



BERD
@NFDI

112. Bibliocon 2024

Die BERD@NFDI Plattform: Eine Forschungsinfrastruktur für unstrukturierte Daten in den Sozialwissenschaften

Dr. Ulrich Krieger
Dr. Renat Shigapov

07.06.2024



<https://berd-platform.de>

BERD steht für
Business, **E**conomic und
Related **D**ata



A Single Place for Data in Business, Economics and Related Fields



Data for Science... created by the community

BERD@NFDI empowers researchers to enhance the full potential of (big) data along the entire research journey.

[Login](#)[Explore](#)

1

Veröffentlichen, Teilen und Finden von (unstrukturierten) Daten

2

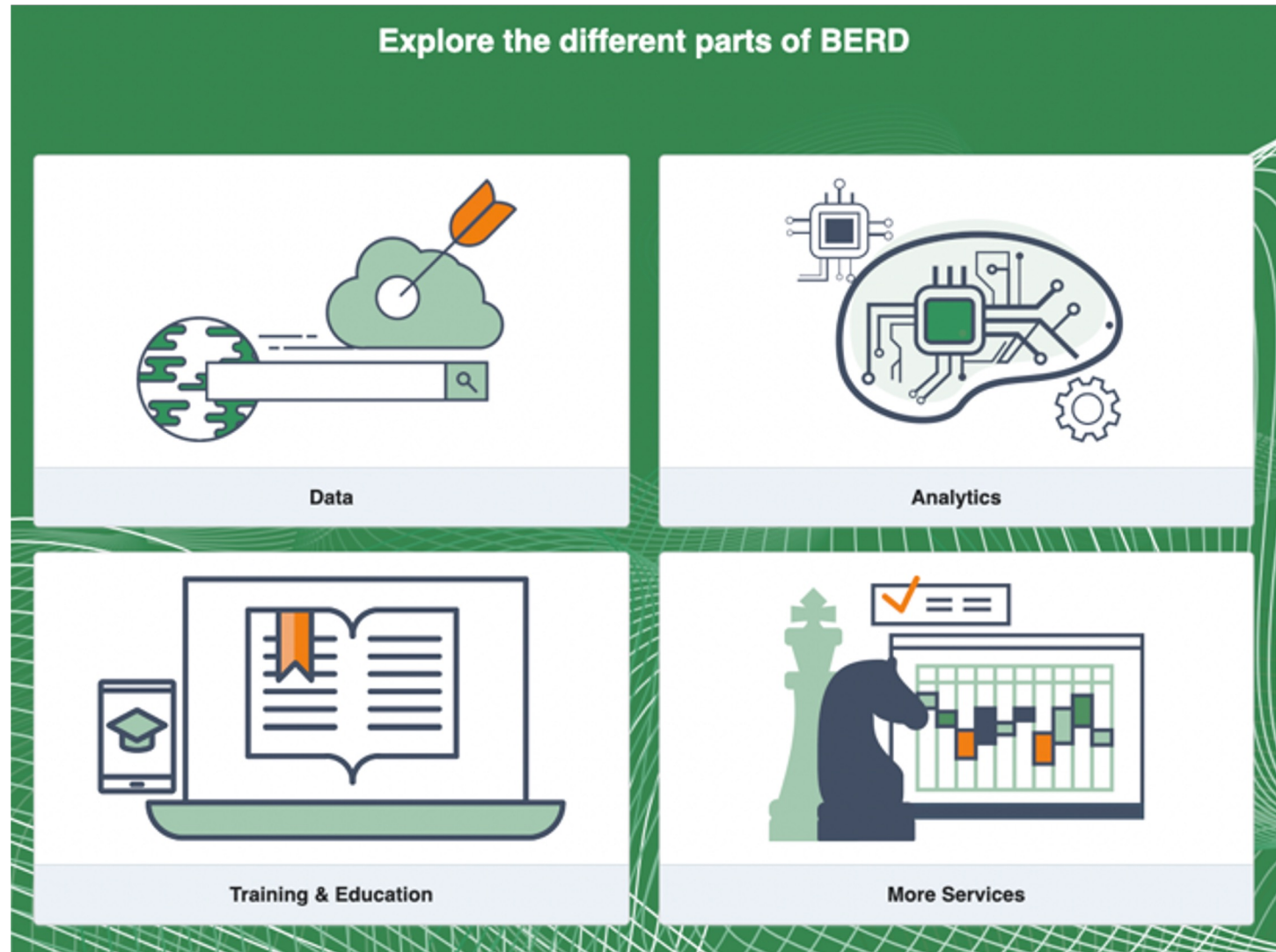
Modernste Methoden und Algorithmen für die Arbeit mit (unstrukturierten) Daten

