



NFDI for Business, Economic and Related Data

Georg Licht

ZEW – Leibniz Zentrum für Europäische Wirtschaftsforschung
Mannheim

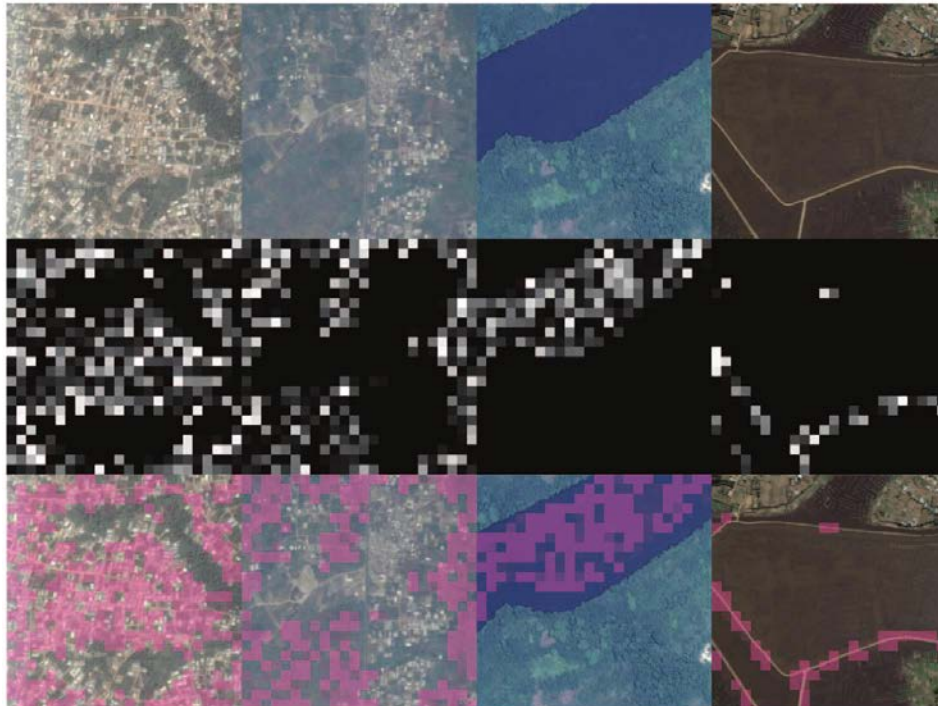
Forschungsinfrastrukturen in BWL und Finance

VHB 2020

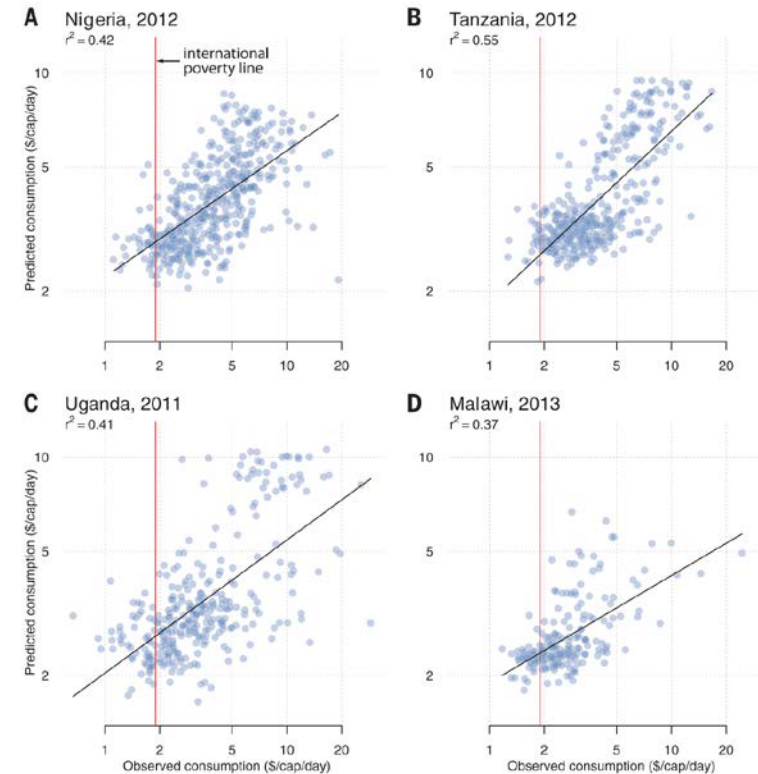
Johann Wolfgang Goethe-Universität Frankfurt am Main

18. März 2020

Automated Feature Extraction



Predicted and Actual Consumption Rates Based on Satellite Image Features



Non-verbale Kommunikation von CEOs verrät strategische Ausrichtung des Unternehmens

ANGER



Guler Sabanci (0:25:41), 0.85
Sabanci Holdings, Turkey

CONTEMPT



Ritu Kumar (0:45:19), 0.71
Ritika Private Ltd., India

DISGUST



Mallika Sarabhai (0:19:41), 0.66
Darpana Academy, India

FEAR



Ela Bhatt (0:60:00), 0.43
SEWA, India



Shyam Benegal (0:29:53), 0.97
Filmmaker, India



Andre Esteves (0:50:21), 0.90
BTG Pactual, Brazil



Merrill Fernando (0:54:34), 0.63
MJF Group, Sri Lanka



Fadi Ghandour (0:52:49), 0.57
Aramex, U.A.E.

