

BERD
@NFDI

8. Bibliothekskongress Leipzig 2022

Bibliotheken als Partner für die Nationale Forschungsdateninfrastruktur am Beispiel von BERD@NFDI

Dr. Sabine Gehrlein | Prof. Dr. Klaus Tochtermann

2. Juni 2022

Bibliotheken können Forschungsdaten



Bibliotheken in der NFDI



Sprint oder Marathon?



1. BERD@NFDI: Ausgangslage und Mission
2. Bibliotheken als zentrale Akteure im Konsortium
3. Aktuelle Herausforderungen und Lösungsansätze

1. BERD@NFDI: Ausgangslage und Mission

1930er



Source: Archives for the History of Sociology in Austria (Graz), »Marienthal« Virtual Archives

1980er



Source: ISR Archive

seit 2010



Source: IAB SMART Study, Kreuter et al.

- + Detailliert
- Beobachtungsfehler
- Relativ geringer Umfang

- + Standardisiert
- + Großer Umfang
- Teuer
- Hohe Belastung
- Falschmeldungen

- + Standardisiert
- + Großer Umfang
- + Günstig
- + Niedrige Belastung
- **Komplexes Prozessieren**
- **Tools und Infrastruktur fehlen**

Sentiment-Analyse von Twitter-Tweets: Werden Meinungsumfragen überflüssig?

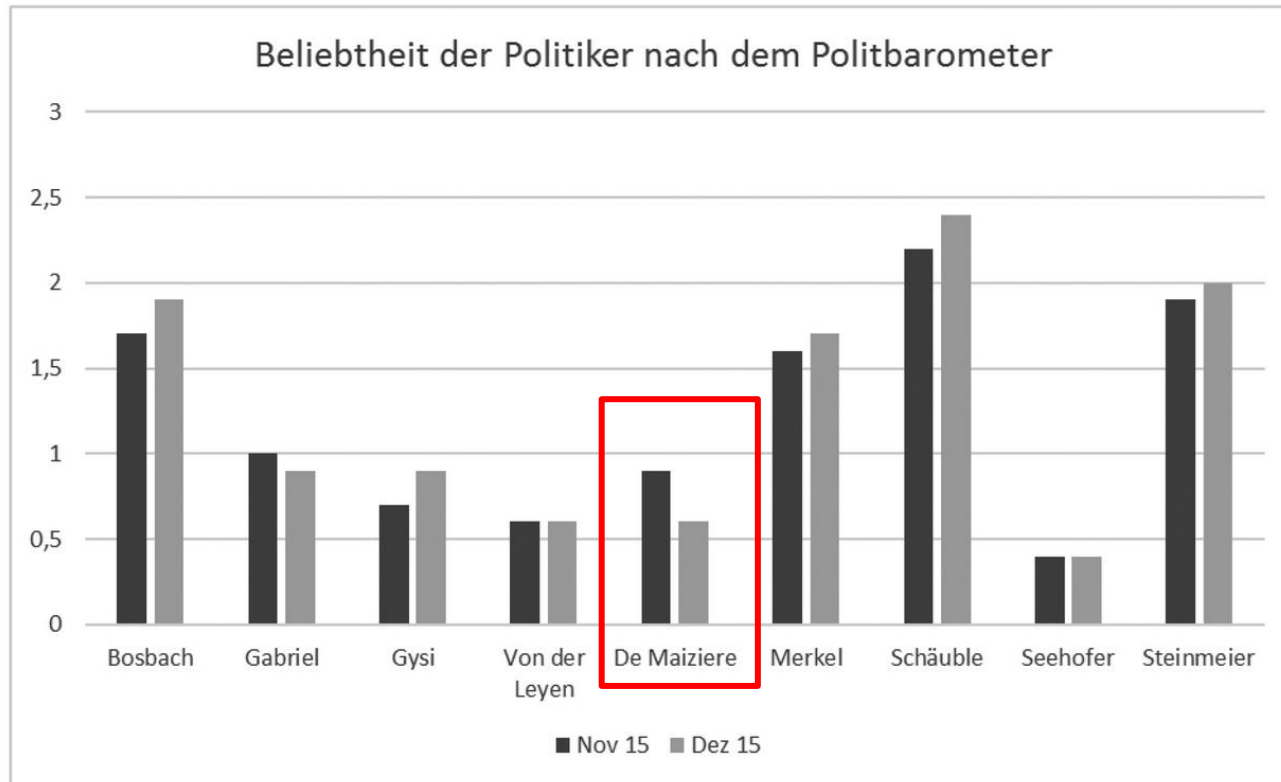


Abbildung 6:
Beliebtheit der Politiker im November und Dezember 2015.

