Science in the Age of AI: Is Your Research Ready for the Digital Revolution?

Leibniz ScienceCampus
Digital Transformation of Research

Sonja Schimmler

Harald Sack, Anna Jacyszyn, Felix Bach

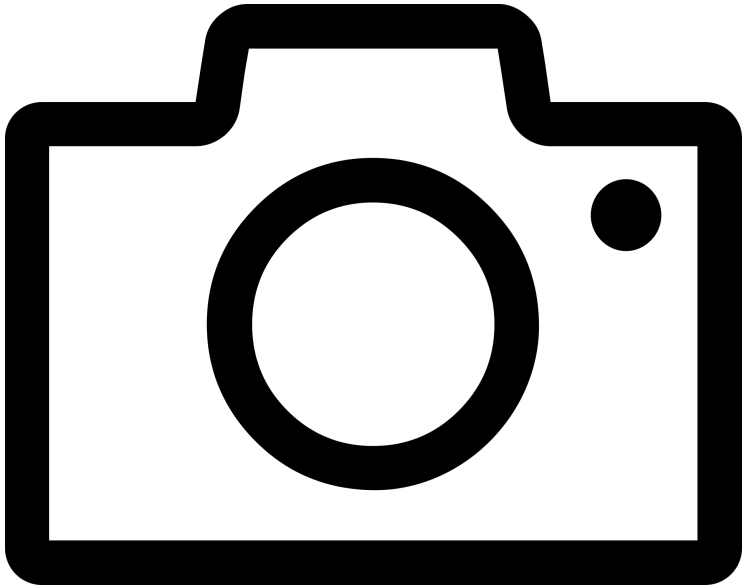
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

KIT
Karlsruhe Institute of Technology

funded by Leibniz Competition

Interdisciplinary Colloquium on Digitalisation of Research, FIZ Karlsruhe, 17 July 2025

Photos and recording



Pixabay, ste_phania

YouTube

www.youtube.com/@DiTraRe

Agenda

Science in the Age of AI: Is Your Research Ready for the Digital Revolution?

Sonja Schimmler
Fraunhofer FOKUS, Technische Universität Berlin

Anna Jacyszyn
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure

Harald Sack
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure
KIT AIFB

Felix Bach
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure

1. Welcome and introduction
2. DiTraRe - who are we?
3. Is Your Research Ready for the Digital Revolution?
4. Data Infrastructures as a Foundation for AI Projects
5. Discussion session
6. Lunch @ Casino

Interdisciplinary Colloquium
on Digitalisation of Research

We want to use the colloquium as a **regular** possibility to highlight and **discuss** the manyfold aspects of digitalisation of research with experts and interested scientists.

- Monthly event: every first Thursday at **11:00** (~~10:30~~) with rare exceptions
- Hybrid format
- Different on-site locations to foster interdisciplinary exchange

www.ditrare.de/en/ditrare-colloquium

Schedule

www.ditrare.de/en/ditrare-colloquium

Slot available 2 October.

07.08.2025 **11:00–12:00**	TBA (AI in Biomedical Engineering)	**Till Keller**, Department of Cardiology and Angiology, Medical Center, University of Freiburg, Bad Krozingen + Justus Liebig University Giessen, Giessen
04.09.2025 **11:00–12:00**	TBA (PUNCH4NFDI LLM-based chatbot for creating workflows for physicists and astronomers)	**Arman Khalatyan**, Leibniz Institute for Astrophysics Potsdam (AIP)
02.10.2025 **11:00–12:00**	Slot available	**Slot available**, suggest your speaker in discussion forum (top of the webpage) or ditrare@fiz-karlsruhe.de
06.11.2025 **11:30–12:30**	Machine Learning in Professional Soccer Abstract	**Daniel Memmert**, German Sport University Cologne, Institute of Exercise Training and Sport Informatics; **Katja Keller**, Karlsruhe Institute of Technology, Institute of Sports and Sports Science (KIT IfSS); **Claudia Niessner**, Karlsruhe Institute of Technology, Institute of Sports and Sports Science (KIT IfSS)
04.12.2025 **11:00–12:00**	TBA	**Stefan Dietze**, GESIS Leibniz Institute for the Social Sciences + Heinrich Heine University Düsseldorf

Discussion forum

Continue discussions after the event:

- Introduce yourself
- Meet others
- Exchange ideas
- Collaborate
- Discuss the talk
- Suggest speakers
- …

www.ditrare.de/en

Newsletter

www.ditrare.de/en/newsletter

Forum

- General discussion
- The last colloquium
- Upcoming events

www.ditrare.de/en/forum

Leibniz ScienceCampus *Digital Transformation of Research* (DiTraRe)

DiTraRe
Leibniz ScienceCampus *Digital Transformation of Research*

- Growth core to **establish new research branch**.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

+

KIT
Karlsruhe Institute of Technology

- Planned as a **4+4 years** project (start: September 2023).
- Funded by the Leibniz Association + FIZ KA + KIT.
- Analyse the process of **digitalisation of research**.
- **Multilevel interdisciplinary approach**. → We start with 4 specific use cases.

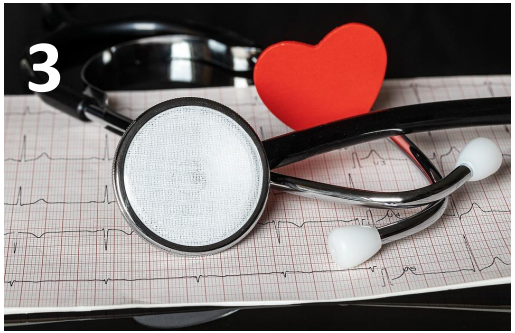
DiTraRe Use Cases

Sensitive Data in Sports Science

*KIT Institute of **Sports and Sports Research***

Chemotion Electronic Lab Notebook

*KIT Institute of **Biological and Chemical Systems***



AI in Biomedical Engineering

*KIT Institute of **Biomedical Engineering***

Publication of Large Datasets

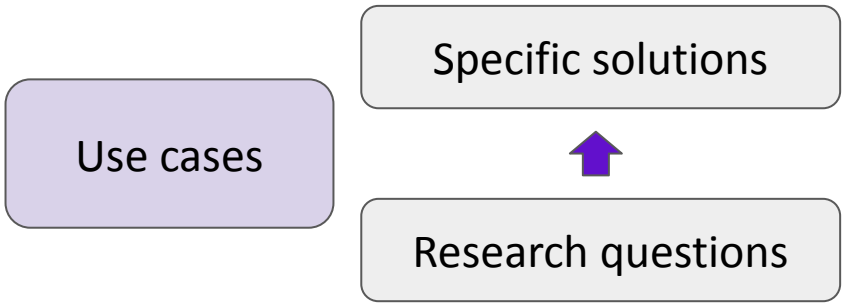
*KIT Institutes of **Meteorology and Climate Research***

From Use Cases to Research Clusters

Use cases

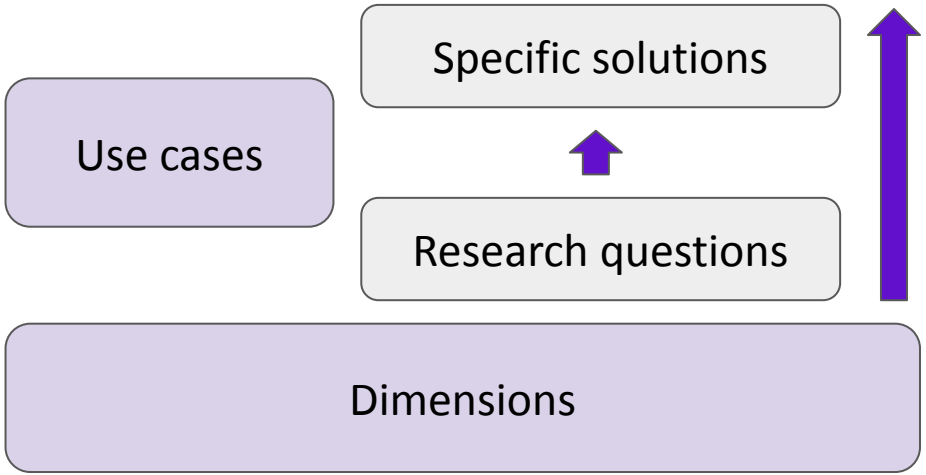
- Impact of digitalisation and AI on science
- Impact of digitalisation and AI on reception of science in society
- Overarching challenges of a technical, legal, and ethical nature