HAPPINESS



Eva Muraya (0:03:28), 1.00
BSD Group, Kenya

SADNESS



Seema Aziz (0:33:30), 0.95
SEFAM / Care Foundation, Pakistan

SURPRISE



Zia Mody (0:13:51), 0.99
AZB & Partners, India

NEUTRAL



Mallika Sarabhai (1:19:22), 1.00
Darpana Academy, India



Erling Lorentzen (0:22:17), 1.00
Aracruz Celulose, Brazil



Jaime Zobel de Ayala II (0:24:40), 0.97
Ayala Corporation, Philippines



Cem Boyner (0:01:37), 0.99
Boyner Holdings, Turkey

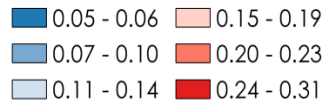


Yusuf Hamied (0:33:06), 1.00
Cipla, India

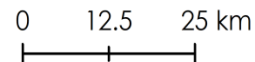
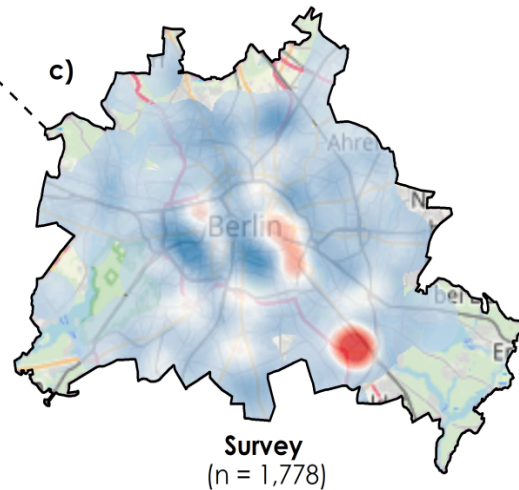
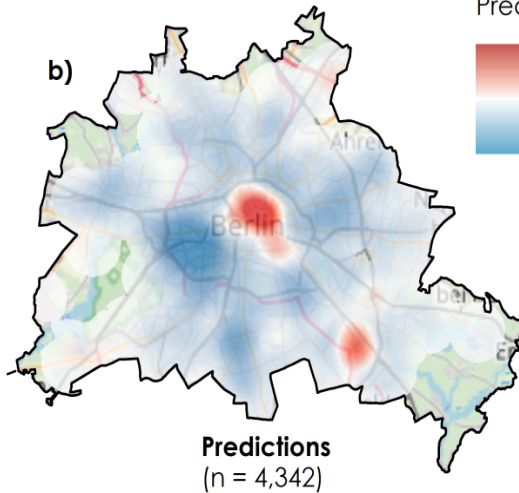
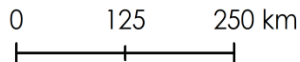
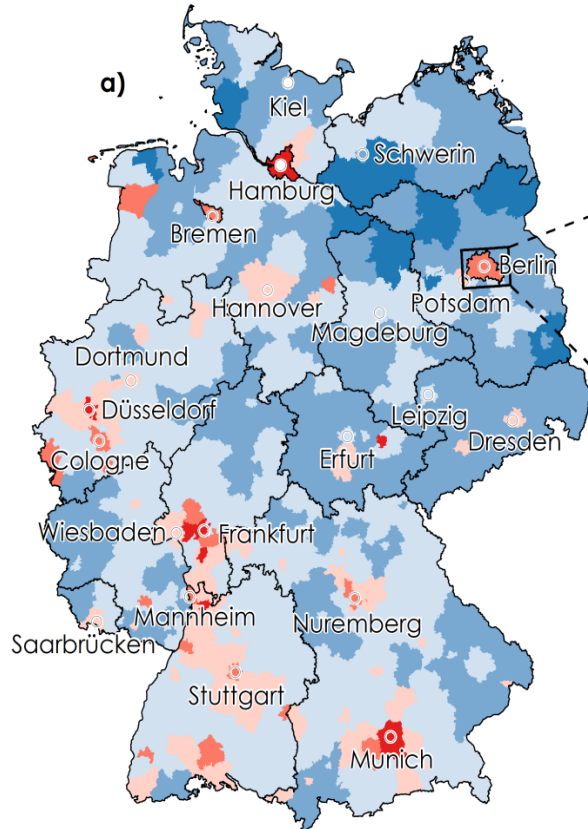
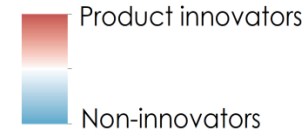
Eine dramatischere Kommunikation eines CEOs (Anger, Disgust, Happiness, Sadness) ist verknüpft mit weniger Unternehmensrestrukturierungen (M&A)