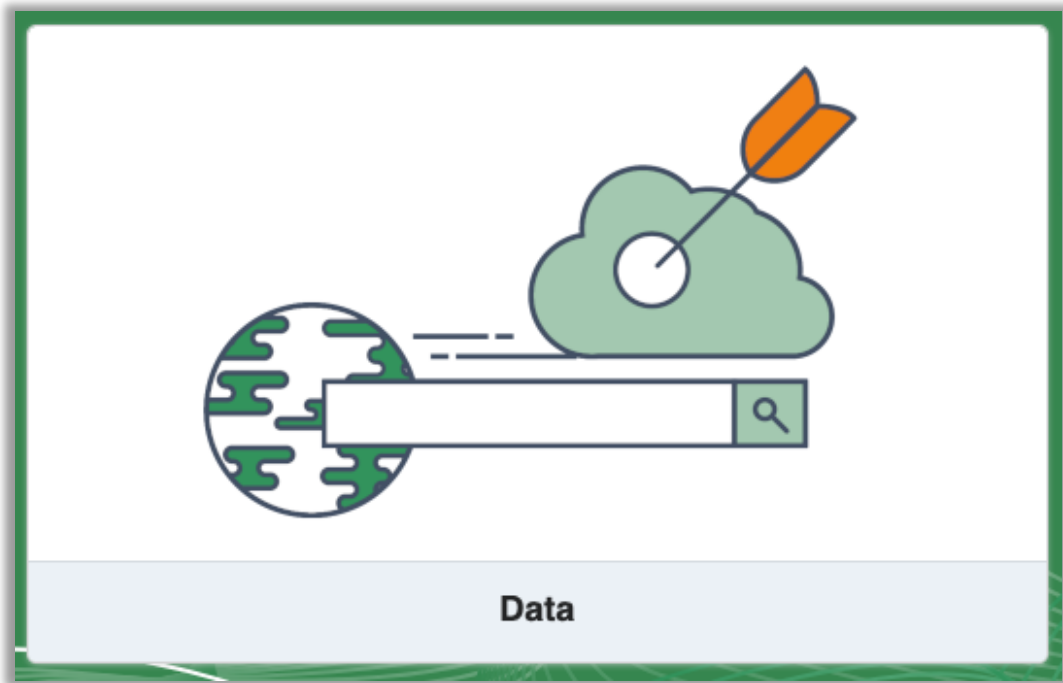
3

Lernmaterialien und Veranstaltungen zum Thema (unstrukturierte) Daten

4

Umfassende Werkzeuge und Anleitungen für die Arbeit mit (unstrukturierten) Daten





- ✓ Hochwertige Datensätze finden, sammeln und kuratieren
- ✓ Zusammenarbeit mit Industrieorganisationen erleichtern
- ✓ Forschern das Teilen und Veröffentlichen von Daten ermöglichen

Durchsuchen und
browsen Sie in einer
Auswahl hochwertiger,
offener (Big) Datensätze
mit hoher
Forschungsrelevanz

October 15, 2019 (v2) Metadata-only Open dataset

CrowdTangle Platform and API

Garmur, Matt ; King, Gary ; Mukerjee, Zagreb

This document describes the CrowdTangle API and user interface being provided to researchers by Social Science One under its collaboration framework with Facebook. CrowdTangle is a content discovery and analytics platform designed to give content creators the data and insights they need to succeed. The CrowdTangle API surfaces stories, and data to measure their social performance and identify influencers. This codebook describes the data's scope, structure, and fields.

Uploaded on June 28, 2023

social sciences Facebook data codebook

2019 (v2) Metadata-only Open dataset

AmazonQA

Gupta, Mansi ; Kulkarni, Nitish ; Chanda, Raghuv eer

We introduce a new dataset and propose a method that combines information retrieval techniques for selecting relevant reviews (given a question) and "reading comprehension" models for synthesizing an answer (given a question and review). Our dataset consists of 923k questions, 3.6M answers and 14M reviews across 156k products. Building on the well-known Amazon dataset, we collect additional annotations, marking each question as either answerable or unanswerable based on the available reviews.

Uploaded on June 28, 2023

Amazon Question Answering multiple categories open-ended questions yes/no-questions Product reviews Machine learning Information retrieval

2014 (v2) Metadata-only Open dataset

Flickr30k

Young, Peter ; Lai, Alice ; Hodosh, Micah

To produce the denotation graph, we have created an image caption corpus consisting of 158,915 crowd-sourced captions describing 31,783 images.