From Use Cases to Research Clusters



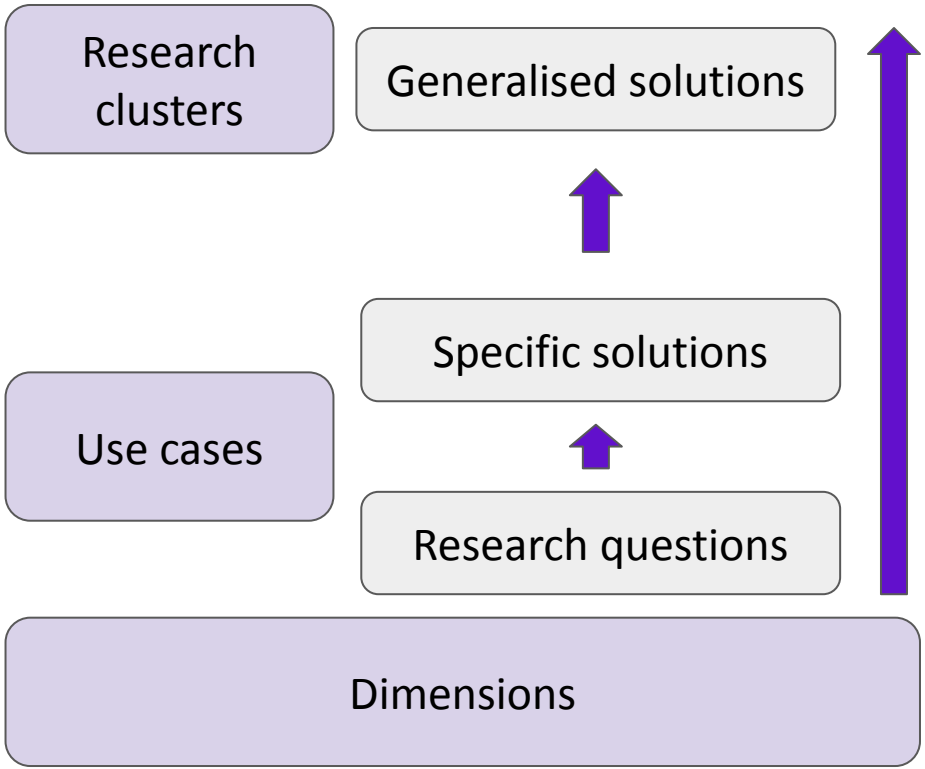
- Impact of digitalisation and AI on science
- Impact of digitalisation and AI on reception of science in society
- Overarching challenges of a technical, legal, and ethical nature

From Use Cases to Research Clusters



- Impact of digitalisation and AI on science
- Impact of digitalisation and AI on reception of science in society
- Overarching challenges of a technical, legal, and ethical nature

From Use Cases to Research Clusters



- Impact of digitalisation and AI on science
- Impact of digitalisation and AI on reception of science in society
- Overarching challenges of a technical, legal, and ethical nature

DiTraRe Dimensions

A. Reflection and Resonance
a dialogue between research and society, interactive process

B. Exploration and Knowledge Organisation
applied AI: represent, organise, and manage domain specific and procedural knowledge

C. Legal and Ethical Challenges
data ethics, data protection, copyright and data law, Task Force “Security”

D. Tools and Processes
digital tools tailored precisely to the needs of researchers

Planned outcomes

- Specific solutions to use cases problems
- Generalised solutions, i.e. best practices, esp. in ethical, legal, security context
- Intensify interactions with society
- Identify new potential threats which come with digitalisation
- Spin-off projects
- Create a networking environment: **collaborate with others**!

Leibniz ScienceCampus
Digital Transformation of Research

Science in the Age of AI: Is Your Research Ready for the Digital Revolution?

- paradigm shift towards **digital** data collection, -management and analysis
- **high quality data** is the basis for good science and ML/AI
- research infrastructures needed for all **stages of the research cycle**
- research **processes are changing** through digitalisation
- **AI** may lead to even more **radical changes** - if we look at the **speed of development**
- How can we **reflect** on the way **how science is conducted** at this speed?
  - … in all of DiTraRe's **dimensions**, starting with our **use cases**?
  - What are **effects** the **on science system, epistemology, trust** in science by the public?
  - what role plays a **human in the loop** if we can't **scale** our understanding as quickly as AI?

Short history of digitalisation in science

Leibniz ScienceCampus
Digital Transformation of Research

Philosophical Foundations of AI

- **350 BCE - Aristotle**:
  a. philosophical foundation of AI; introduced “psyche” (soul) and symbolic reasoning, unique to humans
  b. Developed early theories of perception, reasoning, and consciousness as “actuality”

- **17th c. Leibniz** asked “what kind of machine is the mind?” and developed logical calculus

- **20th c.**: **Wiener, Shannon, Church, Turing, von Neumann** formalised thought via control, information, and computation theories

- **Mid-20th c.**: Cognitive architectures by **Minsky, Piaget, Simon** establish models of learning and reasoning

- **Modern AI:** Intelligence seen as computational model-building; evolution as learning algorithm

Short history of digitalisation in science

Leibniz ScienceCampus
Digital Transformation of Research

1703 **Leibniz develops the Logical Machine, binary number system.** Publishes *Explication de l'Arithmétique Binaire*

1936 **Alan Turing** publishes "*On Computable Numbers*", introducing the **Turing machine**

1938 - 1977 **Introduction of computers and application in Research**

- 1938-1941 First computers by Konrad Zuse: Z1 (1938, binary mechanical), Z3 (1941, first digital computer).
- 1945 Von Neumann Architecture
- 1975 First Personal Computers: Altair 8800 (1975) Intel's 8080 processor, Apple I (1976), Commodore PET (1977), …
- 1977 First computational DNA sequencing

1970s-1980s **Networking & the Internet and early internet protocols.**

- 1969 ARPANET goes online, connecting four U.S. universities (UCLA, Stanford, UCSB, University of Utah).
- 1983 Internet (TCP/IP protocol) is standardised.
- 1989 WWW: Tim Berners-Lee launches the World Wide Web, revolutionising information access (web browsers…)

1990s **Open Science movement -** enhances transparency and accessibility. e.g.: 1991 arXiv.org starts

- 2001 Open Knowledge, platforms like Wikipedia democratise knowledge and enable collaborative content creation.

2000s **Digitisation of Libraries starts**

- 2002 Internet Archive's Million Book Project starts (still relevant: Websites and data disappear - e.g. now in the US)
- 2004 Google and major libraries begin digitising books at scale (Project Gutenberg already started in 1971)

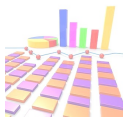
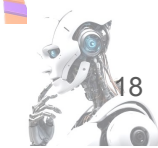
2000-2010s **Data-Driven Research advances**

- 2000 Cloud computing and Big Data infrastructures transform data storage and analysis.
- 2008 CERN's Large Hadron Collider produces vast datasets.
- 2009 Jim Gray's Fourth Paradigm describes the era of data-intensive scientific discovery.
- 2014 Research Tools like Electronic Lab Notebooks (ELNs) improve structured data acquisition and reproducibility.

recent years **Artificial Intelligence (AI) & Machine Learning (ML) become central to scientific data analysis**





Today's Symposium Talk by *Sonja Schimmler*

Sonja Schimmler
Fraunhofer FOKUS, Technische Universität Berlin

Leibniz ScienceCampus
Digital Transformation of Research

- Prof. Dr. Sonja Schimmler is **Professor for Research Data Infrastructures** at **TU Berlin**, holding the Chair of Research Data Infrastructure.
- **Member** of the **Scientific Advisory Board (SAB)** of the Leibniz ScienceCampus **DiTraRe**.
- Key leadership **roles in NFDI**:
  - Spokesperson of **NFDI4DataScience** (data science & AI infrastructures).
  - Co-Spokesperson of **NFDI4Cat** (catalysis data management).
  - Co-Spokesperson of **Base4NFDI** (core foundational services).
- Research focuses on **digitalisation and openness of scientific workflows**, semantic web, linked data, AI-supported data management, and scalable infrastructures.
- Plays a crucial role in shaping interoperable, trustworthy, and sustainable data infrastructures foundational for AI-driven research across disciplines.

