Proposal
Leibniz ScienceCampus

Digital Transformation of Research

[DiTraRe]

Leading Leibniz Institute
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure (FIZ KA)

Participating University/Universities
Karlsruhe Institute of Technology (KIT)

Designated Speaker
Prof. Dr. Franziska Boehm (FIZ KA)

Scientific and Administrative Coordinator
Dr. Felix Bach (FIZ KA)
Dr. Christiane Noe (FIZ KA)

B) Quality of the project and track record of the partners

B.1 Project summary and introduction to the academic field

The digital transformation of research affects both academia and society as such. It comprises the adaptation of epistemological processes as well as the transparency and communication of findings. New data-driven methods not only open up innovative approaches for knowledge generation, but also raise legal, ethical and societal questions. This transformation has been underway for many years, but is by no means complete. Hey et al. showed the massive influence of digitalisation and discussed data exploration as a new science paradigm (Hey et al., 2009). The Council for Information Infrastructures stated: "The digital transformation has taken hold of research data and research methods, it is fundamentally changing science" (RfII, 2016). The National Research Data Infrastructure (NFDI) demonstrates the importance of digital transformation and the associated research data management (RDM) for all disciplines. The Weizenbaum Institute's group "Digitalisation and Research" understands digitalisation "as a change in research and publication processes that promises to increase the benefits of research".[1] The Deutsche Forschungsgemeinschaft refers to the term 'digital turn' as "all relevant changes and effects of an epistemic, ethical, legal, technical, infrastructural, organisational, financial and social nature resulting from the development and use of digital technologies" (Katerbow et al., 2020). Following this definition, in the planned LSC we investigate the influence and consequences of digitalisation on scholarly cultures, the trust in artificial intelligence (AI), security and privacy awareness, legal and ethical aspects, data quality, Open Science, and the change in relevance of scientific evidence in society.

B.2 Innovative character of the project/subject field in relation to the state of the art

The LSC takes up the motto of the Leibniz Association (LA) "theoria cum praxi", combining reflective with practical elements. In line with current federal and state policy goals, it employs a transdisciplinary and holistic methodology to address highly relevant research questions along with exemplary use cases in four dimensions (see B.3). DiTraRe will complement the NFDI with a much broader, dimension-based view on digitalisation of research and thus provides new impetus at both practical and science policy levels. This innovative approach leads to relevant insights for decision-makers in the science system including funders.

(i) **Knowledge organisation** requires formally sound methodologies to represent, organise, and manage domain specific as well as procedural knowledge. Semantic technologies are key enablers for findability, accessibility, interoperability, and re-use (Wilkinson et al., 2016) of ever-increasing amounts of research data. Knowledge graphs (KG) based on formal ontologies are connecting heterogeneous sources within single domains and beyond (Farhalla et al., 2017; Hogan et al., 2021). Hybrid AI methods combining symbolic representations and subsymbolic machine learning are increasingly exploited to index unstructured data and to explain deep-learning-based classifications and decisions (Xu et al., 2019; Sack 2021). The LSC focuses on the distributed generation of KG and therein on the interconnection of freely accessible with protected information.

(ii) **Infrastructures** such as data repositories, data archives, virtual research environments and smart lab solutions have been conceptualised and deployed (Schomburg et al. 2011; Lin et al., 2020; OECD, 2020) to enable data-driven science (Hey et al., 2009). They make scholarly output accessible to both academia and the public, promoting the digitalisation of the research system (Borgwardt, 2016). Generic infrastructures and services do not yet fully meet

[1] https://www.weizenbaum-institut.de/en/research/rg11/ [last visited on July 21, 2022].

the needs of researchers and are not well integrated into their processes (RFII, 2016; BMBF, 2019). Disciplinary digital tools tailored precisely to the needs of researchers are currently being developed, but their acceptance and impact are not yet foreseeable.

(iii) **Security awareness** has been studied in the private and organisational context, e.g. interviews to identify mental models (Volkamer/Renaud, 2013) and user studies to evaluate security awareness measures (Volkamer et al, 2020; Reinheimer, 2020). Recently, researchers have started studying the security awareness of software/app developers (Naiakshina, 2017; Tuladhar 2021) and individuals applying machine learning (Bieringer et al., 2022; Boenisch et al., 2021). In the LSC, we design measures to increase the security awareness of researchers developing scientific software. This ultimately touches ethical, trust related, and legal aspects.