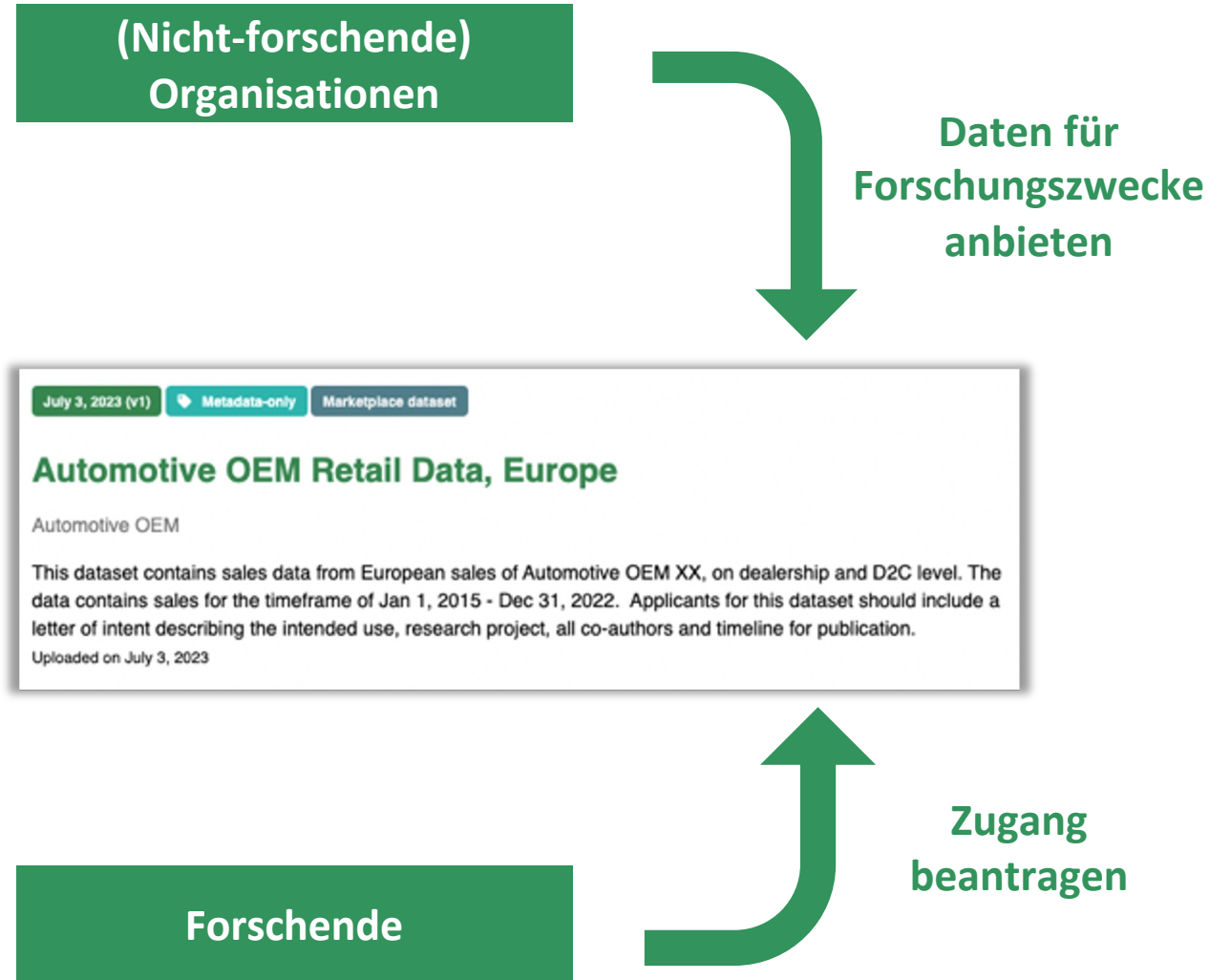
Uploaded on June 28, 2023

1. Daten in BERD: Data Marketplace

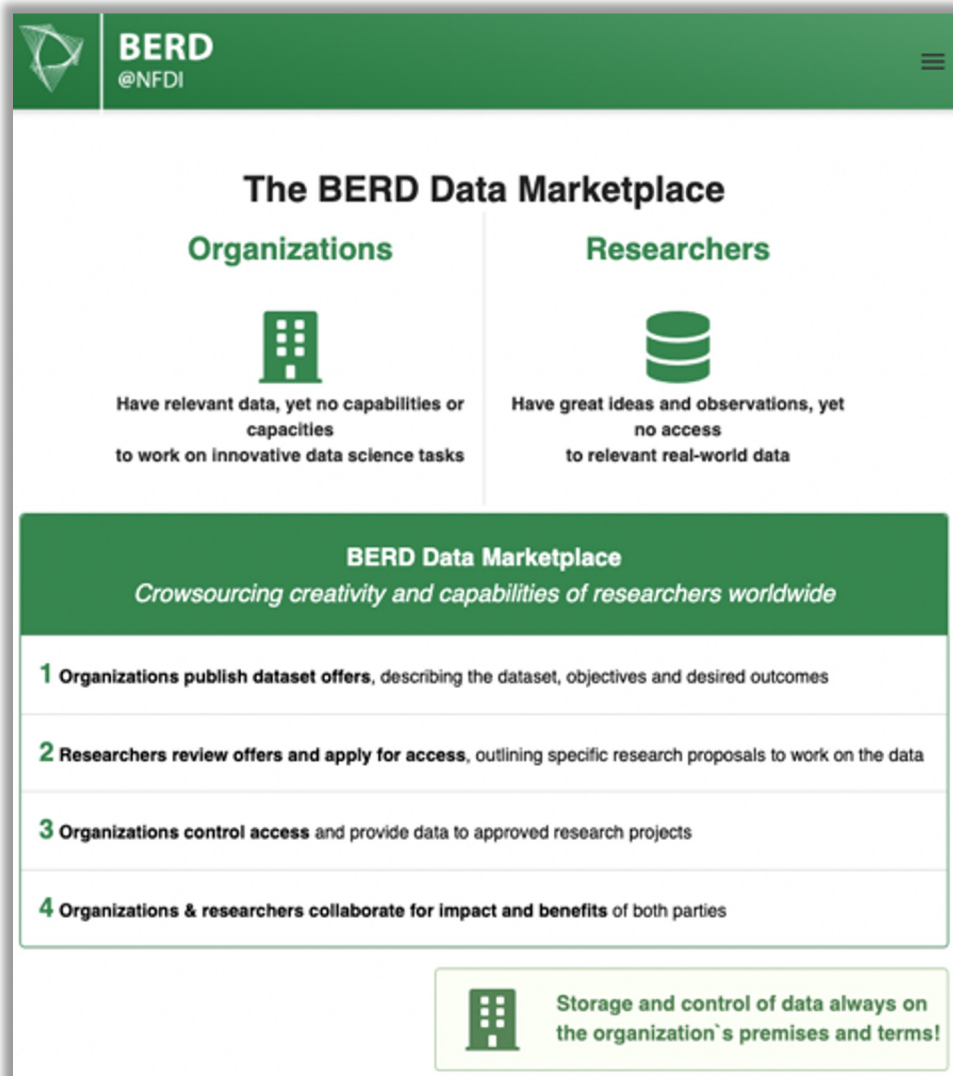
Ein zentrales Ziel:



den Austausch und die Zusammenarbeit zwischen Industrieorganisationen und Forschern bei relevanten datenbasierten Problemen zu ermöglichen und zu fördern



1. Daten in BERD: Data Marketplace




The screenshot shows the BERD Data Marketplace interface. At the top left is the BERD@NFDI logo. The main heading is "The BERD Data Marketplace". Below this, there are two columns: "Organizations" and "Researchers".