Indikatoren durch Analyse von Unternehmens-Web-Seiten **BERD** NFDI

Ratio of product innovator firms:



Predominant type of firm:



Indicator Validation:
Regional Innovation indicators

Correlations:

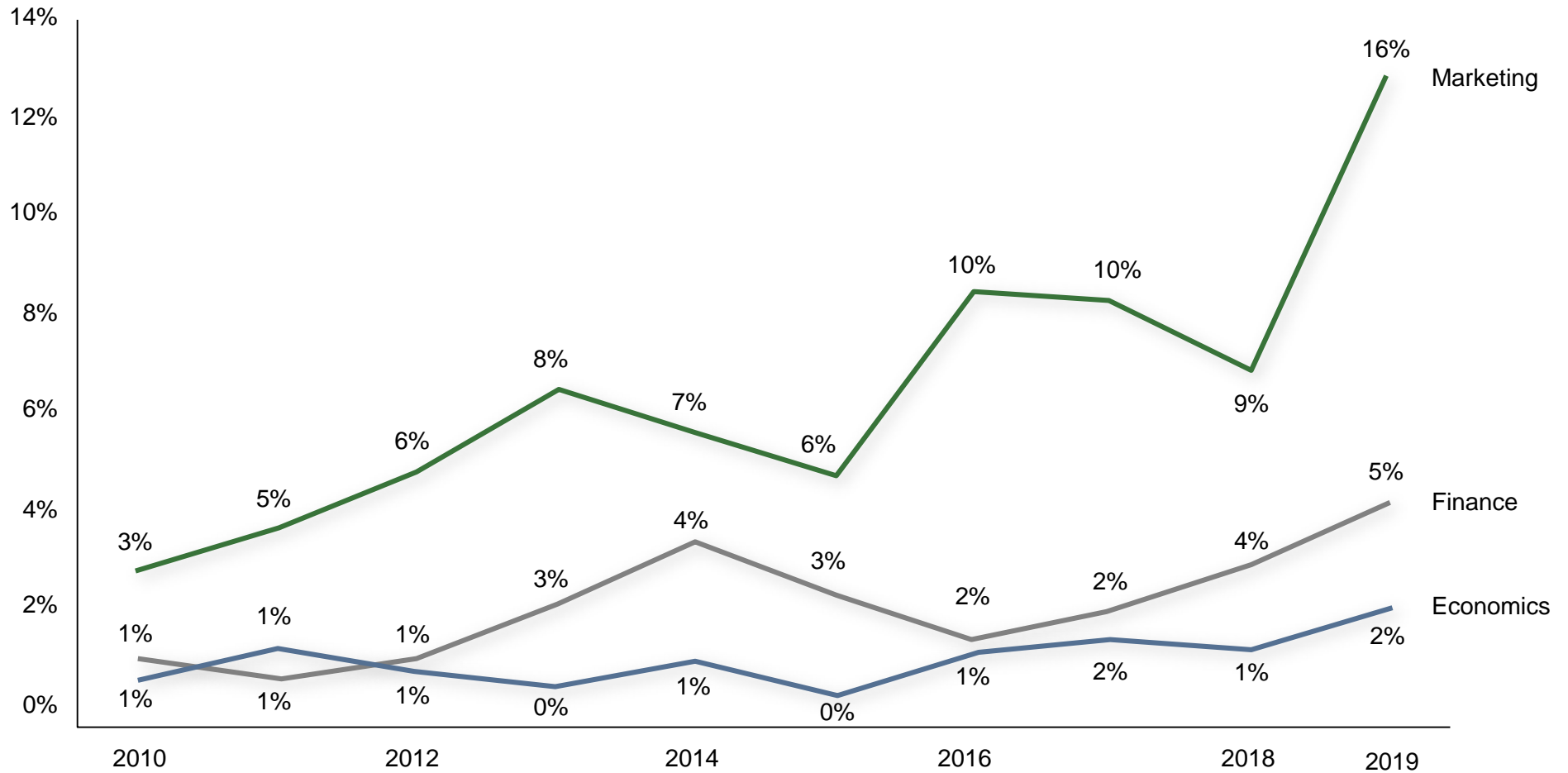
- 0.62 population density
- 0.72 regional R&D staff
- 0.67 regional high-tech patent applications

