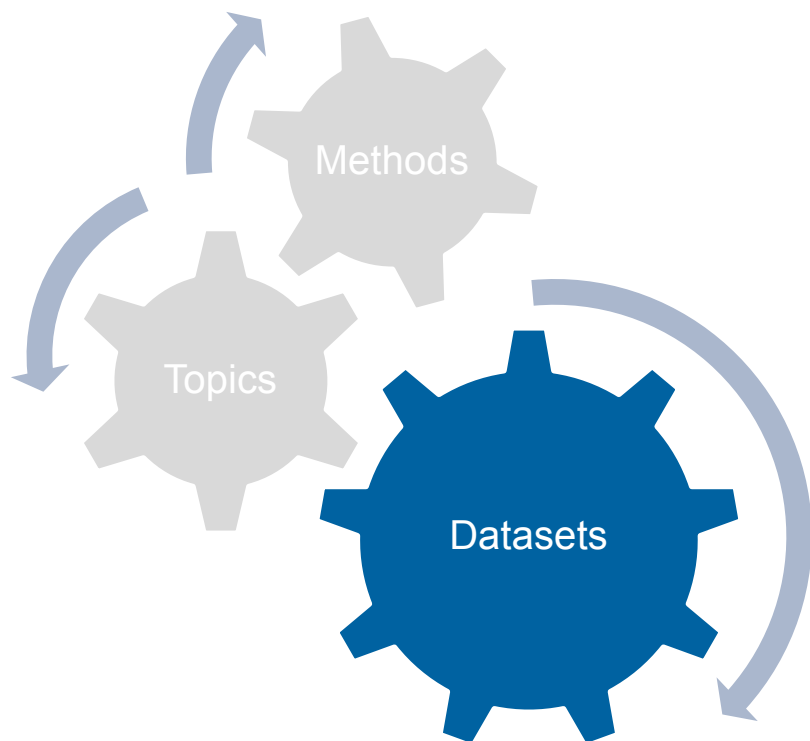
Knowledge Life Cycle (IDIS-R)II



How do we do this?

- **Step 1:** Create the set of corpora and usage side metadata (machine learning techniques) - Competition
- **Step 2:** Figure out how you learn from it and automate it (machine learning techniques) – Dataset recommendation engine
- **Step 3:** Engagement – recognize and emphasize patterns (with human curation) – Rinse and repeat

Step 1: competition design



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.¹⁹²⁰

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 non-construction 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.

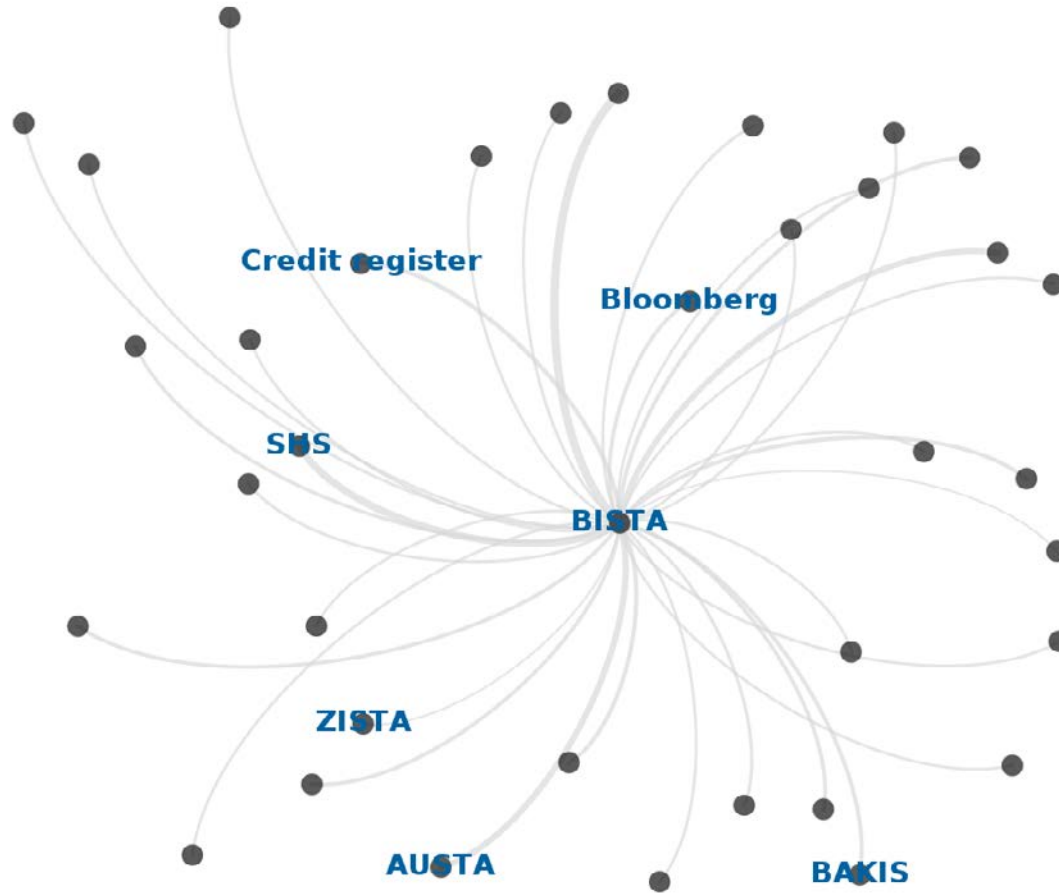
Our analysis is at the bank-state level. The regional differentiation of bank loans is possible because of a feature of the German banking system: the presence of Sparkassen (savings banks) and Volksbanken (cooperative banks). Each bank is only present in one German state. Sparkassen are geographically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