Organizations
Have relevant data, yet no capabilities or capacities to work on innovative data science tasks

Researchers
Have great ideas and observations, yet no access to relevant real-world data

BERD Data Marketplace
Crowdsourcing creativity and capabilities of researchers worldwide

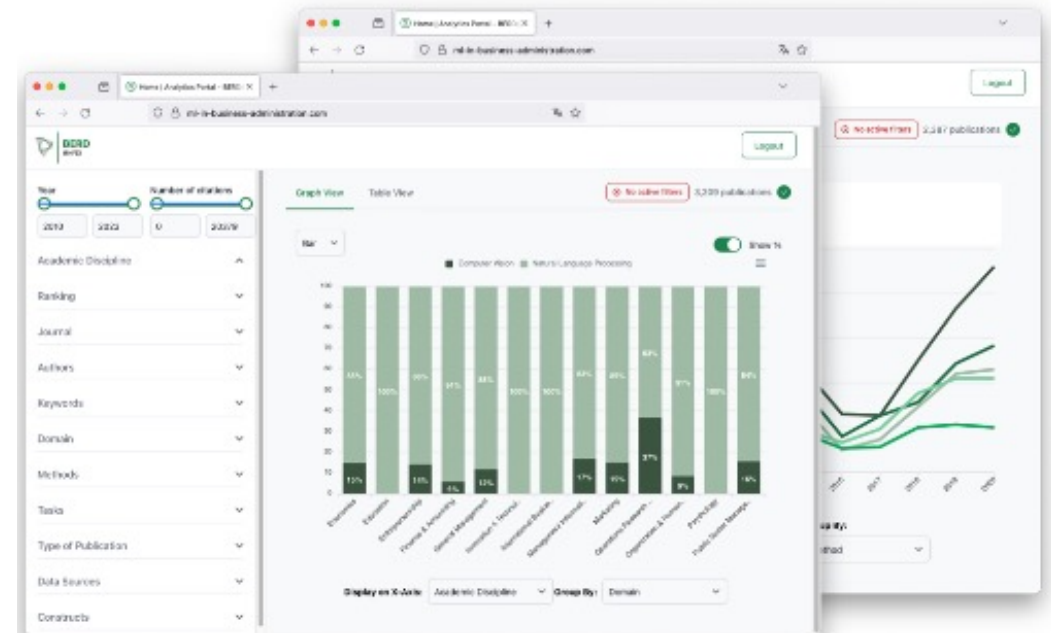
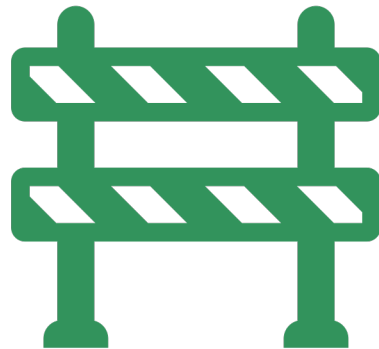
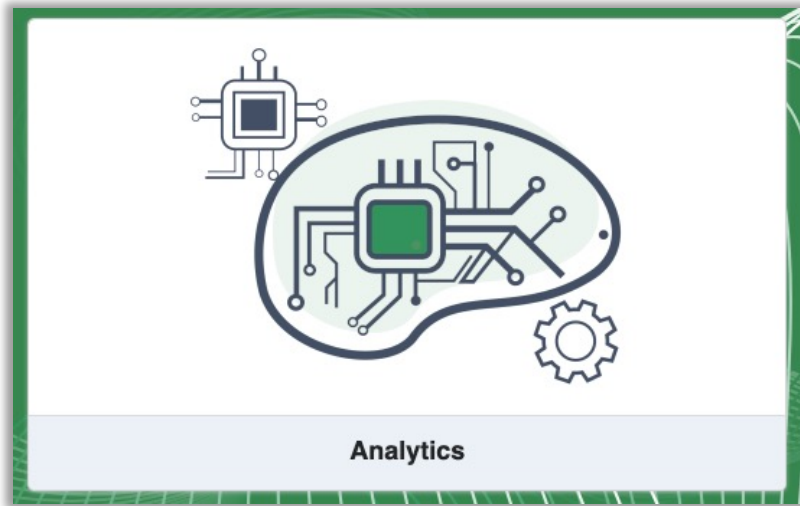
- 1 Organizations publish dataset offers**, describing the dataset, objectives and desired outcomes
- 2 Researchers review offers and apply for access**, outlining specific research proposals to work on the data
- 3 Organizations control access** and provide data to approved research projects
- 4 Organizations & researchers collaborate for impact and benefits** of both parties

 **Storage and control of data always on the organization's premises and terms!**



<https://berd-platform.de/for-companies>

2. Analytics in BERD: im Aufbau



Pre-Alpha Testing

3. Training & Education in BERD



- ✓ Erfahren Sie, wie Sie vom (ungenutzten) Potenzial unstrukturierter Daten in Wirtschaft, Ökonomie und den Sozialwissenschaften profitieren können



Forschende



Datenanbieter

Synchron



Asynchron



Flipped Classroom

Kurse & Workshops

Online



Vor Ort



Hybrid

3. Training & Education in BERD: BERD Academy



In-depth Workshops

Data Science with Telecommunication Data
With Göran Kauermann
Munich, June 12th 2024

Only a few seats left for this workshop!