(iv) **Reflection and resonance** on digitalisation and data usage in research become increasingly relevant. Fundamental changes in the coordination of and the communication within societal processes take place (Schrape, 2021). Issues like responsibility and risks gain in importance (Bogner et al., 2022; Grunwald, 2022; Hosseini et al., 2022; Huijstee et al., 2022; Orwat et al., 2022). Technology Assessment (TA) deals with the uncertainty of knowledge and the inclusion of relevant societal stakeholders, an important aspect regarding science communication (Böschen et al., 2021; Bonn et al., 2021). An analysis of the cooperation of different stakeholders in transdisciplinary settings facing the digital transformation of research is needed. In the LSC, we will extend the knowledge and practices of TA towards the specific requirements concerning data usage by addressing ethical and social aspects.

(v) Digital research, data sharing and RDM pose **legal and ethical challenges** which are not adequately addressed by existing legal instruments, necessitating changes to existing laws and new legislation. This pertains data protection law, intellectual property (IP) law, public sector information regulations and other data regulating legislation at EU (i.e. Data Governance Act or the EU Data strategy EU Com, 2022 and EU Com, 2021) and at national level (i.e. German Data Strategy, Datennutzungsgesetz, BMBF 2019, Globocnik/Richter, 2022). Although the interests of research, education and innovation have been partly taken into account (Specht-Riemenschneider/Wehde, 2022; Richter, 2018), the concrete transformation of the legal provisions still needs to be elaborated. The FAIR principles raise issues with IP rights, data protection law or research cooperation agreements (Graber-Soudry et. al., 2021, Hoffmann/Begoña, 2021). Analysing data with methods of AI, machine learning or text and data mining require a legal and ethical framework (Chiou, 2020; Surblytė-Namavičienė, 2020). Legal, technical and ethical considerations need to go hand in hand to find adequate techniques and models, e.g. anonymisation and data trustees. We will propose appropriate changes in law and develop guidelines on using personal data for research purposes.

B.3 Work programme with milestones and work packages

“In the future, scientific work in almost all disciplines will be significantly characterised by digital research practices and information infrastructures" (Katerbow et al., 2020). Shaping this future is the overarching goal of DiTraRe. It follows a transdisciplinary approach, and central research questions will be investigated along four dimensions:

- **Reflection and Resonance (D-RR, Sociology and TA)**: Research findings are increasingly shaped by algorithms. All stakeholders require transparent and reproducible epistemological processes. In order to create trust, communication of findings demands transparency and comprehensibility - within the science system and

into society. This shift must therefore not only be examined from a technical, but also from a societal and ethical perspective.

- **Exploration and knowledge organisation (D-KO, Computer Science)**: Semantic technologies provide a formal representation of knowledge contained in research data, thus facilitating the efficient integration of heterogeneous data sources. The growing adoption of AI-based knowledge mining technologies requires comprehensible and trustworthy AI algorithms ("explainable AI").
- **Legal and ethical challenges (D-LE, Law)**: The digital transformation is shaped by data ethics, data protection, copyright and data law. Emerging legal framework conditions need to be evaluated, as they may provide the means to make sensitive data collections available for subsequent use in a legally secure manner. This requires a thorough analysis of the involved research data and its legal and political context.
- **Tools and processes (D-TP, Software Engineering)**: The entire process from collection to publication and archiving of research data is affected by the transformation. Smart tools such as Electronic Lab Notebooks (ELN) support researchers in generating and providing interoperable and documented data with the ultimate goal of a data continuum. At the same time, the security of the processes and the security awareness of those involved must be intensified.

The work programme of DiTraRe is based on a unique combination of "digital enablers" such as computer science with application disciplines. Accordingly, it is organised as a matrix: Four research clusters (RC) each start from a scientific use case that raises concrete questions. These questions are examined along the four dimensions. Subsequently, (i) an accordingly coherent solution is developed which (ii) is evaluated by the use case partners, (iii) the questions are generalised and placed in the overall context of the DiTraRe in order to move from the concrete to a more abstract level, (iv) ensuring the transferability of the research results to other disciplines. Thus, the work programme points far beyond the focus of the NFDI, looking at both the effects of digitalisation on research and society and at related processes before and after the actual RDM.

B.3.1 Work Package 1: Research Cluster "Protected Data Spaces"

Use Case: Sensitive data in sports science *(**KIT-IfSS**, Dr. Claudia Niessner)*

The MO\|RE data platform makes physical fitness data derived from sports science studies available to both academia and the public. Research would strongly benefit from linking health with physical fitness data, e.g. in longitudinal data sets. However, publishing sensitive health (e.g. BMI, blood pressure) and other personal data (e.g. geolocation, social status) is challenging. An overarching concept for the secure handling of sensitive data is lacking, ranging from a trustworthy IT environment to sophisticated access management and auditing mechanisms, which ensures compliance with legal regulations.