Data Infrastructures as a Foundation for AI Projects

Prof. Dr. Sonja Schimmler

Fraunhofer FOKUS and TU Berlin

Data Infrastructures as a Foundation for AI Projects

Background

About us

Groups

- Fraunhofer: Research Group Lead
- TU Berlin: Guest Professor

Research

- Digitalisation and Opening up of Science
- Research Data Infrastructures

About us

Projects

- Weizenbaum Institute
  - Group Digitalisation and Opening up of Science
- Berlin University Alliance (BUA)
  - Project Berlin Open Science Portal
- National Research Data Infrastructure (NFDI)
  - Consortium NFDI4DataScience
  - Consortium Base4NFDI
- Data Competency Center (DCC)
  - QUADRIGA

Data Infrastructures as a Foundation for AI Projects

Research Focus

Research Focus

Modern research on an internationally competitive level …

… relies on more and more data

… is no longer possible without digital support

Live Sciences	Natural Sciences
Engineering Sciences	Humanities and Social Sciences

This leads to a paradigm shift

Research Focus

Scientific progress will only be furthered, if …

… data becomes available at large scale

… data is linked and machine-interpretable

… specificities of the different disciplines are very carefully taken into account

Live Sciences	Natural Sciences
Engineering Sciences	Humanities and Social Sciences

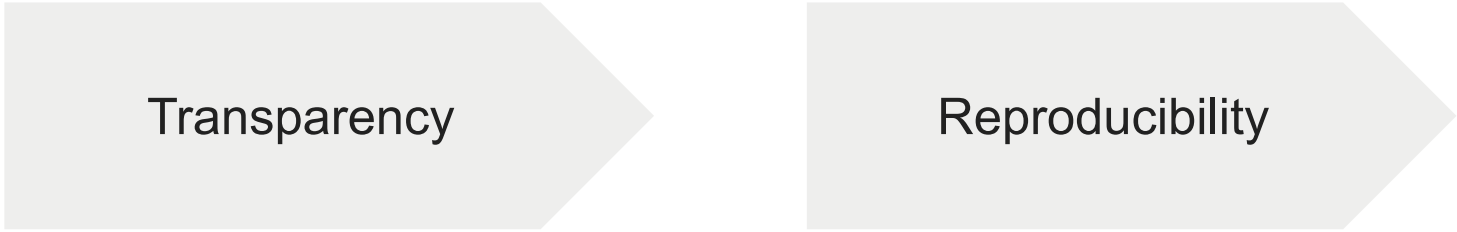
This leads to a paradigm shift

Data Infrastructures as a Foundation for AI Projects

Data and Data Infrastructures

Good Scientific Practice

- Increase **transparency and reproducibility** in research



Good Scientific Practice

- Foster **open science**

Open Data

Open Research Processes

Open Access

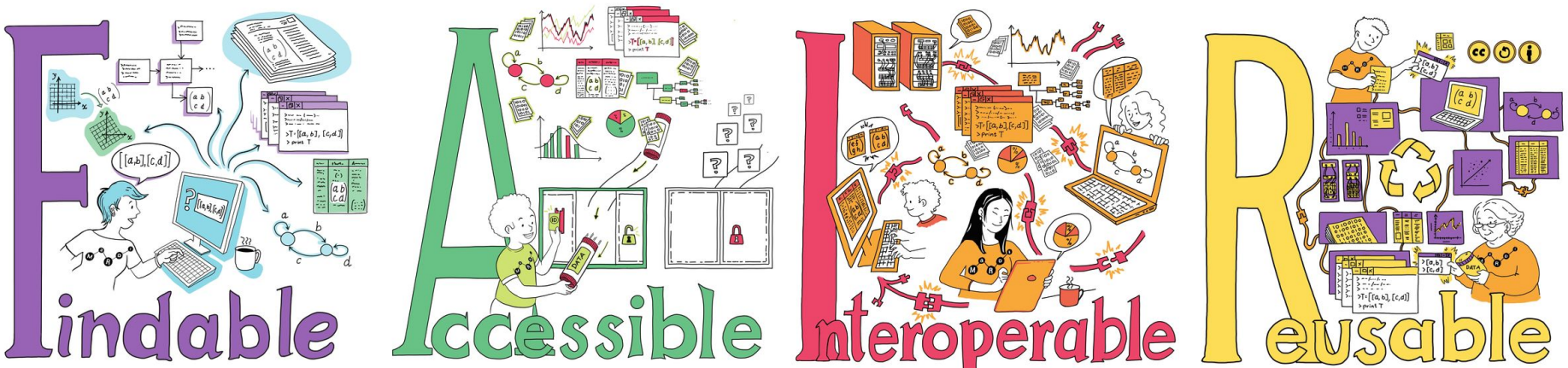
Open Source

Open Educational Resources

Good Scientific Practice

- Consider **FAIR** principles

  (findable, accessible, interoperable, reusable)



by Constanza Rojas-Molina

Good Scientific Practice

- Consider **CARE** principles

  (collective benefit, authority to control, responsibility, ethics)

from gida-global.org

Data Infrastructures as a Foundation for AI Projects

AI

Evolution in AI

LLM

- Language model is used in isolation
- Lacks factual knowledge

Retrieval Augmented Generation

- Retrieves unstructured information to augment answers

Knowledge Augmented Generation

- Access to structured information sources, such as databases, knowledge graphs

Evolution in AI

LLM

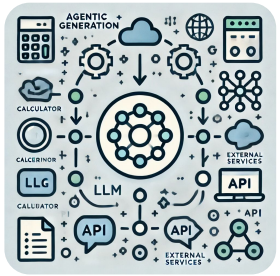
- Language model is used in isolation
- Lacks factual knowledge

Retrieval Augmented Generation

- Retrieves unstructured information to augment answers

Knowledge Augmented Generation

- Access to structured information sources, such as databases, knowledge graphs



Agentic Generation

- Use of external tools such as reasoners, APIs, computation services

Artificial General Intelligence

- Self-reflection, goal setting

Challenges in AI

Flood of research artefacts
paper, data, models, scripts/code, …

Changing environment
data science & AI hype

Challenges in AI

Flood of research artefacts
paper, data, models, scripts/code, …

Changing environment
data science & AI hype

Bias
of training and evaluation data

Trustworthiness and **responsibility** of AI

Challenges in AI

Flood of research artefacts
paper, data, models, scripts/code, …

Changing environment
data science & AI hype

Bias
of training and evaluation data

Trustworthiness and **responsibility** of AI

Good scientific practice
transparency & reproducibility

Digital scientific practice
tools & services

Data Infrastructures meet AI

Generative AI

- Is there a difference in the use of LLMs in the different communities?
- How can we support researchers and the general public that use LLMs?

Data Infrastructures meet AI

Generative AI

- Is there a difference in the use of LLMs in the different communities?
- How can we support researchers and the general public that use LLMs?

Infrastructures and Services for Generative AI

- How to adapt infrastructures to support researchers that use LLMs?

Data Competencies for Generative AI

- What new competencies are needed by researchers that use LLMs?