Registration
Free Ticket
currently available: 3

Register

Ticketing powered by [prolix.io](#)

Informationswebinare

[IAB-SMART 1] The IAB-SMART Study: Collecting Behavioral Smartphone Sensor Data for Social Research
With Georg-Christoph Haas
Online, June 28th 2023, 12pm noon - 1:30pm CEST

Click here to download the slides of the webinar:

[IAB-SMART-Webinar-1_Study_Haas-G-C](#)

Do you own a smartphone? Probably. Smartphones have become an ubiquitous tool of our

Selbstlernmodule

[AutoML – Automated Machine Learning](#)

Do you want to become a Machine Learning (ML) developer? Then you may have encountered the challenge of designing functional and efficient (ML) pipelines. Sign up for this course now to be able to tackle this problem effectively in the future!

With Bernd Bischl, Frank Hutter, Janek Thomas, Lars Kotthoff, and Marius Lindauer

[Data Science with Python](#)

You want to learn Python for Data Science, but don't find the time to visit synchronous courses regularly? Register for this self-paced course and learn all you need to start with Python on your own schedule!

With Nicolas Heist and Sven Hertling

[Deep Learning for NLP \(DL4NLP\)](#)

Do you have an interest in Deep Learning algorithms and Natural Language Processing? This course covers all of the basics required for understanding current research, relevant state-of-the-art architectures and application areas.

With Andreas Stephan, Benjamin Roth, Christian Heumann, Hinrich Schütze, Leonie Weißweiler, and Matthias Aßenmacher

Mehrere bevorstehende Veranstaltungen!

[BERD@NFDI Research Symposium](#)

Register for the first BERD@NFDI Research Symposium for conference-style sessions and a young researchers' colloquium, fostering collaboration and exchange of information in research in business, economics, and related social sciences.

Mannheim, June 10th to 11th 2024

[Deep Learning with Humans-In-The-Loop: Active Learning for NLP](#)

The abundance of text data and the advent of powerful deep learning models has led to rapid advancements in Natural Language Processing. However, adapting models to specialized tasks still requires a nuanced understanding of the data. Join this hands-on workshop to dive into Active Learning!

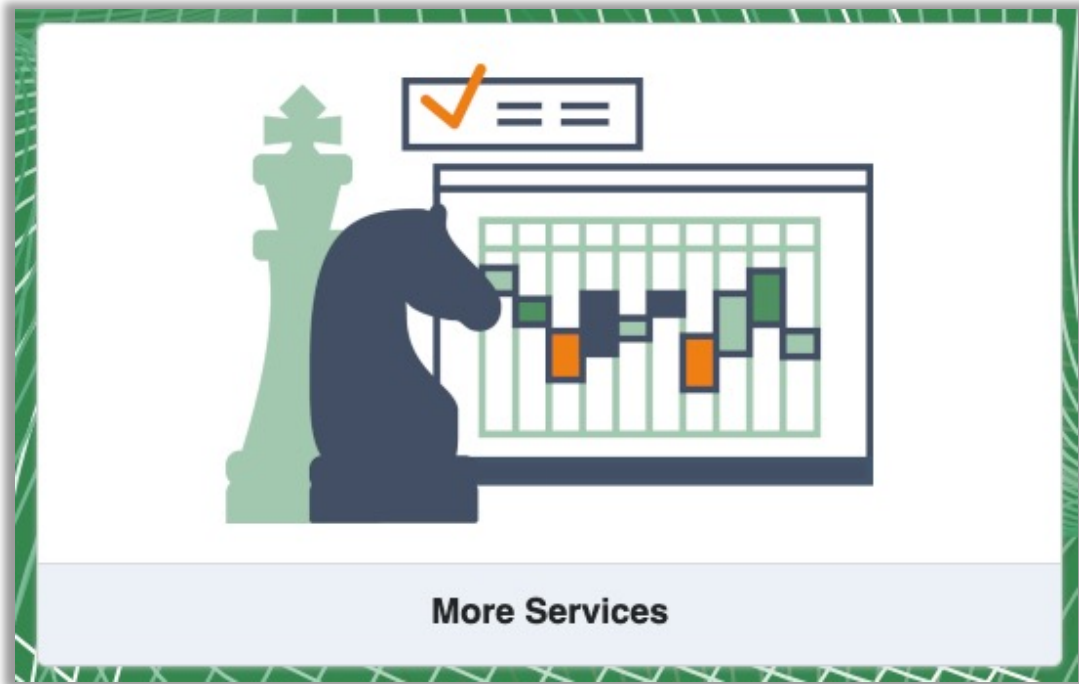
With Lukas Rauch
Kassel, July 4th 2024

[Turning PDFs into Research Data](#)

Do you ever feel that the data you need for your research is accessible but it's not in a convenient table? Join this online course to learn how to scrape data from PDFs and make it usable by using OCR and NLP!