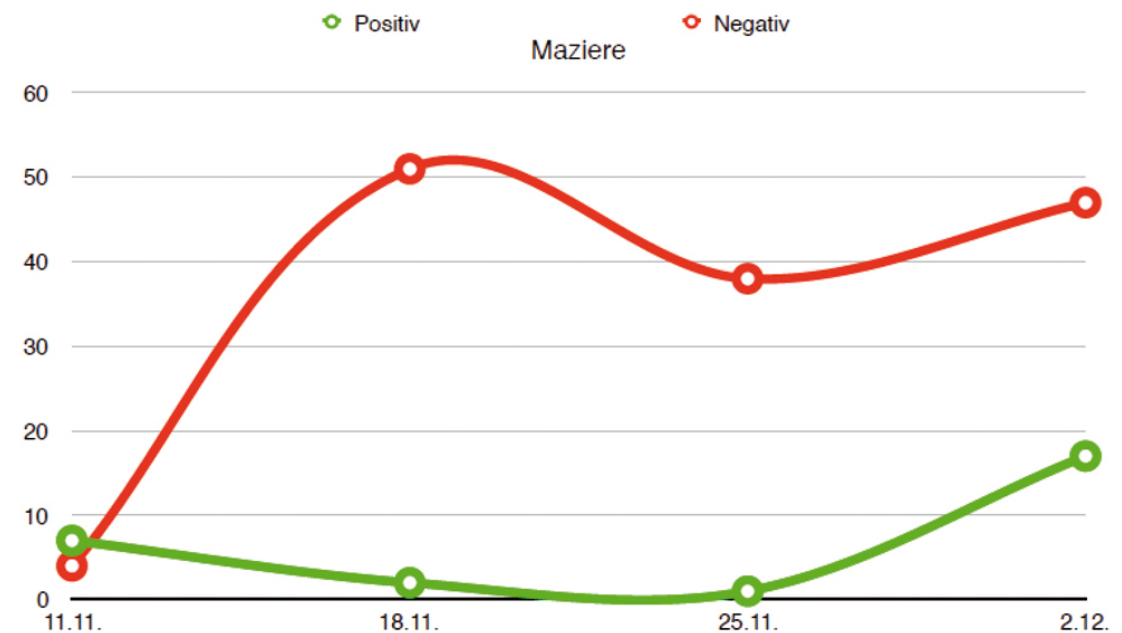


Abbildung 8:
Tweets zu De Maiziere im November 2015.

Wie steht es um die Anwesenheit von Abgeordneten im Landtag?