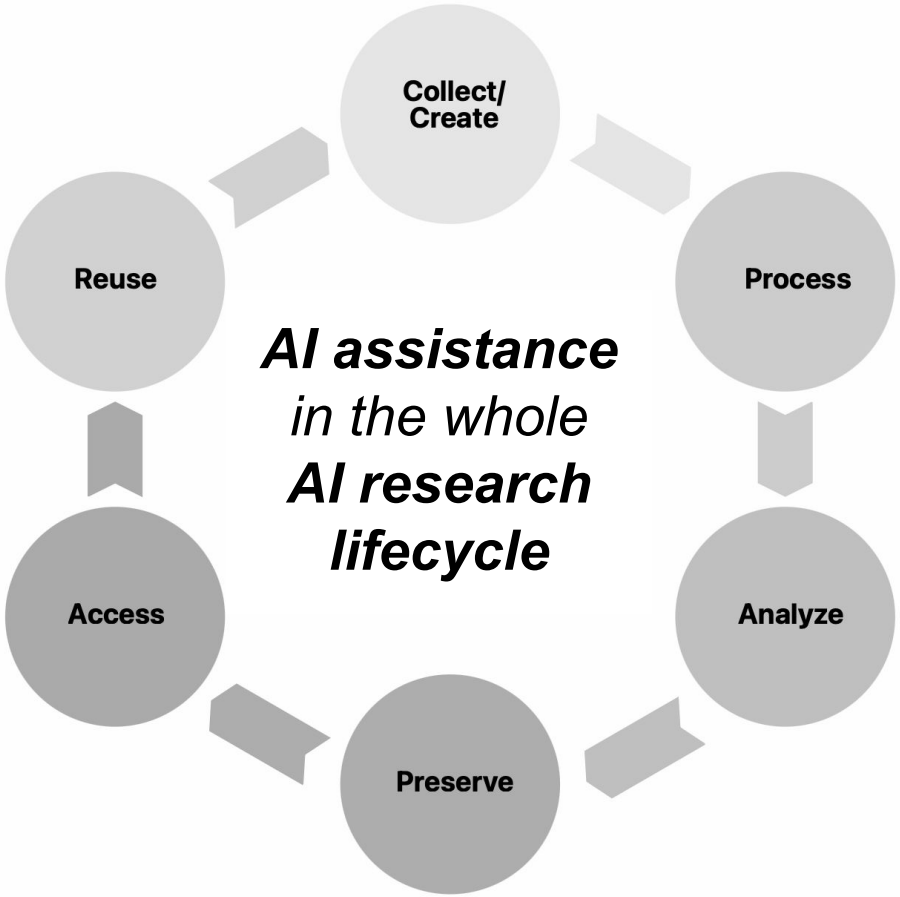
Data Infrastructures meet AI

AI for Data Infrastructures

Data Infrastructures for AI

Data Infrastructures meet AI

AI for Data Infrastructures

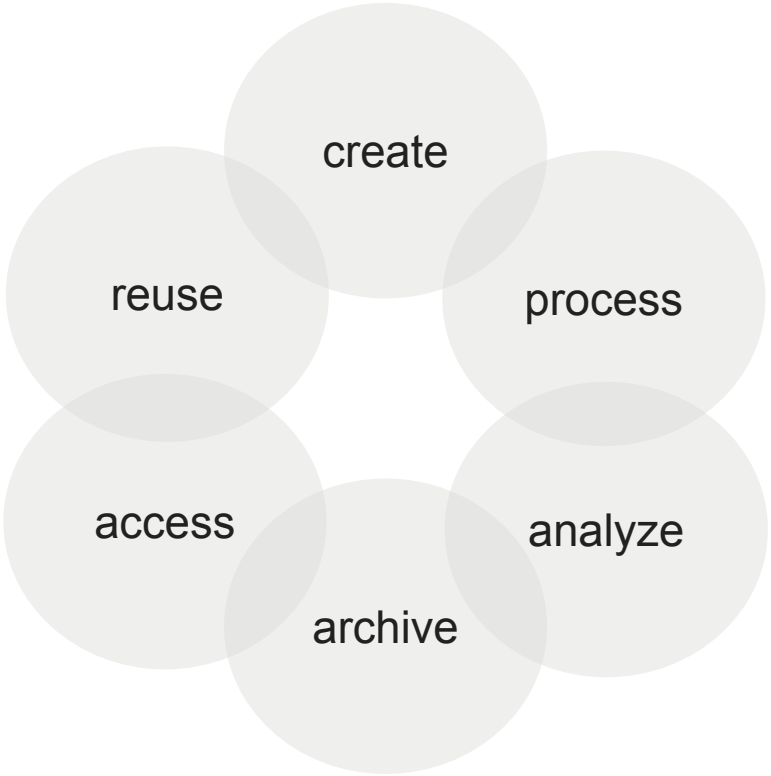


Data Infrastructures for AI

Data Infrastructures as a Foundation for AI Projects

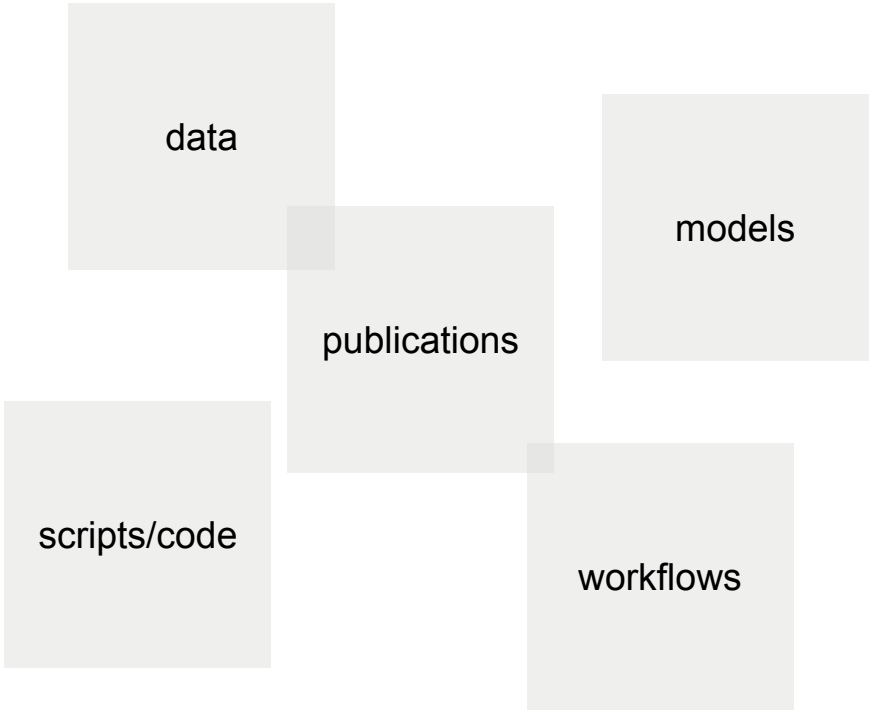
Data Infrastructures

Data Infrastructures



- The whole **research lifecycle** should be considered

Data Infrastructures

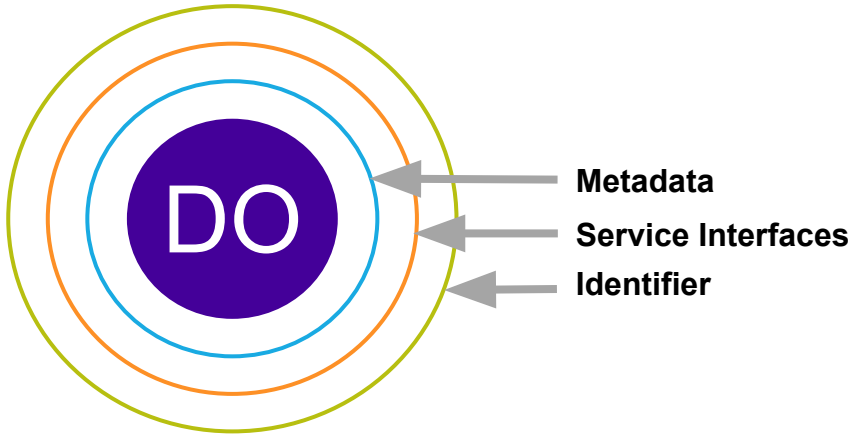


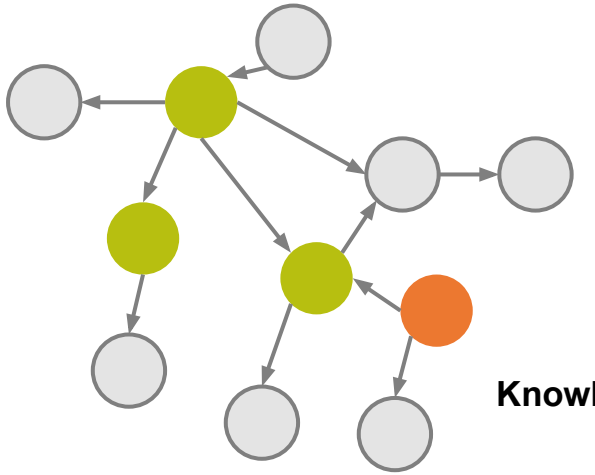
- Publications, data, and other **digital artefacts** (e.g. models, workflows, scripts/code) should be taken into account and should be properly represented

Data Infrastructures

- Digital artifacts should be enriched with **metadata** and should be **linked** for contextualization
- The technical foundation is formed by (**research) knowledge graphs** and **FAIR digital objects**

FAIR Digital Objects (FDOs)