Various categories of research data are subject to legal restrictions such as data protection laws, personal rights or copyrights. Ethical restrictions for data sharing include e.g. the geolocation of sensitive sites or politically or socially unacceptable content. Nevertheless, there is a legitimate research interest in such data. In the LSC, we clarify which categories of data can be re-used for research purposes. We propose legal, ethical and technical solutions by taking into account different levels of data sensitivity. We investigate procedures for pseudonymisation and anonymisation and linking of sensitive data with non-critical data in

distributed systems of knowledge organisation. We study researchers' awareness of the associated security and privacy risks and possible consequences.

- **D-RR**: explores ethical issues in the handling of sensitive data, e.g. necessary negotiation processes and trade-offs between transparency and data protection, as well as the trust of data subjects and researchers in the completeness, integrity, anonymity and traceability of data.
- **D-KO**: focuses on the representation of (partially) protected information in open KG and on interweaving protected with non-critical data.
- **D-LE**: focuses on data protection and pseudonymisation/anonymisation of different categories of sensitive data and conceptualises re-use models by taking into account ideas such as data intermediation services and data altruism. Based on the results, we develop security and privacy awareness measures.
- **D-TP**: focuses on requirements for secure storage of personal data with regard to integrity, confidentiality, authenticity and availability. We design security and privacy solutions that consider (socially) acceptable risks.

B.3.2 Work Package 2: Research Cluster "Smart Data Acquisition"

Use Case: Chemotion Electronic Lab Notebook *(KIT-IBCS, Dr. Nicole Jung)*

Chemistry labs in academia make limited use of lab automation and device integration. Despite current research data guidelines by funders and positive examples in industry, there is reluctance to adopt technologies such as ELNs. Concerns include dependencies on software and technologies not under control of scientists, faulty methods for data assignment and data analysis, and missing control over re-use of their data.

The RC investigates innovative technical and societal methods as well as quality criteria for data acquisition as well as partially automated procedures for documentation, analysis and interpretation of data, thus fostering the acceleration of research processes. It assesses associated opportunities and risks, including legal challenges related to IP protection. The Chemotion ELN will serve as a testbed to investigate the efficiency of data acquisition and analysis as well as the establishment of trust and accountability.

- **D-RR**: focuses on the appropriate involvement of society in virtual spaces, e.g. define the conditions for acceptance of data from a stakeholders' perspective in transdisciplinary research processes.
- **D-KO**: focuses on the application of novel methods of data acquisition, analysis and interpretation, and their evaluation in comparison to the results gained by traditional intellectual processes.
- **D-LE**: focuses on IP legislation and licences for cooperatively created data and data resulting machine-based analyses.
- **D-TP**: focuses on automation and virtualisation of laboratory research ("smart lab"), including the design and implementation of next-generation ELN with innovative functions to facilitate accelerated research processes, such as directly connected digital measuring devices.

B.3.3 Work Package 3: Research Cluster "AI-based Knowledge Realms"

Use Case: Artificial Intelligence in Biomedical Engineering *(KIT-IBT, Dr. Axel Loewe)*

KIT-IBT develops computer models of the human heart to predict cardiovascular diseases earlier and more accurately using software engineering, algorithmics, numerics, signal

processing, data analysis, and machine learning. We employ AI methods trained on purely synthetic or hybrid (simulated + clinical) datasets to help decipher disease mechanisms. Simulated data are often essential to overcome issues of data privacy and existing bias in most available datasets, but raise questions of explainability of AI decisions and trust.

Machine learning and AI hold great promise to enable new discoveries and innovation. They help address issues of ever-increasing amounts of data and offer opportunities to semantically link currently separated information. However, they are accompanied by risks, ranging from legal assessment of the use of synthetic training data for AI systems, limited or biassed training data and quality problems in indexing to a lack of acceptance by users due to unverifiable decisions by AI systems. This applies in particular to the social, political and economic consequences of AI-based decisions made by models that can no longer be explained or understood ("black boxes").