With Jack Collins
Online, Starting Aug 27, 2024

4. Mehr Dienste bei BERD: Anwendungsfälle



- ✓ Erhalten Sie zusätzliche rechtliche Unterstützung zur korrekten Datenerhebung und Datenverarbeitung während Ihrer Forschung
- ✓ Erhalten Sie Anleitungen zur Digitalisierung Ihrer analogen Dokumente mithilfe modernster OCR-Algorithmen
- ✓ Zugriff auf Wissensgraphen zu deutschen Unternehmensdaten

Unsere interaktiven Systeme helfen Ihnen zu bestimmen, ob Sie der DSGVO oder Bundesgesetzen unterliegen und wie Sie die korrekte Einwilligung für die Datennutzung einholen können



Applicability of the GDPR

iVA1 shows you if the General Data Protection Regulation (GDPR) applies to your research data.

[Start here \(DE\)](#)

[Start here \(ENG\)](#)



Consent as legal basis

iVA2 helps you identify what you have to keep in mind to get a lawful consent for data processing.

[Start here \(DE\)](#)

[Start here \(ENG\)](#)



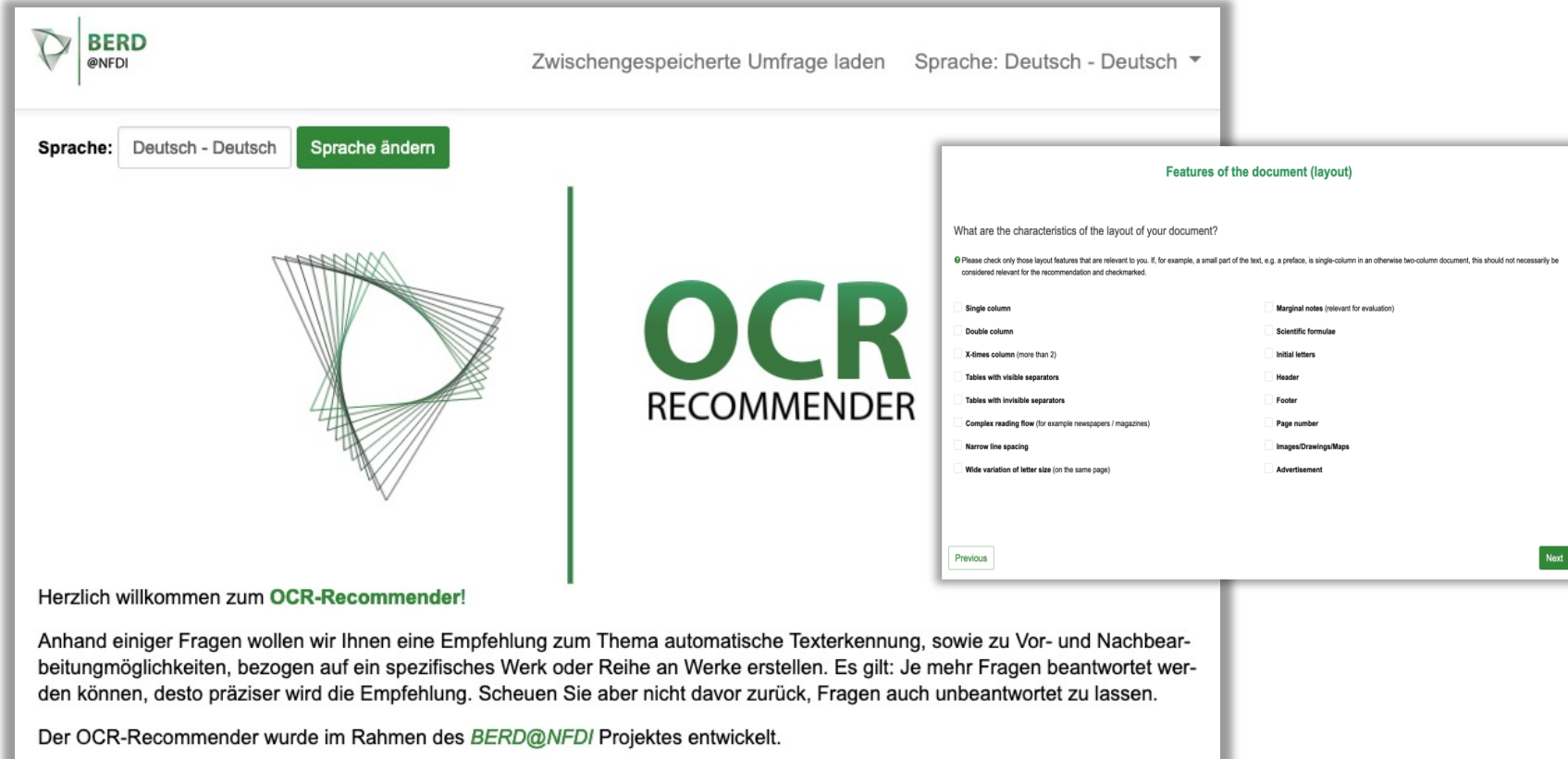
Research relevant legal bases

iVA3 guides you through the process of determining whether data processing in the context of your research project meets a federal or state legal basis.