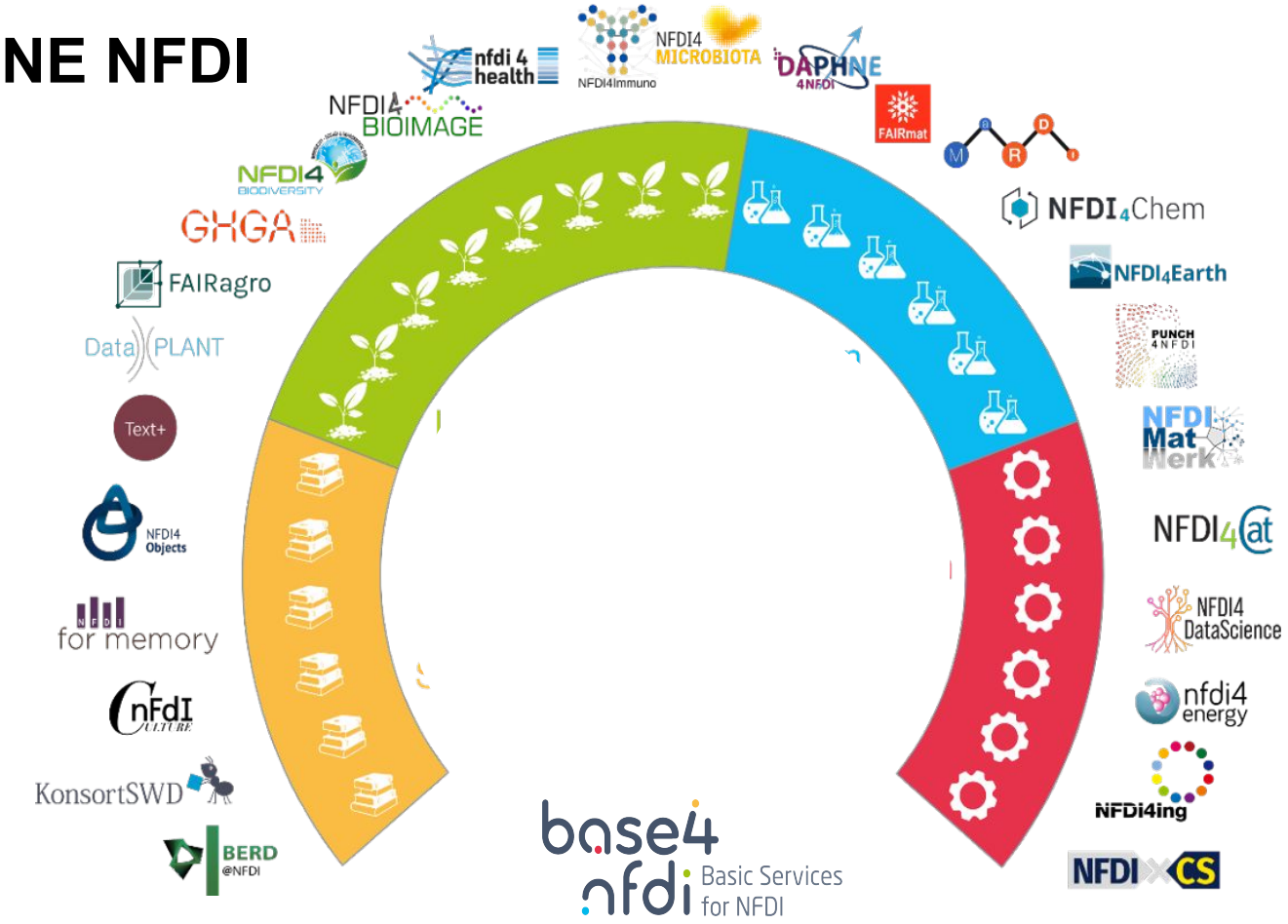
Knowledge Graphs (KGs)

Data Infrastructures as a Foundation for AI Projects

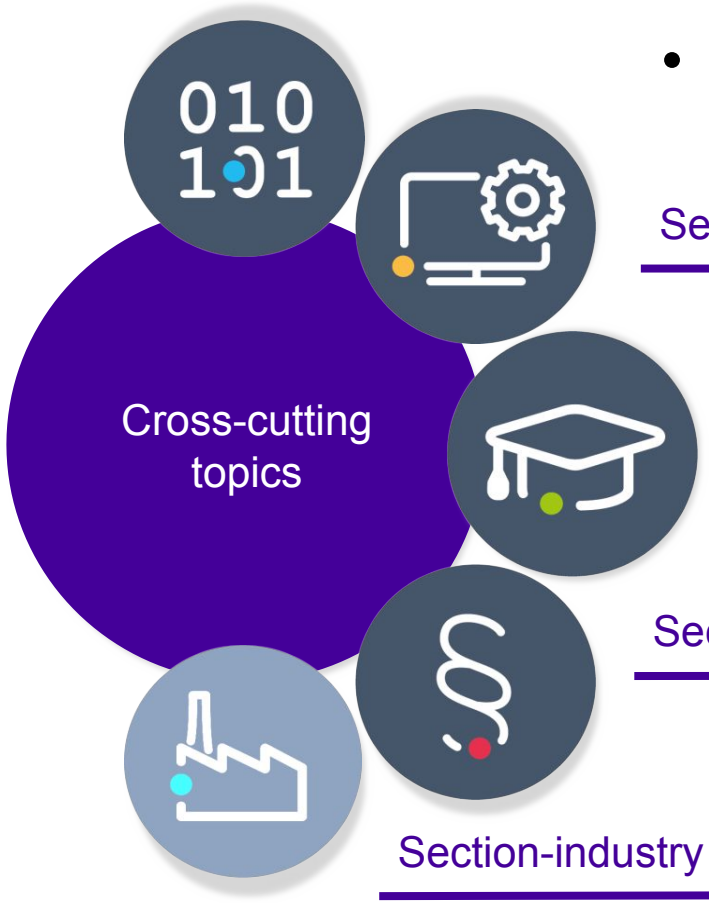
Projects: NFDI and NFDI4DS

National Research Data Infrastructure (NFDI)

- NFDI to set up a **German National Research Data Infrastructure**
  - The goal is to build **ONE NFDI**
- 26 NFDI consortia



National Research Data Infrastructure (NFDI)

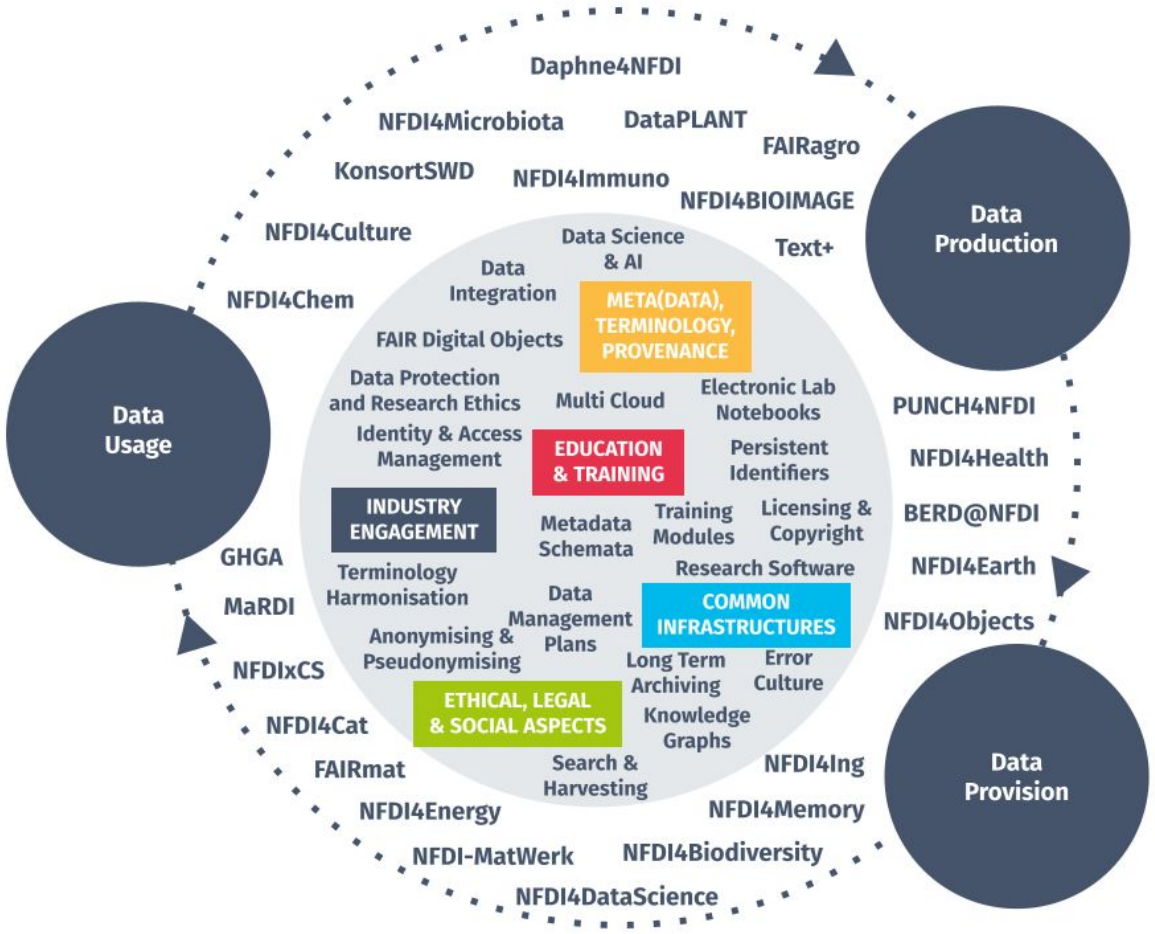


- Sections to work on cross-cutting topics

Section-infra	Common Infrastructures
Section-edutrain	Training & Education
Section-ELSA	Ethical, Legal and Social Aspects
Section-industry	Industry Engagement

National Research Data Infrastructure (NFDI)

- Base4NFDI to develop basic services



IAM4NFDI
Identity & Access Management
Supported by: Section Common Infrastructures