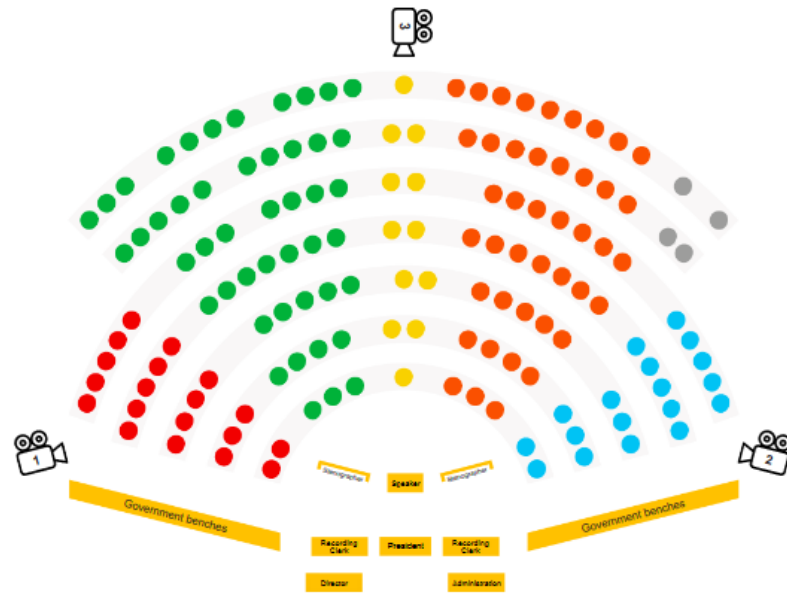
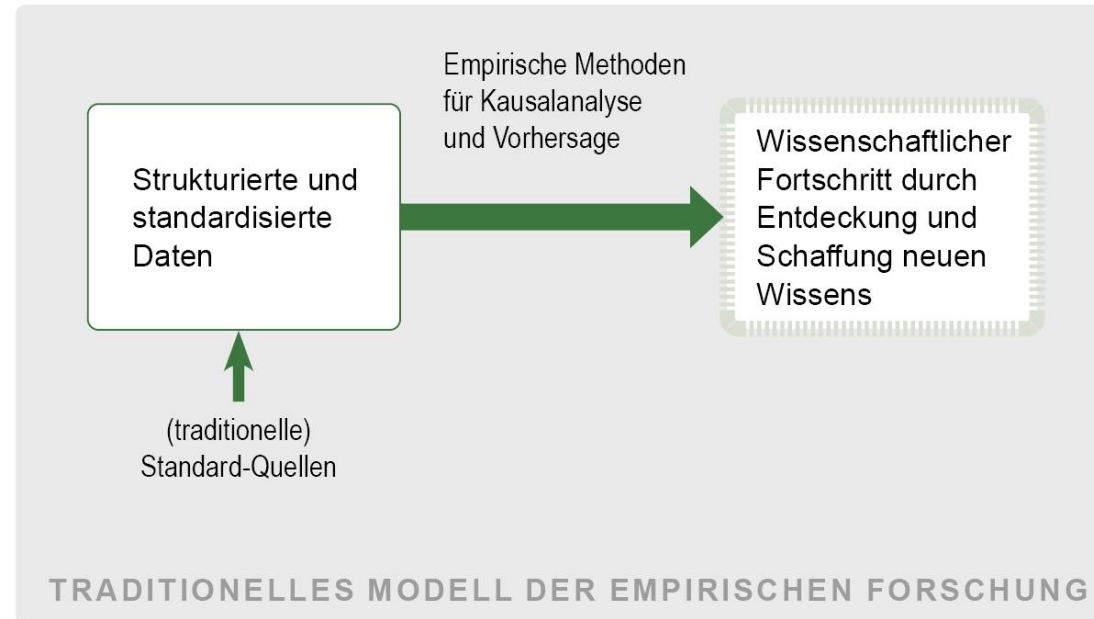