¹⁹Schufa 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.

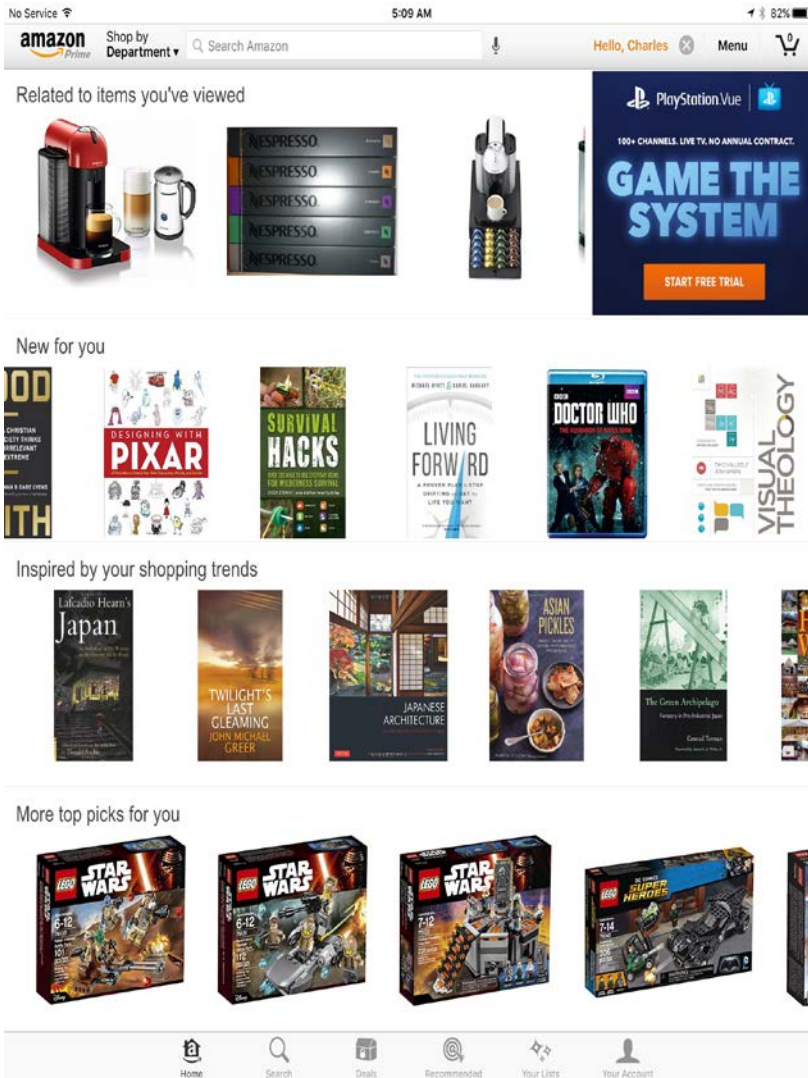
²⁰For reasons of data confidentiality, Auxmoney provides its credit intermediation by month and state only if five or more loans were made in that month in that state.

²¹The 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.

Step 2: use of user specific knowledge - input



Step 2: use of user specific knowledge



Related to data you've viewed

New data similar to data you've used

What others have done with similar data (recipes)