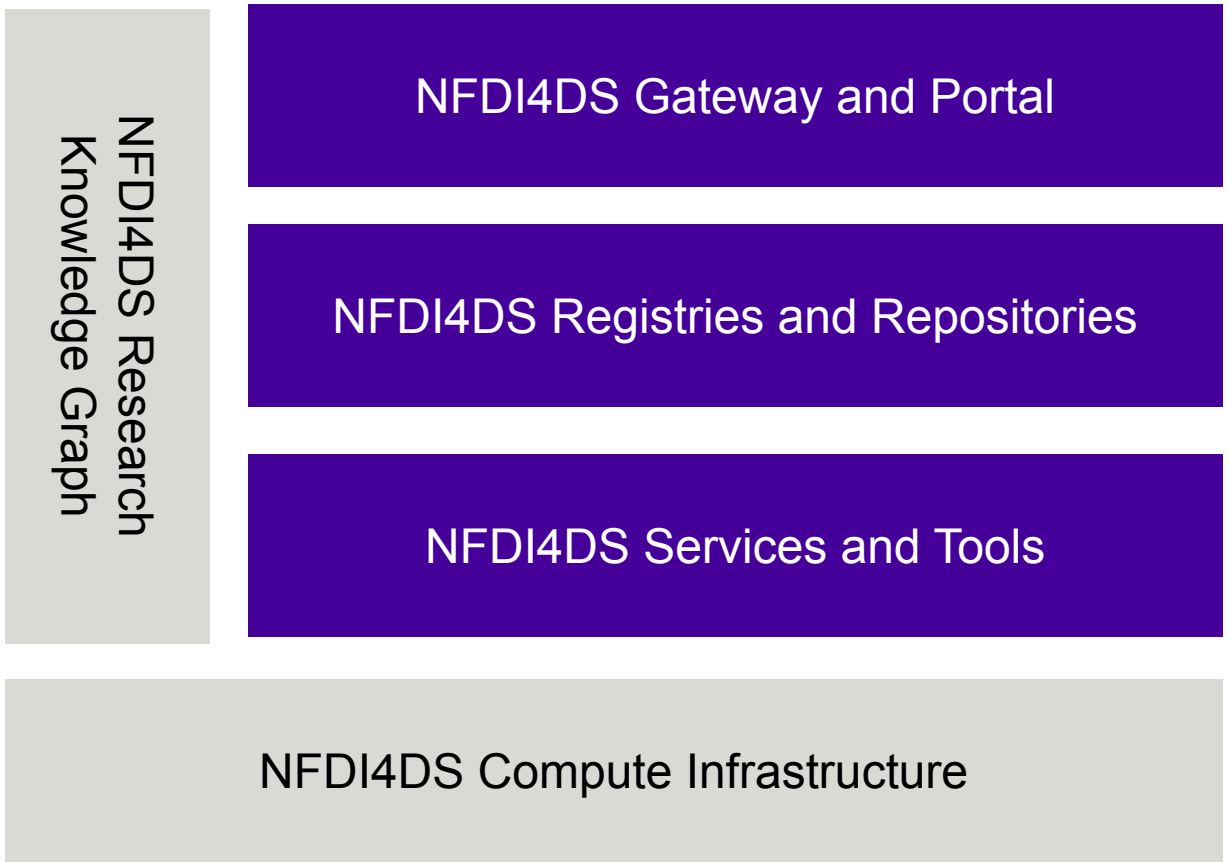
PID4NFDI
Persistent Identifier Service
Supported by: Section Common Infrastructures

TS4NFDI
Terminology Services
Supported by: Section Metadata, Terminologies, Provenance

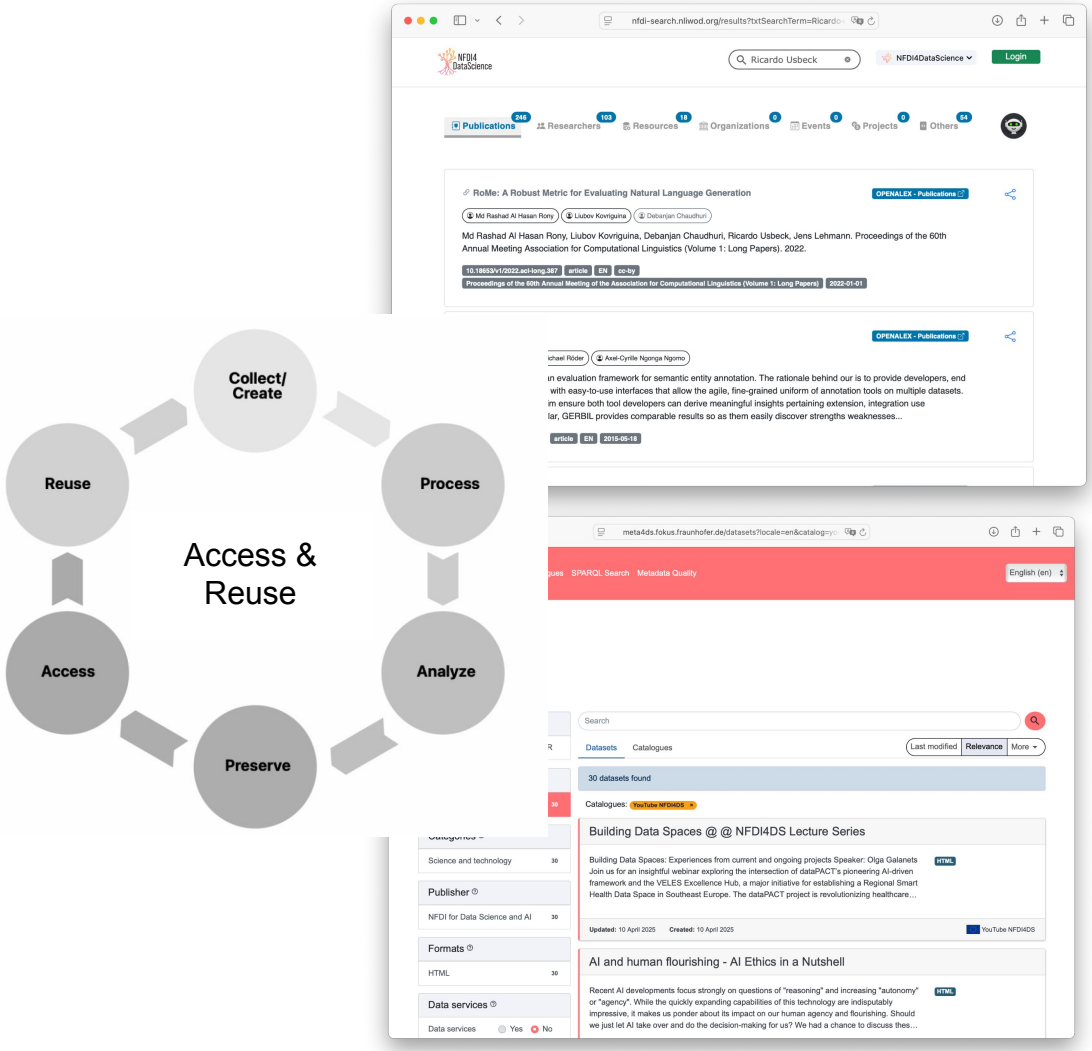
Data Infrastructures as a Foundation for AI Projects