Visual match to microgeographic pattern of special innovation survey in Berlin

Quelle: Kinne/Lenz (2019) Predicting Innovating Firms using Web Mining and Deep Learning, ZEW DP 12-2019

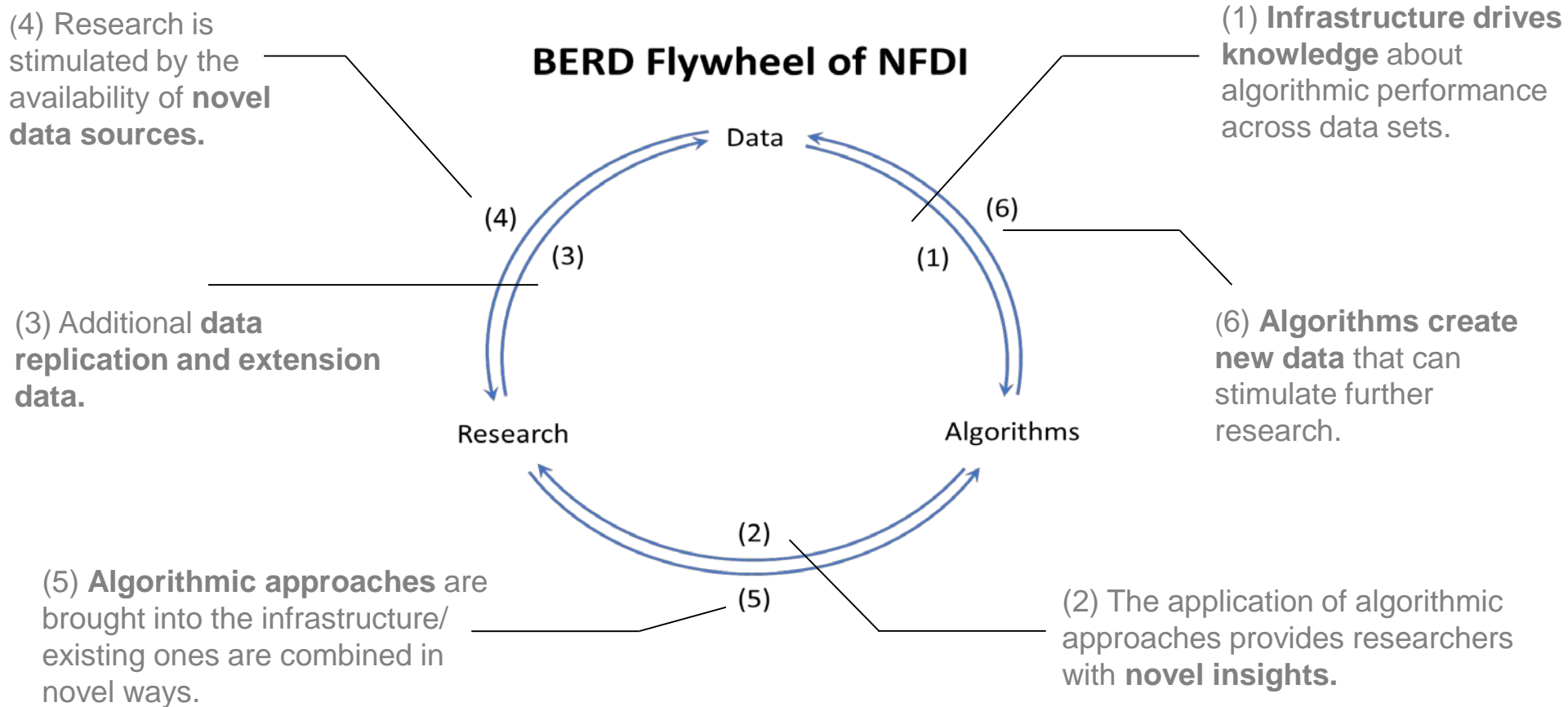
Unstrukturierte Daten und KI-Algorithmen werden relevanter und populärer

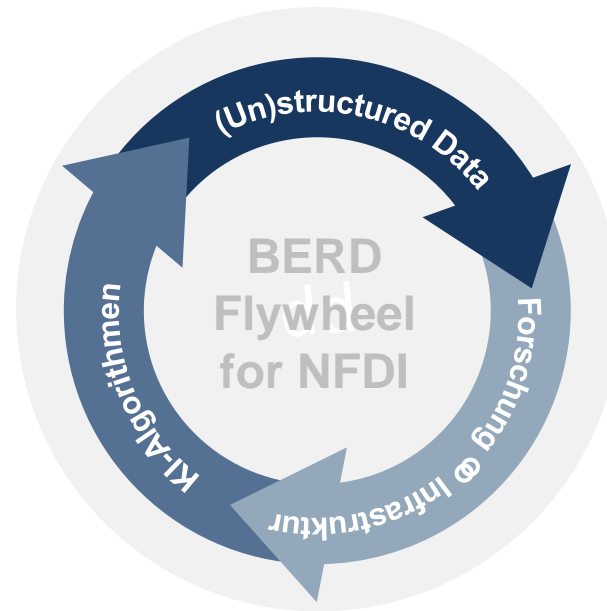
Publikationsaufkommen basierend auf unstrukturierten Daten und KI-Algorithmen*



* Top journals in VHB Jourqual 3.1 category A+ (VHB online org.)