[Start here \(DE\)](#)

4. Mehr Dienste bei BERD: OCR Recommender

Nutzen Sie unseren OCR Recommender um Empfehlungen für automatische Transkription, Texterkennung sowie Vor- und Nachbearbeitungsoptionen zu erhalten



The screenshot shows the OCR Recommender web interface. At the top left is the BERD@NFDI logo. To the right, there are links for 'Zwischengespeicherte Umfrage laden' and a language dropdown set to 'Sprache: Deutsch - Deutsch'. Below this, there is a language selector showing 'Deutsch - Deutsch' and a 'Sprache ändern' button. The main content area features a large green logo for 'OCR RECOMMENDER' next to a stylized green graphic of overlapping lines. Below the logo, a green banner reads 'Herzlich willkommen zum OCR-Recommender!'. The main text explains that the service provides recommendations for automatic transcription and editing based on user questions, and that more questions lead to more precise recommendations. It also states that the service was developed within the BERD@NFDI project. On the right side, a 'Features of the document (layout)' form is visible, asking for characteristics of the document layout with various checkboxes.

OCR RECOMMENDER

Herzlich willkommen zum **OCR-Recommender!**

Anhand einiger Fragen wollen wir Ihnen eine Empfehlung zum Thema automatische Texterkennung, sowie zu Vor- und Nachbearbeitungsmöglichkeiten, bezogen auf ein spezifisches Werk oder Reihe an Werke erstellen. Es gilt: Je mehr Fragen beantwortet werden können, desto präziser wird die Empfehlung. Scheuen Sie aber nicht davor zurück, Fragen auch unbeantwortet zu lassen.

Der OCR-Recommender wurde im Rahmen des **BERD@NFDI** Projektes entwickelt.

Features of the document (layout)

What are the characteristics of the layout of your document?

Please check only those layout features that are relevant to you. If, for example, a small part of the text, e.g. a preface, is single-column in an otherwise two-column document, this should not necessarily be considered relevant for the recommendation and checkmarked.

- Single column
- Double column
- X-times column (more than 2)
- Tables with visible separators
- Tables with invisible separators
- Complex reading flow (for example newspapers / magazines)
- Narrow line spacing
- Wide variation of letter size (on the same page)
- Marginal notes (relevant for evaluation)
- Scientific formulae
- Initial letters
- Header
- Footer
- Page number
- Images/Drawings/Maps
- Advertisement

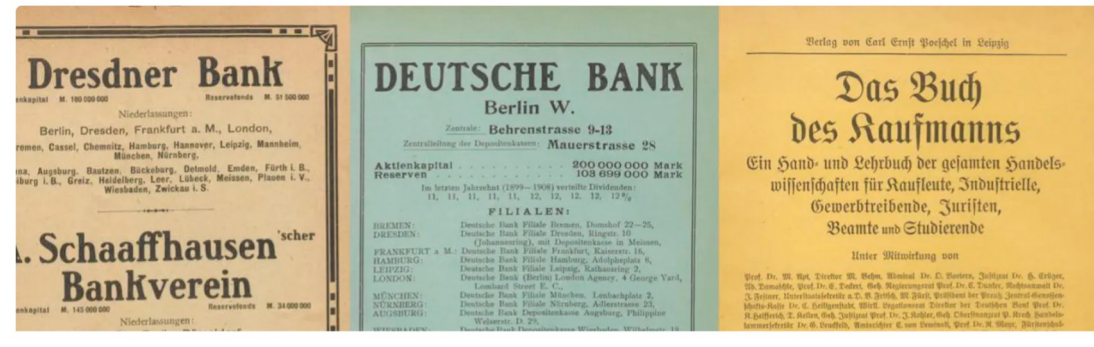
Previous Next

4. Mehr Dienste bei BERD: BERD Knowledge Graphs

Zugriff auf unsere Wissensgraph-Infrastruktur zu deutschen Unternehmensdaten, basierend auf verknüpften Daten aus verschiedenen digitalisierten Büchern

Aktienführer Knowledge Graph

[View source](#) [View history](#)

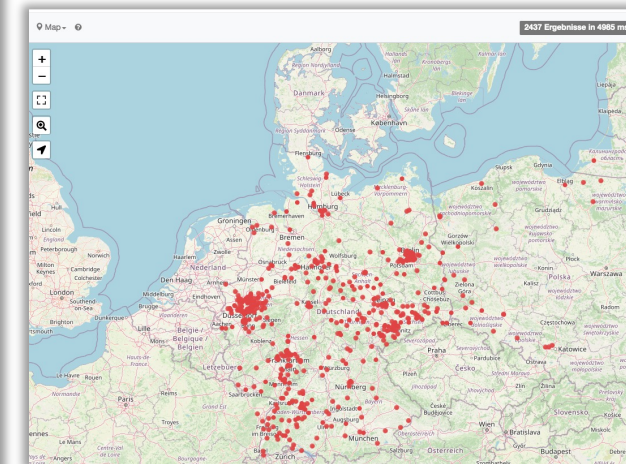
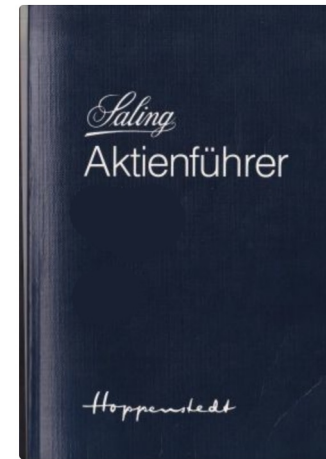


Overview

The Aktienführer (AKF) Knowledge Graph is free and open. It contains structured (meta)data for the German listed stock companies from the Hoppenstedt-Aktienführer from 1956 to 2018. The metadata includes official company names with time ranges, company identifiers (ISIN and WKN) and Aktienführer IDs.