NFDI4DS - Infrastructures

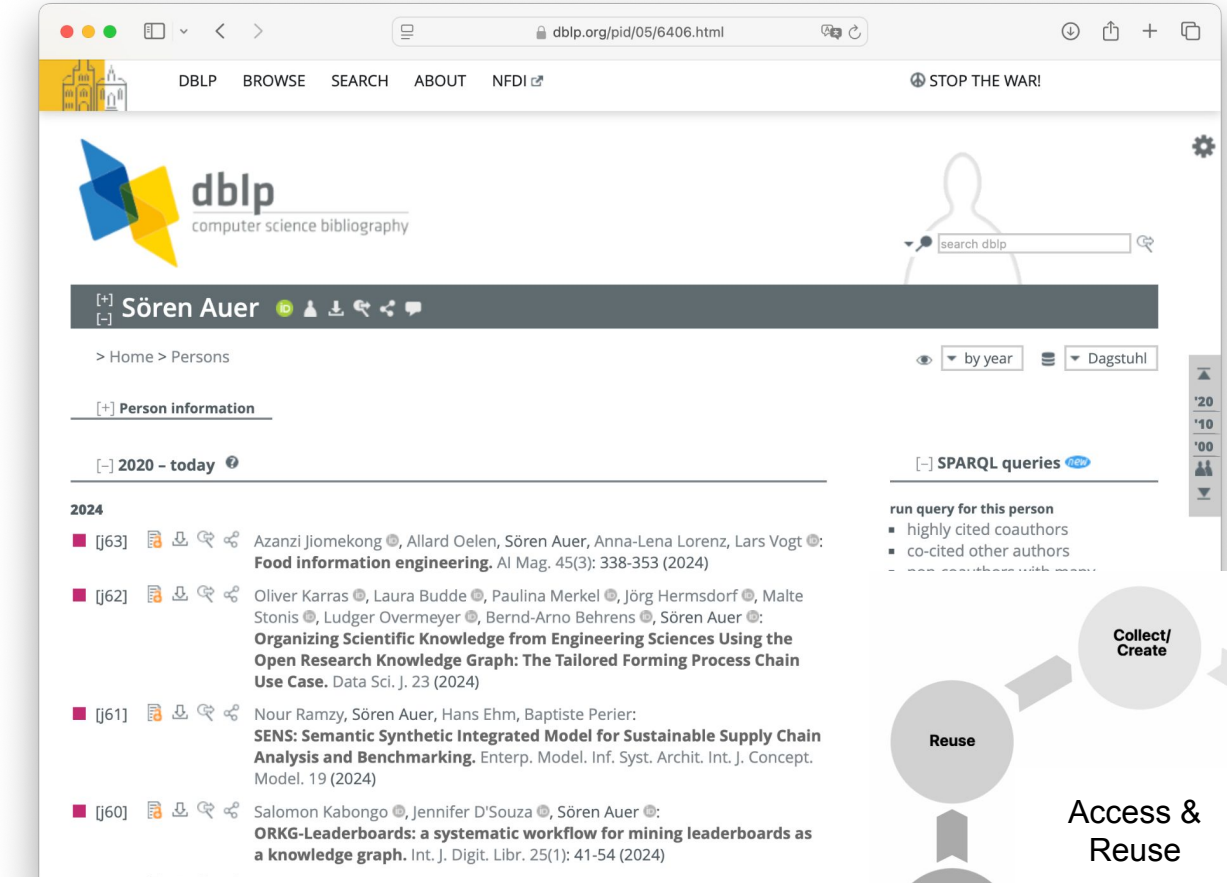
NFDI4DS: Service Portfolio



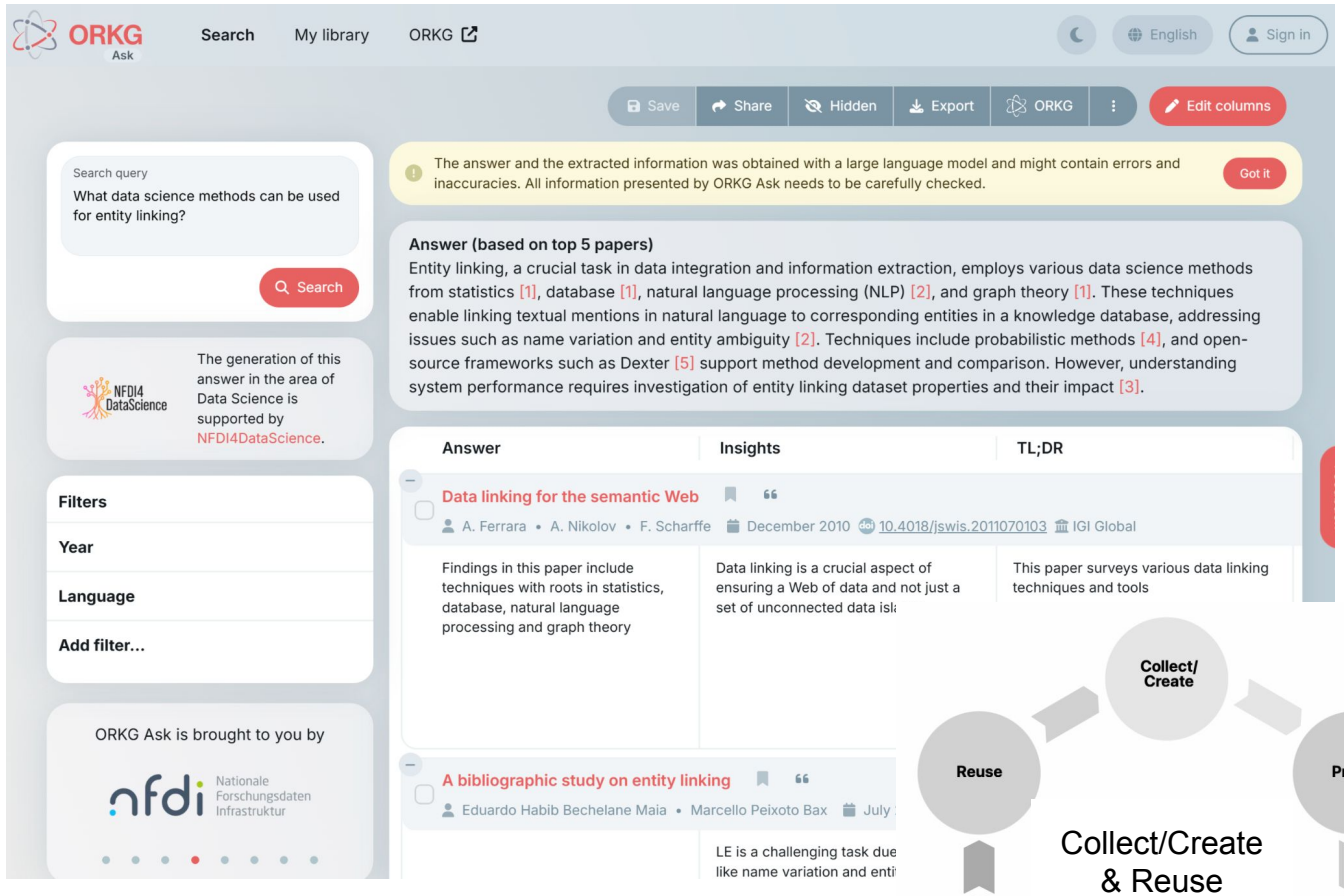
NFDI4DS: Service Portfolio – Gateway and Portal

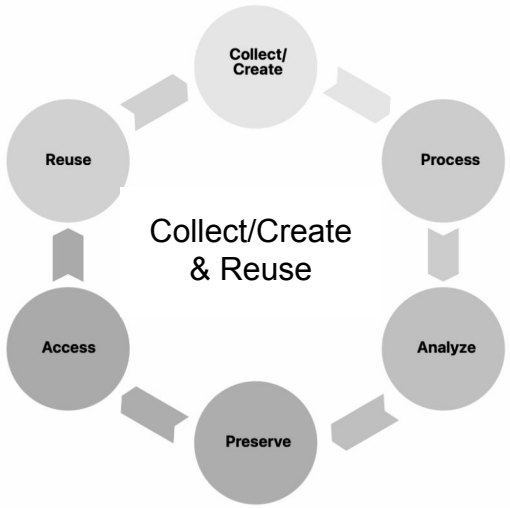


NFDI4DS: Service Portfolio – DBLP



NFDI4DS: Service Portfolio – ORKG and ORKG Ask





Data Infrastructures as a Foundation for AI Projects

NFDI4DS - Shared Tasks

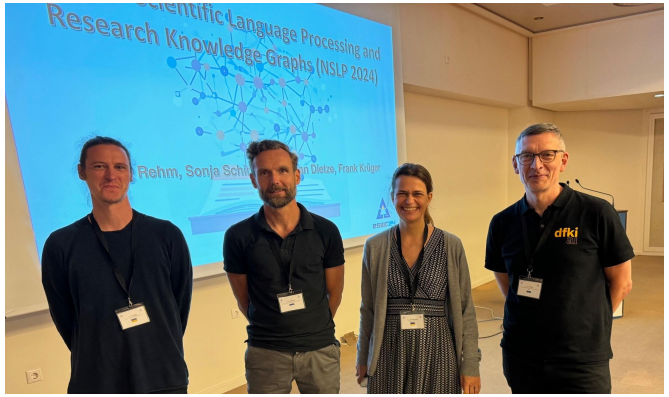
NFDI4DS: Shared Tasks

- Development of new tools and services
- Community involvement

SPGQA - Sample : Sheet1

sample_id	question	questionEntity	answerE
8413	Who was formerly an actor and now a Republican senator?	actor, Republican Party	Arnold S
8414	Who was formerly an actor and now a Republican senator?	actor, Republican Party	Chuck G
8415	Who was formerly an actor and now a Republican senator?	actor, Republican Party	Jeff Flak
8416	Who was formerly an actor and now a Republican senator?	actor, Republican Party	John Mc
8417	Who was formerly an actor and now a Republican senator?	actor, Republican Party	Jon Hunt
8418	Who was formerly an actor and now a Republican senator?	actor, Republican Party	Lindsey
8419	Who was formerly an actor and now a Republican senator?	actor, Republican Party	Mitch Mc
8420	Who was formerly an actor and now a Republican senator?	actor, Republican Party	Orrin Ha
8421	Who was formerly an actor and now a Republican senator?	actor, Republican Party	Rand Pa
8422	Who was formerly an actor and now a Republican senator?	actor, Republican Party	Rick San
8423	Who was formerly an actor and now a Republican senator?	actor, Republican Party	Roy Blun