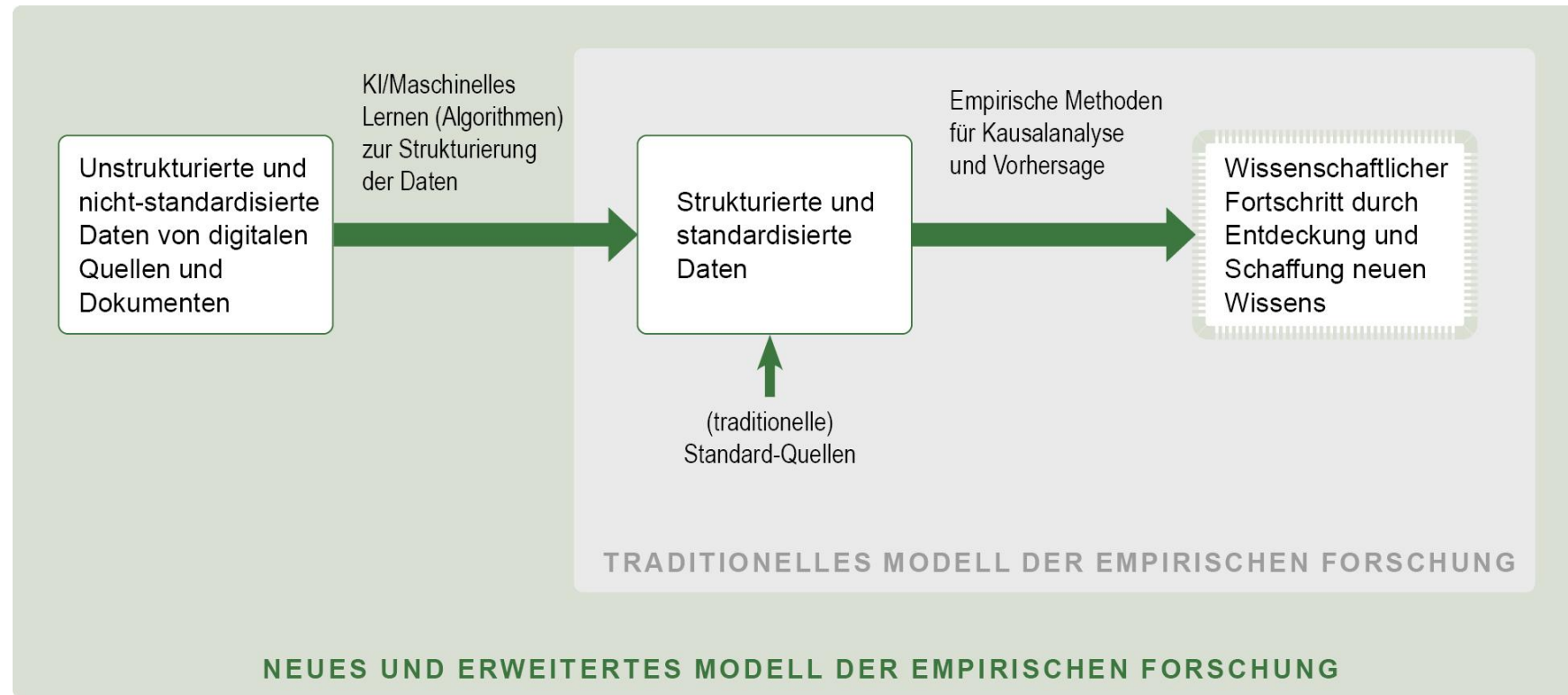
Figure 26.1: Camera locations in the *Landtag Baden-Württemberg*



Figure 26.2: Screenshot from camera 2 of the *Landtag* plenary chamber focusing on the Green party

Aus: "Automated video analysis for social science research" von Gschwend et al.