- [official name](#)
 - [start time](#)
 - [end time](#)
- [ISIN](#)
- [WKN](#)
- [Aktienführer ID](#)
 - [point in time](#)

We enriched metadata via entity linking with other databases: [Credireform](#), [Bureau van Dijk Orbis](#), [GND](#) and [Wikidata](#). The knowledge graph contains [342](#) external identifiers representing various databases, including the valuable company databases: [Crefo number](#), [BvD ID](#), [Orbis ID](#), [Handelsregisternummer](#), [PM20 folder ID](#),

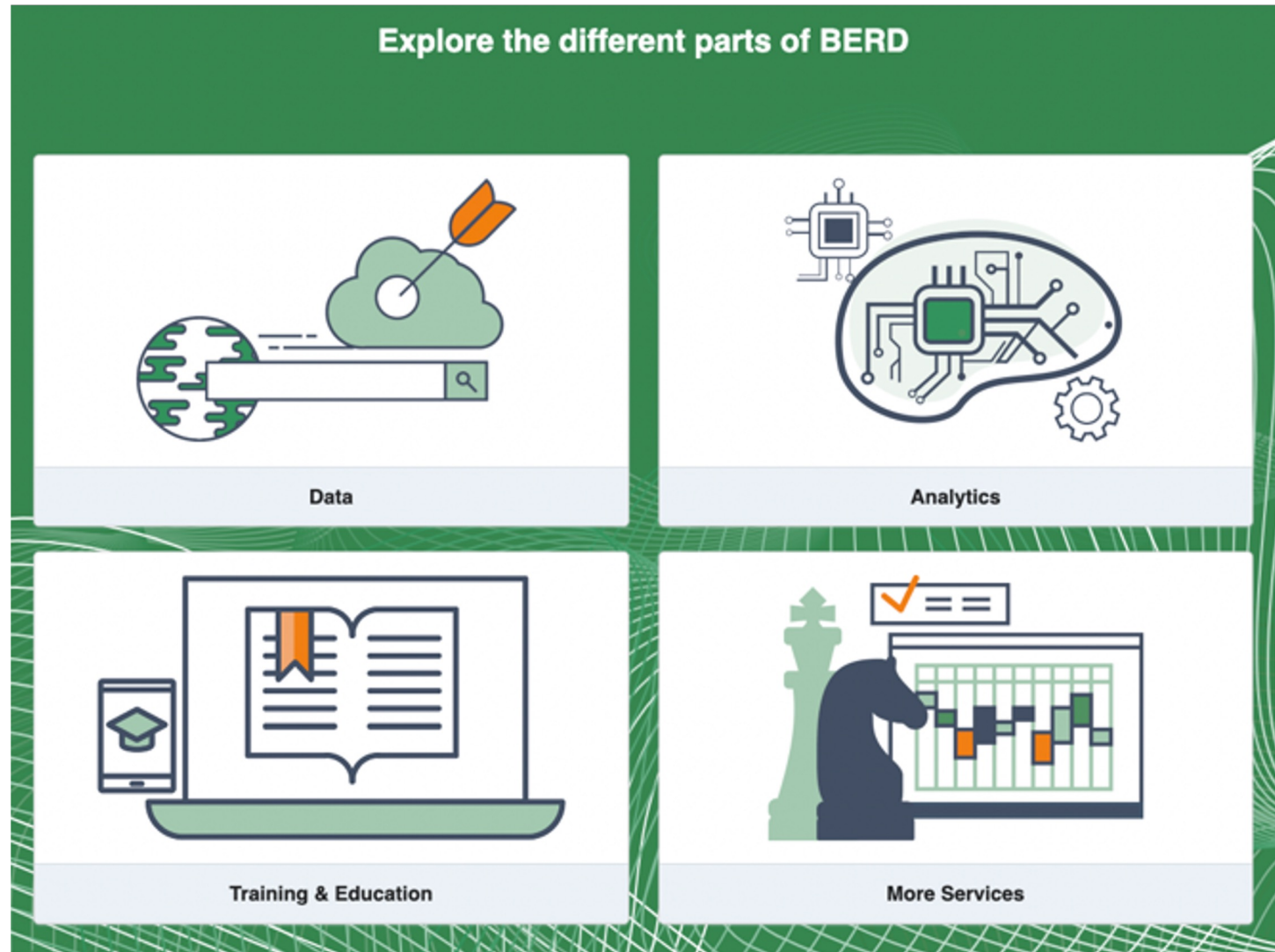


1

Veröffentlichen, Teilen und Finden von (unstrukturierten) Daten

3

Lernmaterialien und Veranstaltungen zum Thema (unstrukturierte) Daten



2

Modernste Methoden und Algorithmen für die Arbeit mit (unstrukturierten) Daten

4

Umfassende Werkzeuge und Anleitungen für die Arbeit mit (unstrukturierten) Daten

Fragen, Kommentare und Wünsche sind willkommen!

BERD-Plattform: <https://berd-platform.de>

Projekt: <https://www.berd-nfdi.de>

Newsletter: <https://www.berd-nfdi.de/newsletter-en>

LinkedIn: <https://www.linkedin.com/school/berd-nfdi>

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 460037581