- **D-RR**: focuses on the design and communication of decision-making processes in view of the uncertainty of knowledge, e.g. knowledge gained by AI methods, as well as on the applicability of machine learning methods with regard to data representation.
- **D-KO**: focuses on explainability and explanatory components based on symbolic knowledge representation (combination of symbolic and sub-symbolic AI, explainable AI), on hybrid AI systems that complement humans, on automatically generated semantic relations in KG.
- **D-LE**: focuses on the potential of synthetic training data for AI systems with regard to the challenges of data protection and legal questions regarding digital twins, e.g. in medicine.
- **D-TP**: focuses on necessary infrastructure to generate, store and disseminate (synthetic) training data, with emphasis on interoperability and re-use.

B.3.4 Work Package 4: Research Cluster "Publication Cultures"

Use Case: Publication of large datasets *(KIT-IMK, Dr. Sabine Barthlott)*

KIT-IMK generates and analyses very large datasets in chemistry-climate simulations or in satellite data for observing the state of the atmosphere. Publication of those data is currently very inefficient due to their size. Re-use is hampered by missing methods for exploring such datasets efficiently in order to evaluate their relevance for other research questions. Selecting subsets of datasets for re-use or peer review is currently not possible.

New publication formats beyond classic peer-reviewed articles are gaining in importance. Data publications make scientific findings reproducible and form the basis for further research (RfII, 2016). Software used to generate or interpret data must be included with data publications as a quality assurance measure. Both should be understood as first-class scholarly outputs (McKiernan et al., 2016). Existing publishing infrastructures are not yet well suited for data and software (RfII, 2021). The dynamically changing legal framework requires an in-depth analysis of European and national data laws and policies (see B.2.iv)) and their impact on new publication formats and researchers' willingness to share data, algorithms and software. The shift towards Open Science must be accompanied by a suitable communication strategy to help prevent misinterpretation of research results. It takes into account new communication formats and stakeholders such as science communicators or decision-makers to improve the exchange between science and society.

- **D-RR**: focuses on the necessity of a cultural change with regard to the needs of research and society for mutual communication, especially in the light of Open Science, and new formats of science communication.
- **D-KO**: focuses on massively parallel authoring and quality assurance of large KG.
- **D-LE**: focuses on the impact and analysis of data laws, policies and data strategies, particularly with respect to Open Science.
- **D-TP**: focuses on the publication and exploration of extensive datasets, on new publication formats, e.g. including actionable source code, partially generated text, software publications.

B.3.5 Deliverables

- Concrete solutions for the problems identified in the use cases in form of published guidelines, analyses, tools, and best practice documents provided
- Compendium of research results on DiTraRe´s website as a "living document" published and continuously updated
- At least 16 peer-reviewed publications by the participating working groups
- Four additional third-party funding applications (one per dimension) approved
- Two international scientific conferences held, establishing the format of a “Karlsruhe Conference on Digital Transformation”
- One transdisciplinary colloquium on digital transformation of research held
- A jointly supervised project course on topics of the LSC for master students organised
- A science communication workshop organised to discuss perspectives from science communicators on knowledge access and evaluation
- A virtual society-science dialogue format, e.g. hackathons, to link citizens' perspectives with the digital transformation in science organised
- A Dagstuhl seminar[2] held

B.3.6 Milestones

Milestones are associated with all research clusters and dimensions together.

Milestone 1 (End of 2023)	Requirements of the use cases elaborated after analysis from the perspective of all four dimensions
	First version of the compendium published
Milestone 2 (Mid 2024)	Solution concepts for use cases finalised
	Research questions going beyond the use cases for all RCs substantiated
	First conference held
Milestone 3 (Mid 2026)	Use case solutions rolled out
	Analysis of potential carried out for all research clusters and dimensions
	Dagstuhl seminar on “Digital transformation of research” held
Milestone 4 (Early 2027)	Final version of the compendium published
	Concluding conference held

B.4 The research project's relevance in terms of its contribution to solving current academic and social, ecological or economic problems

DiTraRe addresses urgent questions in shaping the digital transformation in research and beyond in society. It covers central aspects of the Digitalisation Strategy of the Helmholtz Association: “*Research under the auspices of the digital transformation has an impact on*

[2] https://www.dagstuhl.de/en/program/dagstuhl-seminars/ [last visited on July 20, 2022].

society as a whole. The Helmholtz Association considers it its mission to reflect these societal effects" (Holl-Supra et al., 2019) as well as the LA's self- commitment in the PFI IV:[3] "*focus on science in the digital transformation across the board*". The LSC refers to current government policy priorities (Koalitionsvertrag, 2021): *"data-based solutions across all sectors*" are mentioned as "*crucial field for the future*". DiTraRe provides valuable impulses for "*digital innovations and digital infrastructure including aspects of IT security and digital sovereignty*". Its topics of data access and availability, standardisation, data intermediation services, anonymisation techniques and licences resonate well with supporting "*the development of data infrastructures and launch instruments such as data trustees, data hubs and data donations together with business, science and civil society*".