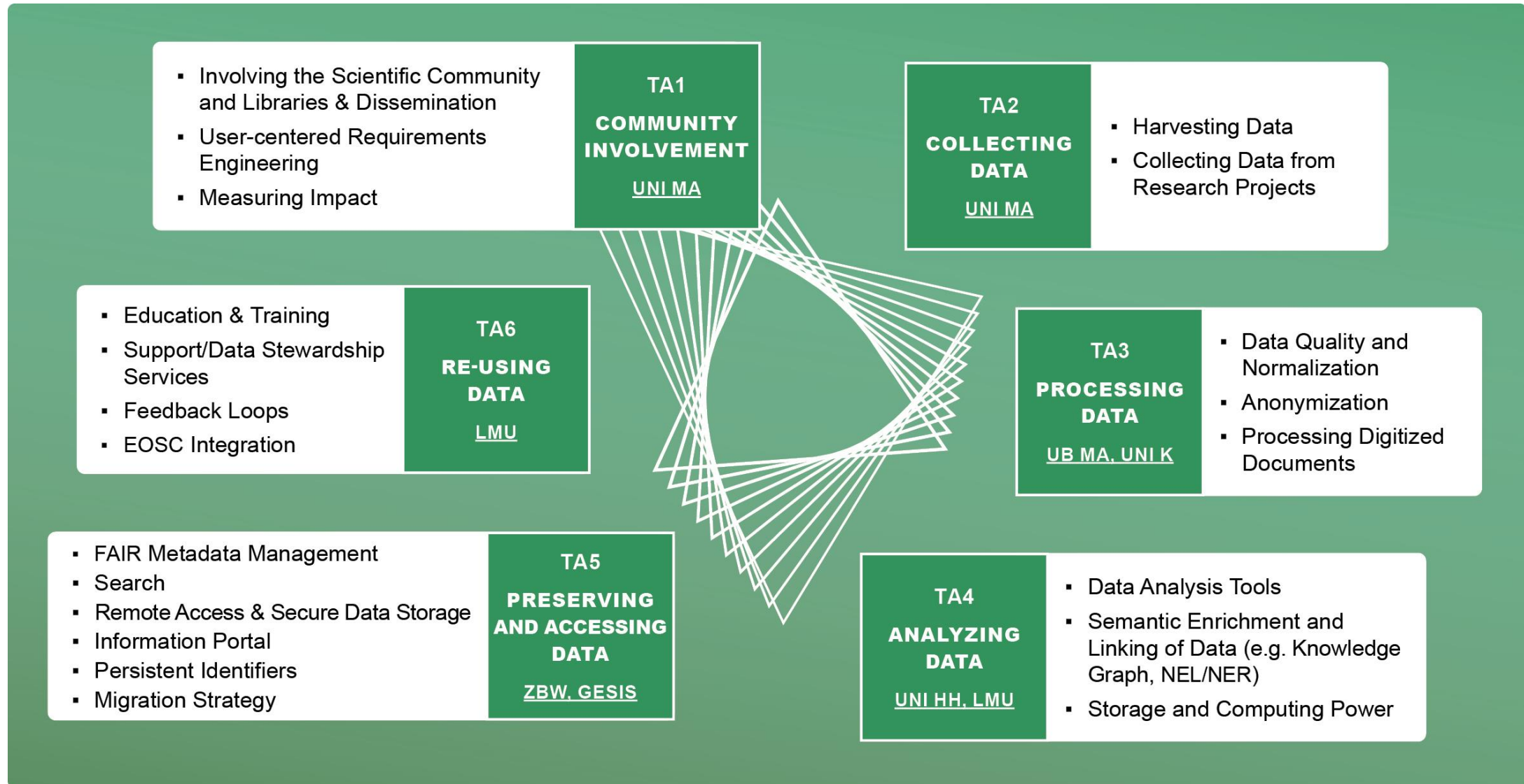


- Fülle neuer komplexer Daten und Datentypen: großes Potenzial für neue wissenschaftliche und gesellschaftsrelevante Erkenntnisse
- "Methodologischer" Aufwand beim Sammeln, Prozessieren und Analysieren der Daten
- Verbunden mit hohem technischem Aufwand
- Risiko von irreführenden und nicht reproduzierbaren Ergebnissen

Ziel: Aufbau einer zentralen Plattform für das Management unstrukturierter und nicht-standardisierter Daten sowie wissenschaftlicher Software in den Wirtschafts- und Sozialwissenschaften

BERD: **B**usiness, **E**conomic and **R**elated **D**ata

Team: Partner aus Wissenschaft und Infrastruktur, u.a. Universitäten Mannheim, Köln, Hamburg, LMU München, GESIS, Leibniz-Informationszentrum Wirtschaft ZBW





“It is difficult to make a **comprehensive list of sources**, where I can find the data I need”

“**Scraping data should be a joint effort**, e.g. to collect annual reports of companies, scraping the text from it, maybe cleaning it [...] so that this work isn't done by ten different people independently from each other”

“It would be helpful to have a platform **where you can talk about your ideas with other people**, who worked on this field to get some tips about codes, sources and databases”

“**Structuring the unstructured data** is very difficult, since I do not have the expertise and I would need some contact person who has more skills on this topic”

2. Bibliotheken als zentrale Akteure im Konsortium

Leibniz Informationszentrum Wirtschaft ZBW

→ Deutschlandweite Infrastrukturen zu Publikations- & Forschungsdaten der Wirtschaftswissenschaften