Recipes like yours

Interaction with INEXDA's metadata schema

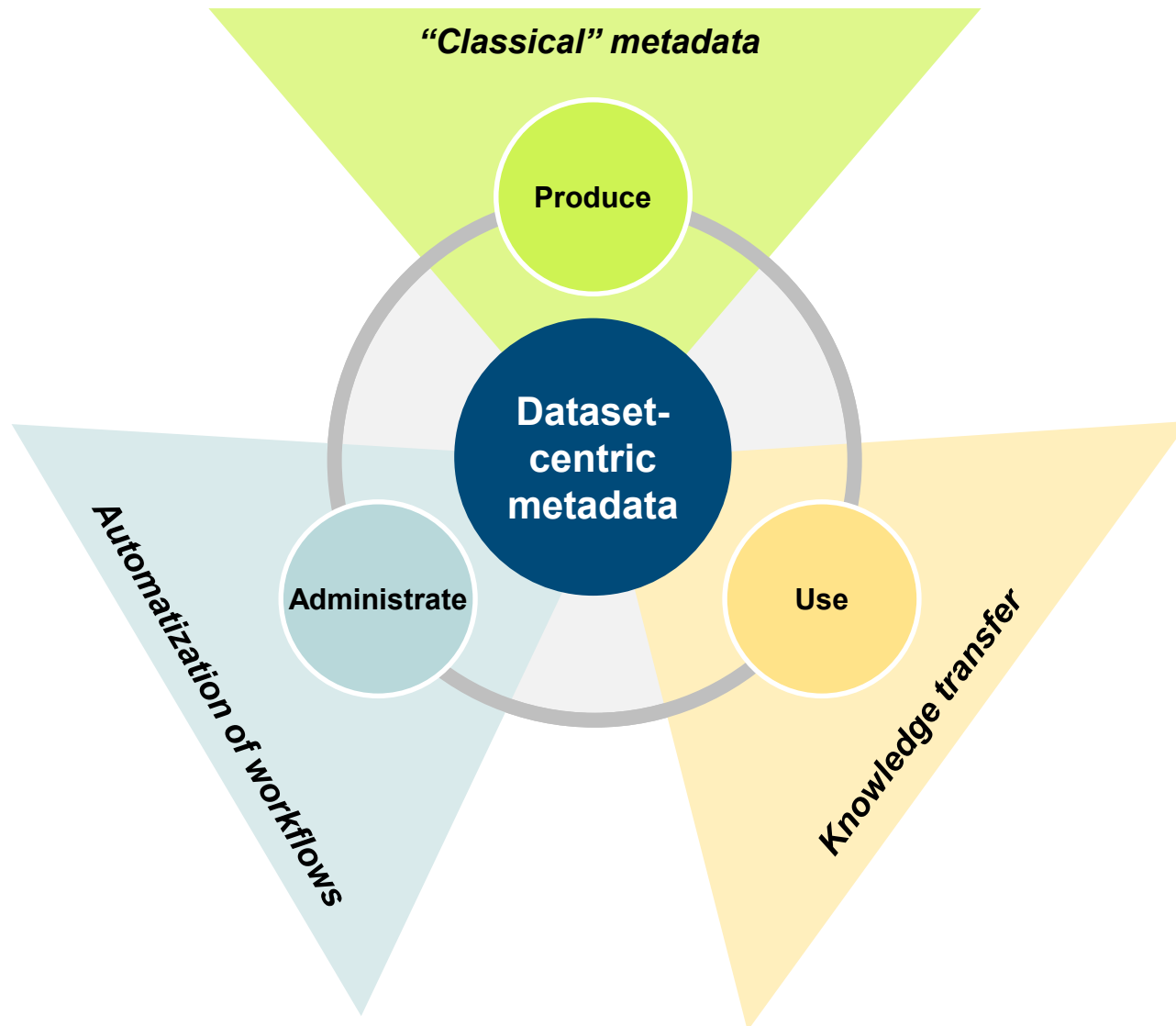
1	Resource Type
2	Resource Identifier
3	Name of Dataset
4	Creator
5	DOI Proposal
6	URL
7	Language of Resource
8	Publication Date
9	Availability
10	Sampled Universe
11	Sampling
12	Temporal Coverage
13	Time Dimension
14	Collection Mode
15	Unit Descriptions
16	Descriptions
17	Geographical Coverage
18	Keywords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

- Is basis for INEXDA metadata database that was established to store and view metadata from INEXDA members.
- Conclusions:
 - Items are helpful for producing meaningful dataset recommendations for researchers (e.g. Temporal Coverage)
 - Need to amend schema to include usage metadata generated from project

Amend existing metadata to include information on use



Conclusion

WHAT WE DID

- Assess feasibility of meaningful dataset recommendations for researchers based on dataset recognition in publications
- Design a mechanism enabling automated knowledge flow back from publications
- Develop a concept to engage users to contribute knowledge to existing metadata

WHAT'S NEXT

End of Aug 2019

Pilot of dataset recommendation tool

15-16 Nov 2019

Rich Context Workshop hosted by NYU

2020

On-board tool to DSM (if possible)

Contact

Christian.Hirsch@bundesbank.de

Website:

<https://www.bundesbank.de/en/bundesbank/research/rdsc>

Contact: fdsz@bundesbank.de