KGQA

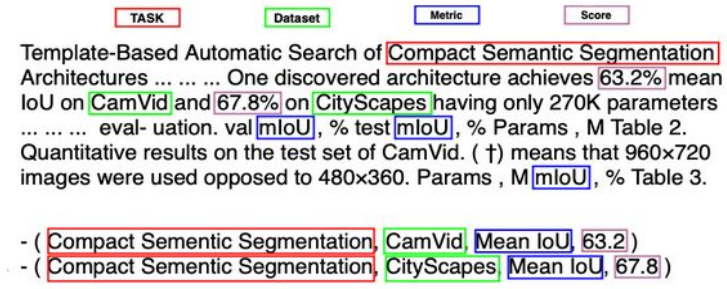


FoRC & SOMD

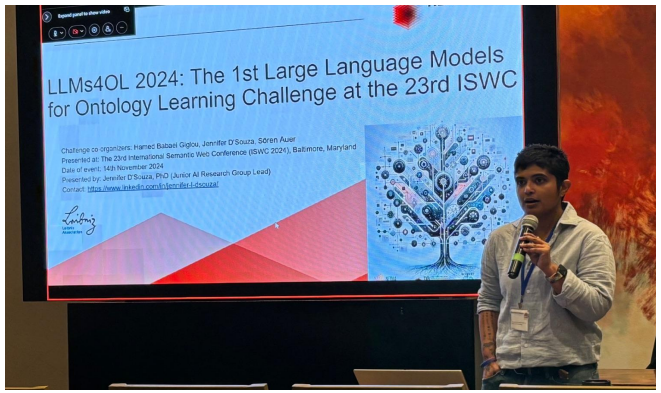
- KGQA **@ACL 2024**
  - Knowledge graph question answering
- FoRC & SOMD **@ESWC 2024**
  - Field of research classification for scholarly publications
  - Software mention detection in scholarly publications

NFDI4DS: Shared Tasks

- Development of new tools and services
- Community involvement



SOTA?



LLMs4OL & Scholarly QALD

- SOTA? **@CLEF 2024**
  - State-of-the-art tracking in AI scholarly publications
- LLMs4OL & Scholarly QALD **@ISWC 2024**
  - Large language models for ontology learning
  - Hybrid question answering over scholarly knowledge graphs and text

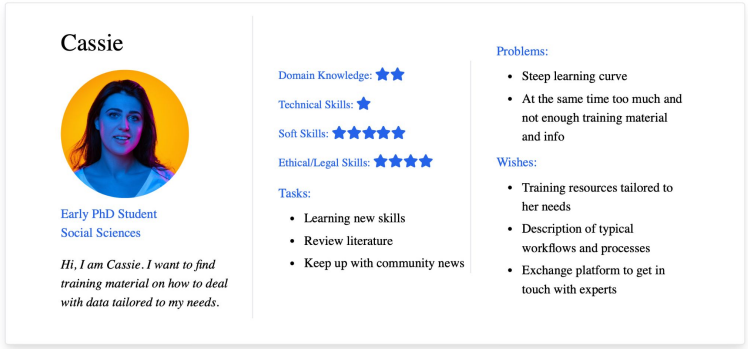
Data Infrastructures as a Foundation for AI Projects

NFDI4DS - Community Involvement

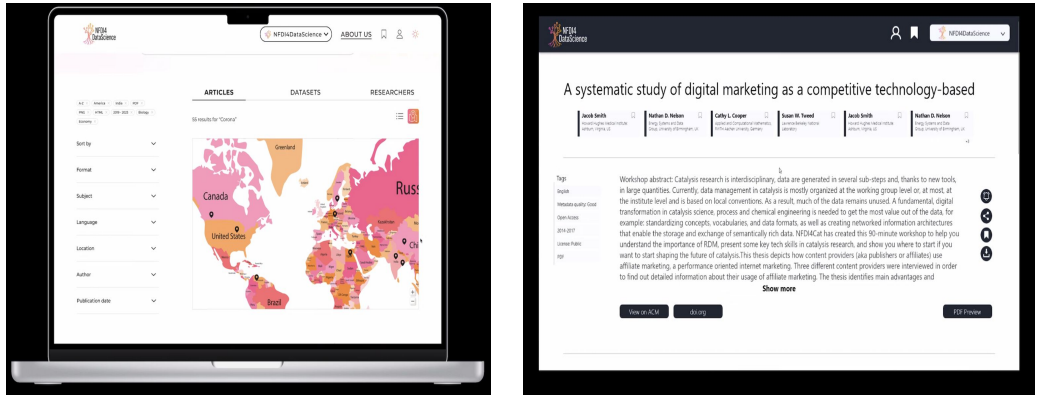
NFDI4DS: Community Involvement

- **Requirements Elicitation**
  - Surveys, interviews, workshops, classes
- **User Testing**

User Testing @ CoRDI



Persona Development @ 4DS Conference



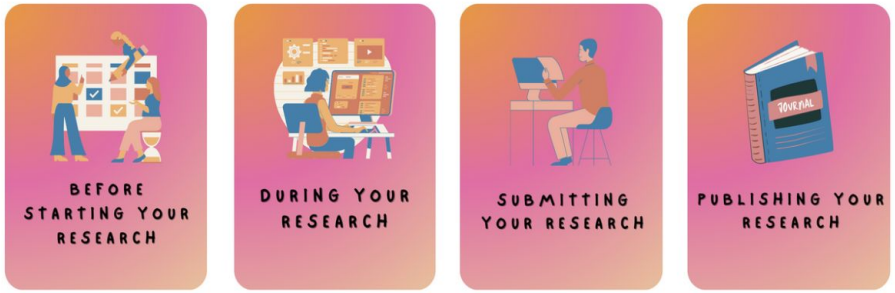
Design of 4DS Access Point @ University Class

NFDI4DS: Community Involvement

- **Training and Education**

Video Series on Ethical Aspects in AI

Book on NSLP



Open Science Best Practices

Summer School

Data Infrastructures as a Foundation for AI Projects

Summary

Challenges in AI

Flood of research artefacts
paper, data, models, scripts/code, …

Changing environment
data science & AI hype

Bias
of training and evaluation data

Trustworthiness and **responsibility** of AI

Good scientific practice
transparency & reproducibility

Digital scientific practice
tools & services

Data Infrastructures meet AI

Generative AI

- Is there a difference in the use of LLMs in the different communities?
- How can we support researchers and the general public that use LLMs?

Infrastructures and Services for Generative AI

- How to adapt infrastructures to support researchers that use LLMs?

Data Competencies for Generative AI

- What new competencies are needed by researchers that use LLMs?

Questions?

Data Infrastructures as a Foundation for AI Projects

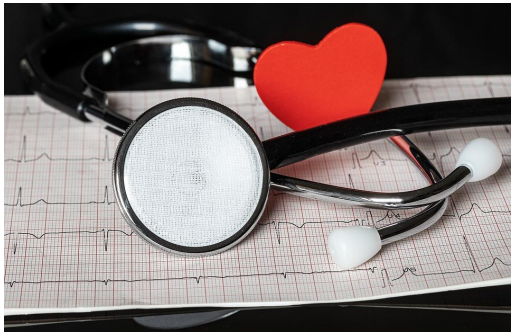
Prof. Dr. Sonja Schimmler
Fraunhofer FOKUS and TU Berlin

Discussion Session

Next Colloquium

- **7 August, 11:00-12:00**
- Join online or at KIT Institute of Biomedical Engineering (IBT), KIT Campus South
- **Till Keller,** Department of Cardiology and Angiology, Medical Center, University of Freiburg, Bad Krozingen + Justus Liebig University Giessen, Giessen
- Topic: **Use of AI in electrocardiogram (ECG) analysis**
- More info + registration: www.ditrare.de/en/ditrare-colloquium

DiTraRe use case: AI in Biomedical Engineering



DiTraRe Symposium

- **2-3 December 2025**
- ZKM Karlsruhe
- Sessions on:
  - Knowledge Representation and AI
  - Law and Ethics in Digitalisation of Research
  - Research Infrastructures
  - Impact on Science and Society
- Call for posters

Website coming soon!

Leibniz ScienceCampus
Digital Transformation of Research

Thank you for joining!

Stay connected

- DiTraRe
  - Website: www.ditrare.de/en
  - Email: ditrare@fiz-karlsruhe.de
  - LinkedIn: www.linkedin.com/company/ditrare
  - Mastodon: social.kit.edu/@DiTraRe
  - YouTube: www.youtube.com/@DiTraRe
  - Zenodo: zenodo.org/communities/ditrare
- Discussion forum: www.ditrare.de/en/forum
- Newsletter: www.ditrare.de/en/newsletter

www.ditrare.de/en

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

KIT
Karlsruhe Institute of Technology