Universitätsbibliothek Mannheim

→ Dienstleistungen zu Forschungsdaten sind ein wichtiger Bestandteil des Serviceportfolios

Beide Bibliotheken verfügen über hohe Expertise in:

- Forschungsdatenmanagement
- Management von Metadaten, Standardisierung, Education
- Entwicklung und dauerhafter Betrieb von Infrastrukturen
- Serviceorientierung

- Partner der Wissenschaft auf Augenhöhe
- Steuerung des operativen und technischen Projektmanagements

ZBW:

Chief Technology Officer CTO

UB Mannheim:

Administrative Projektkoordination



Bild: iStock.com/Pepleimages

3. Aktuelle Herausforderungen und Lösungsansätze

- Zusammenschluss von starken Einzelpartnern und Einzelinitiativen
- Keine Fortführung eines bereits bestehenden Konsortiums
- Konsortium in Förderung seit November 2021
- Onboarding der Projektmitarbeiter*innen derzeit im vollen Gang

- Start-up-Charakter
- Hohe Agilität

- Absolut Wissenschafts-getrieben!
- Intensiver Austausch zwischen Wissenschaft und Infrastruktur



Spannungsfeld zwischen Ergebnisorientierung und Gesamtvision


- Agilität vs. langfristig tragenden Strukturen
- Schnell zur Verfügung stehende Services vs. optimale Vernetzung der Services untereinander
- Unmittelbare Einbindung der Benutzer vs. umfassende bedarfsorientierte Evaluation im Vorfeld
- Einzelservices vs. koordiniertes Angebot



Schnelle Umsetzung von „First Services“

- Open Data Directory
- Data Marketplace

- Rapid prototyping
- Gewinnung von Nutzer*innen, Lernen aus dem Benutzerverhalten
- Interaktion mit der wissenschaftlichen Community
- Inkaufnahme, dass evtl. im Nachgang „händisch“ an Gesamtstruktur angepasst werden muss



animation that shows what this site is about

Search 86 datasets

or browse by:


i

g

u


industry region field of research

Current Dataset Offers




Recommender Competition, 10.2 Mio User

Deadline: 2023-08-01




Customer Journey Optimization

Deadline: 2023-08-01



Online Review Sentiment Analysis


Deadline: 2023 -09-14



Blockchain: Parachain Auction Crowdfund Analysis

Deadline: 2023 -10-01

Success Story



A research team from the chair of quantitative marketing at University of Mannheim wins Robert D. Buzzel MSI **Best Paper Award** for most significant contribution to management practice and thought: "research with lasting value to corporate executive". The research was based on the dataset "2022 Swartz Corporate Executive Survey" that was offered as a competition dataset on our platform.

Congratulations from the BERD team!


Latest

Latest Feed containing certain hashtag or @BERD

Latest related linkedin post

Lastest comment on any dataset on this site

Editorial message to our visitors, posted from backend



The Economist Historical Advertisements – Objects Dataset

c

m

competition dataset metadata only

This dataset contains 191.994 identified object locations and classes for all historical advertisements from all 8.840 issues of The Economist magazine, years 1843 to 2014. We used a state of the art classifier to detect the objects: https://thub.dev/google/openimages_v4/ssd/mobilenet_v2/. You will need this **Master Dataset**, as well, to work with the data.

Keywords: Advertising, Marketing, Historic, The Economist, Object detection.

Attributes
[show more](#)

Get started
6 researchers already applied for this dataset
Deadline: 2023-09-01
[Apply now](#)

Pricing

- Full individual license: 200 €
- Full institutional license: 100 €

[show more](#)



Data Use Agreement

Licensed materials: Licensed material under this agreement is all video data from the L&L advertisement data set in mp4g format.

Authorized user: authorized user under this agreement is the researcher applying for data access and signing this agreement.

Authorized user: authorized under this agreement is the use of the licensed material for personal research purposes of the authorized user. The authorized user is permitted to share parts of the licensed materials with co-authors of his/her research.

Warranty: we guarantee 24/7 access to the data. We guarantee that we have the right to sublicense the data.

Liability: we want to stress that we are not liable for anything, especially not for acts of god. All liability will be transferred to the authorized user.