An open infrastructure that inventories both datasets together with algorithms





Strukturierte & Unstrukturierte Daten

- Offene, forscherezentrierte Plattform für strukturierte und unstrukturierte Daten.
- Community-Prinzip um die Transparenz bzgl. Daten und Software zu erhöhen.

Künstliche Intelligenz Algorithmen

- Infrastruktur für das integrierte Management von (un-)strukturierten Daten und KI-Methoden.
- Cloud-basierte Lösung mit Hochleistungsrecherkapazität.
- Beratung für Forschende hinsichtlich der Auswahl passender Methoden und Algorithmen.

Konsortium erfahrener Wissenschaftler/-innen und Infrastrukturexperten

- Großes internationales Netzwerk und Expertise im Aufbau und Betreiben großer Dateninfrastrukturen.
- Ergänzt NFDI-Initiative um den fehlenden Aspekt "unstrukturierte Daten".



Prof. Kreuter
D. Müller



Dr. Gehrlein



K. Bein
A. Pfister



Prof. Höhle



Prof. Stahl



Prof.
Kauermann



Prof. Fischer



Prof.
Heitmann



Dr. Borst



Prof.
Tochtermann



Dr. Licht



Leibniz-Institut
für ökologische
Raumentwicklung



GfeW



GUG



Standardization and harmonization of terminologies

BERD@NFDI will contribute to the standardization and harmonization of terminologies as it brings in-depth knowledge of the vocabularies and ontologies used in business/economics.



Metadata interoperability

For the services of BERD@NFDI, a way must be found to bring requirements of a particular user community in line with the requirements of interdisciplinary re-use and general guidelines.



Quality assurance of data and algorithms

BERD@NFDI will focus on the development of standards for quality assurance of data such as fake detection and algorithms such as ethical AI.



User involvement

User involvement and user-centered design engineering play an essential role in the whole process of developing, pre- and post-implementing the services of BERD@NFDI.



Legal aspects

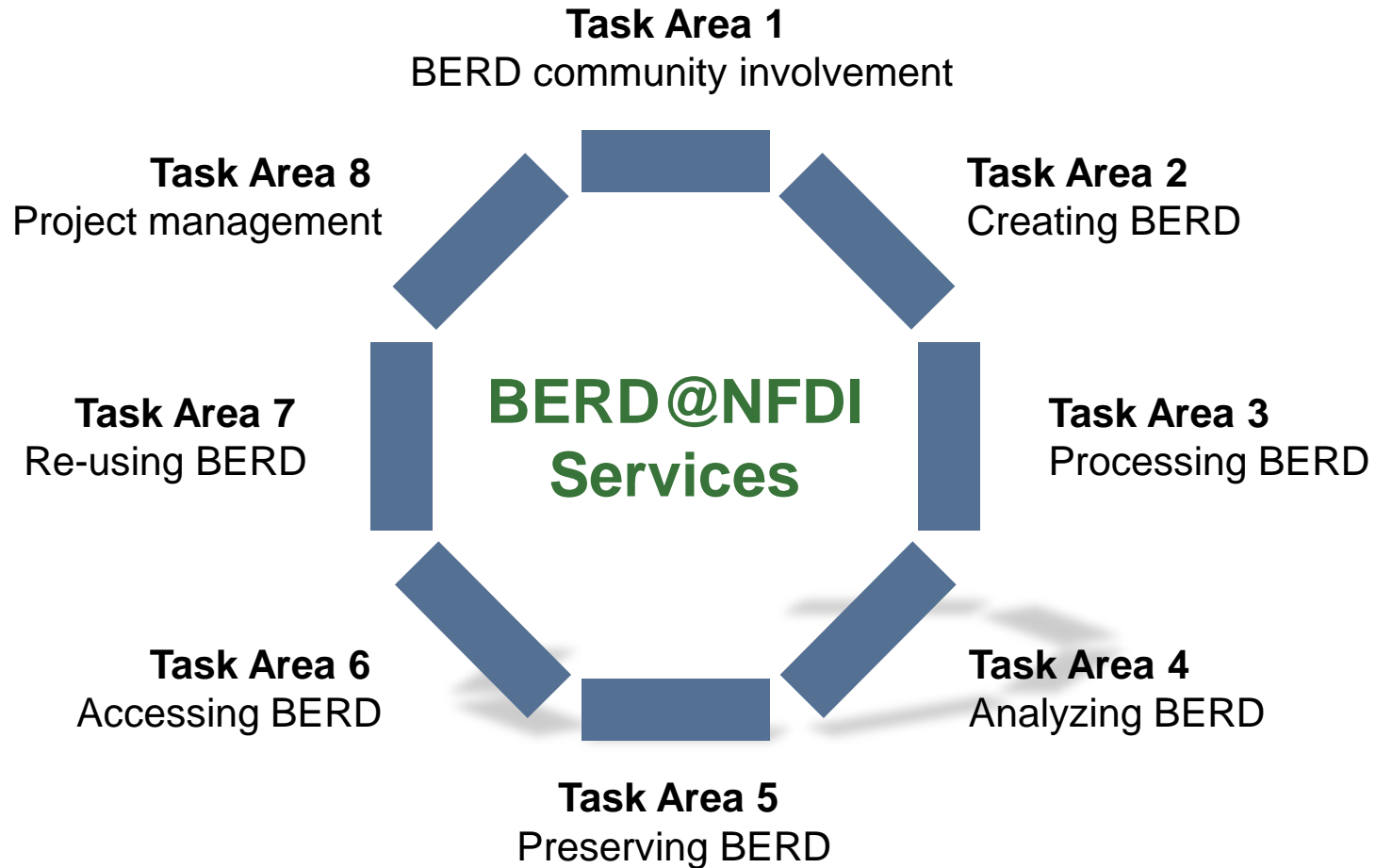
BERD@NFDI will put special emphasis on the legal questions that arise in the context of sharing unstructured data that are relevant for business, economics and related fields.