I agree to comply with all provisions as stated above, especially with the confidentiality clause with all security measures

I agree to refrain from trying to extract data from the platform, e.g. by photographing, screenhulding, screencasting, copying etc.

I will not share my access data to the dataset with other persons than the authorized user(s)

[Confirm and access data](#)

Entwicklung einer Gesamtvision als „Click-through“-Showcase

- Gemeinsame Verständigung über Gesamtvision
- Grundlage für user-driven Software Design mit der Fachgemeinschaft
- Sammeln der Services und der damit verbundenden Bedarfe
- Koordination der Services
- Möglichkeit zur Diskussion und Reflexion der Zielsetzungen



BERD@NFDI: Sprint und Marathon!



Geschafft! Danke!

Fragen, Anmerkungen?

Dr. Sabine Gehrlein | sabine.gehrlein@bib.uni-mannheim.de

Prof. Dr. Klaus Tochtermann | k.tochtermann@zbw.eu

Gehrlein, S., Klein, A., Schumm, I., Tochtermann, K., 2022. BERD@NFDI für unstrukturierte Daten in den Wirtschafts- und Sozialwissenschaften. Zeitschrift für Bibliothekswesen und Bibliographie | ZfBB 69 (1-2), S. 59-67.
<https://doi.org/http://dx.doi.org/10.3196/1864295020691280>.

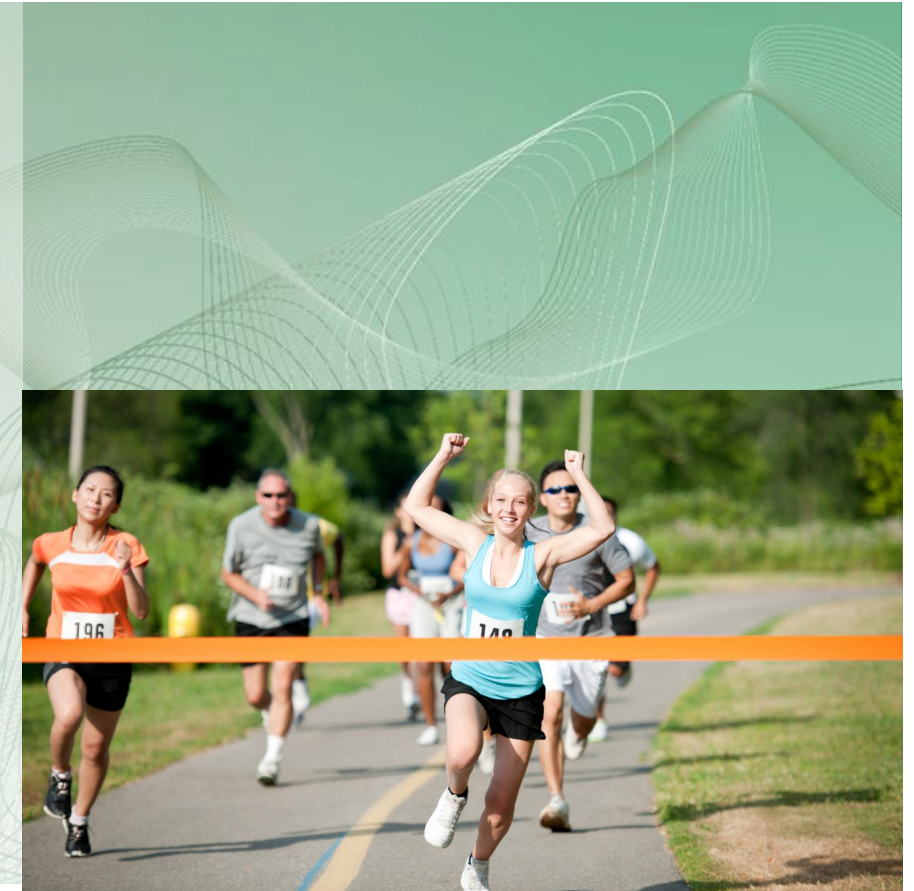


Bild: iStock.com/Fatcamera