Teaching and education

BERD@NFDI can build on extensive experience of several participants in the conception/ implementation of online courses and workshops about data processing and analysis.

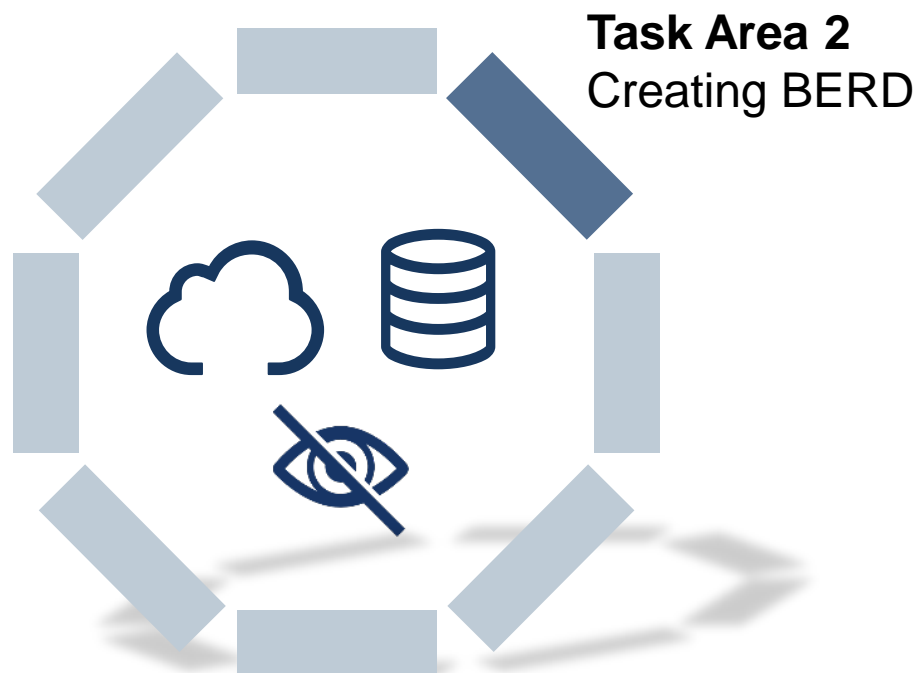
Overview of our Work Program



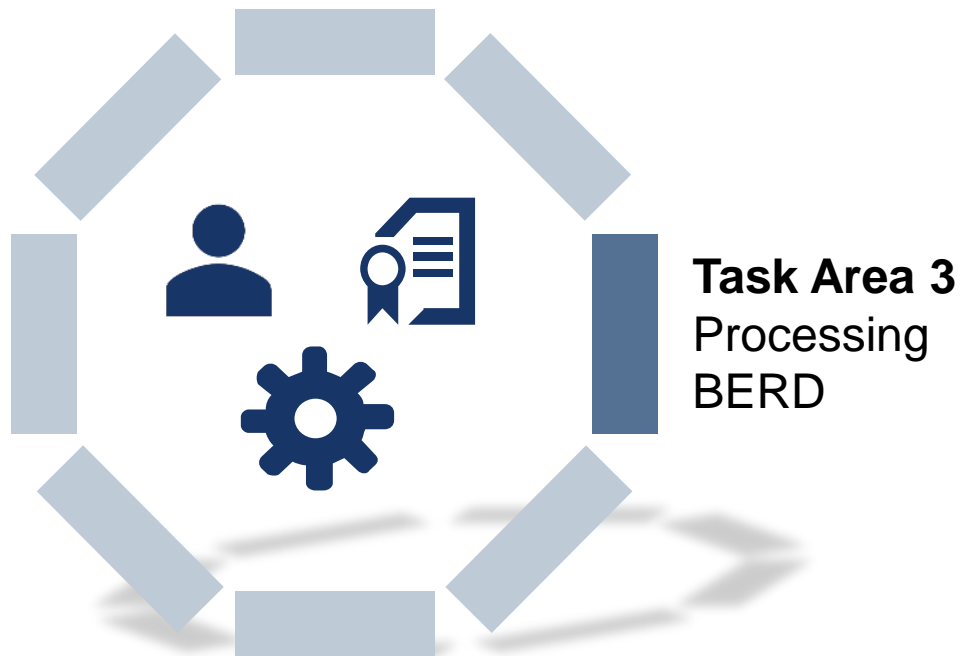
Task Area 1 BERD community involvement



- User-centered design in order to...
 - **Identify quality indicators** for developing infrastructure services
 - React upon users' feedback and **changing user needs**
 - Implement the **FAIR principles** from a **user's perspective**
- **Readily available KPIs based on user involvement**

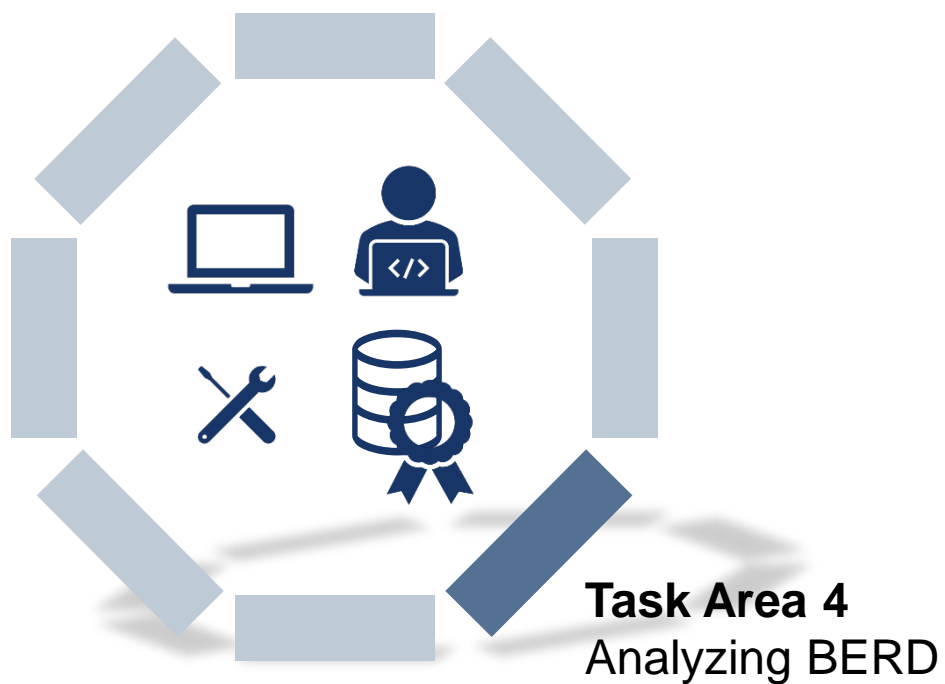


- **Secure environment** for data, (meta)data, and algorithms
- **Storage and use of**
 - raw data
 - derived data
 - results
 - algorithms
- **Legal support for data ingestion**
- **GDPR compliance and privacy protection**



Support **development of standards, best practices, and guidelines** for ...

- **Processing, documentation and normalization** of unstructured data
- **Quality assurance** for data and cross-disciplinary metadata
- **Anonymization** of data
- Conversion of **historical data sources**

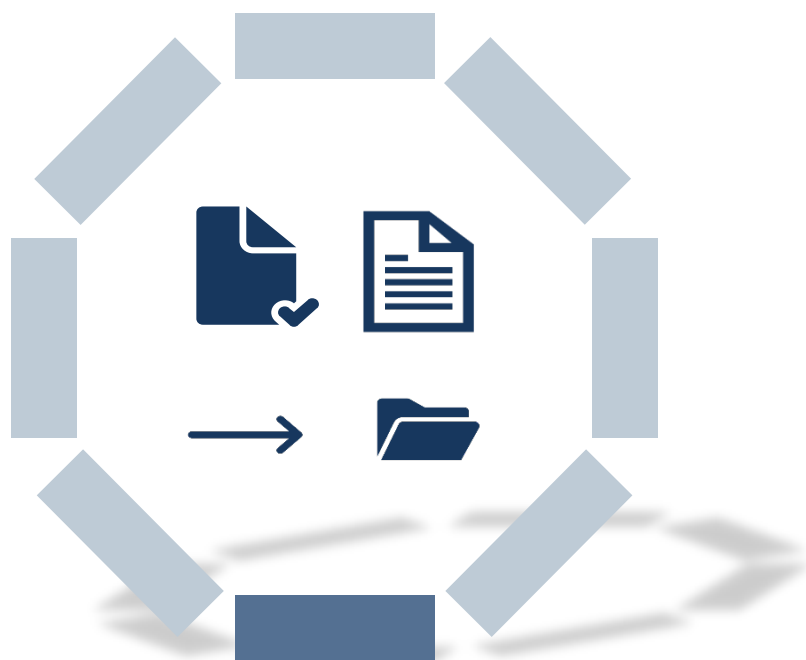


Data analysis tools

- Improving analytical efficiency, capability and accuracy
- Access to algorithms facilitates
 - improvements
 - continuous updating
 - transfer learning

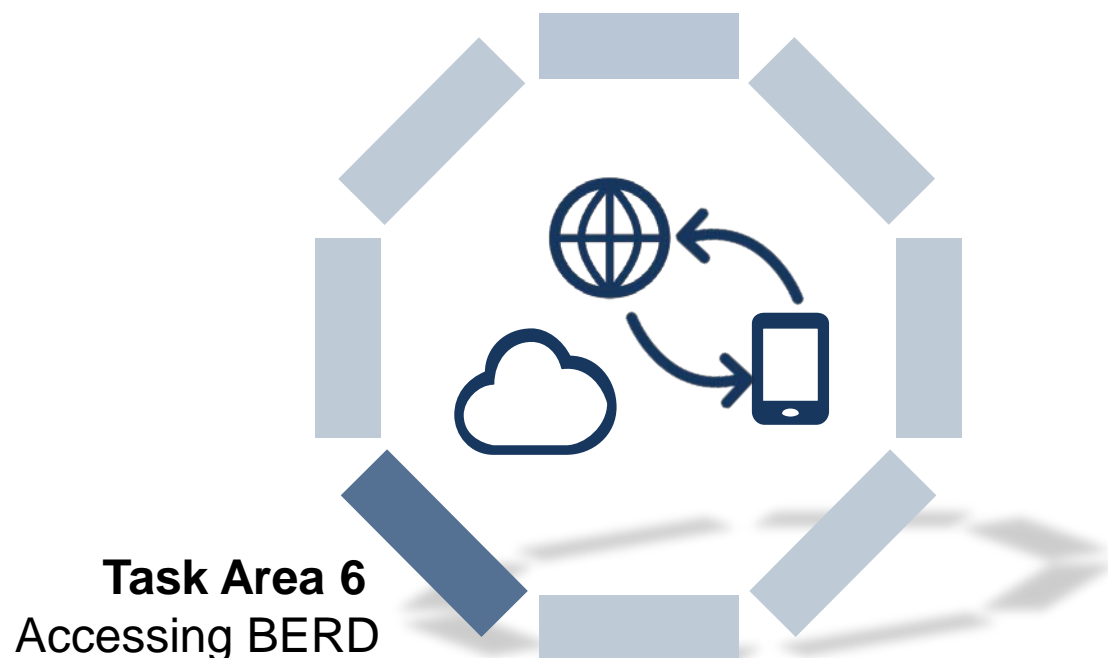
Enrichment and linking of data

- Deep indexing
- Ontologies, semantic search



Task Area 5
Preserving BERD

- **Preservation** of (derived) data **paired with**
 - preprocessing code
 - analytic code
 - trained algorithms
- Implement checks for **data integrity**
- • Standardize and add metadata for stored digital objects
- Develop **migration strategy**



- **Discovery interface** for data, metadata, code, text documentation
- **Identity management**
- **Persistent identifiers**
- **Remote access**
- **Metadata component** that represents resources uniformly at the appropriate granularity level

Task Area 7
Re-using
BERD



- **Education and tailored training programs**
- Support / data **stewardship service**
- (Automated) **feedback loops** from research applications
- Exchange with EOSC, IPSDS etc.

FAIR Data Principles⁹



FINDABLE



ACCESSIBLE



INTEROPERABLE



RE-USABLE

- Not a specific standard, but formulates high-level rules that may help publishers in determining to what extent their meta(data) are compliant with the FAIR principles
- No binary distinction between FAIR or non-FAIR data, but maturity levels between 0% and 100% fair data

0 %

100 %

Vielen Dank für Ihre Aufmerksamkeit