B.5 Track record of the participating researchers in their respective field

Prof. Dr. Franziska Boehm is Vice-President IP rights at FIZ KA and W3 professor of law at KIT. As one of the few law professors at an infrastructure institution, she has a unique insight into the practical as well as theoretical legal issues of RDM. With her two research groups, she is involved in national and international studies and research projects focusing on data law and IP rights. She is a sought-for consultant at top EU level. She is member of the Leibniz Strategy Forum on Digital Transformation, participates in several NFDI consortia[4] and acts as coordinator of the ethical, legal and social section of the NFDI association (ELSA). She is the designated speaker of the LSC and PI of D-LE.

Dr. Linda Nierling is head of the research group "Digital Technologies and Societal Change" at ITAS at KIT. She holds a PhD in Sociology and has broad research experience in TA with a focus on societal changes. She is PI at the Helmholtz Topic "Knowledge for Action". She has led or was involved in several research projects dealing with the impact of technologies in society with a focus on digital transformation in work as well as AI's relation to public communication. She has intensively worked on the interface of research and society with a focus on policy advice. She is PI of D-RR and the RC "Publication Cultures".

Matthias Razum is Vice-President e-Research at FIZ KA. He has a master in Business Informatics and has long been concerned with digital transformation in research and RDM. He has initiated or played a leading role in a large number of innovative interdisciplinary projects, e.g. the SAW projects TOPORAZ/TRANSRAZ, focused on tools and processes tailored for the "digital turn" e.g. in the Digital Humanities. He is responsible for software development and operation of the generic research data repository RADAR and is significantly engaged in several NFDI consortia.[5] He is PI of D-TP and the RC "Smart data acquisition".

Prof. Dr. Harald Sack is Vice-President Information Engineering at FIZ KA and a W3 professor of computer science at KIT. With his two research groups, he is involved in national and international projects focused on KG construction, management, and applications in different scientific domains ranging from mathematics and data science over material sciences up to cultural heritage and digital humanities. His research focus lies in the intersection of symbolic and subsymbolic knowledge representation. He is co-spokesperson of several NFDI consortia[6] and is PI of D-KO and the RC "AI-based Knowledge Realms".

---

[3] Joint Initiative for Research and Innovation.

[4] E.g., NFDI4Culture, NFDI4Chem, NFDI4DS, FAIRAgro, and NFDI4CS.

[5] Co-applicant in NFDI4Chem and NFDI4Memory, participant in NFDI4Objects, and member of the Operations Coordination Committee of Text+.

[6] NFDI4Culture, NFDI-MatWerk, NFDI4DataScience, and MaRDI.

Prof. Dr. Melanie Volkamer is W3 professor at KIT. She leads the SECUSO research group and is principal investigator at the Helmholtz Topic Engineering Secure Systems as well as at the KD2lab. She was involved in the KASTEL BMBF project. Her research interests are on usable security and security awareness. In particular her security awareness concept NoPhish is internationally and nationally well known, applied by many organisations, and recommended by institutions like the BSI. She is PI of the RC “Protected Data Spaces”.

B.6 Involvement of partners and disciplines relevant to the project's success

FIZ KA has had a prominent role in coining the term as well as the concept of “information infrastructures” at the governmental level of the Joint Science Conference (GWK). It has recognised “digital expertise” (Katerbow et al., 2020) and international networks in knowledge engineering, RDM, e-research infrastructures, and law. KIT as a university of excellence complements these competencies ideally with research on security awareness, sociology and technology assessment. With their inter- and transdisciplinary orientation, both partners have an important prerequisite for dealing with this topic holistically.

B.7 The participating partners' previous collaboration

The two partners have been successfully cooperating for a long time. Since 2014, we have formalised the partnership in a cooperation agreement. Two joint professorships (see B.5) in the areas of "Intellectual Property Rights in Distributed Information Infrastructures" and "Information Service Engineering" have unique denominations to date. There are long-standing, successful cooperative relationships between FIZ KA and various KIT institutes on topics such as data protection, information/knowledge modelling and RDM.[7] In 2019, FIZ KA and KIT jointly established the NFDI Founding Directorate and played a decisive role in the establishment of the NFDI association in Karlsruhe and Baden-Württemberg. The Vice-President Research at KIT is a member of the Scientific Advisory Board of FIZ KA.

C) Structural and strategic efficacy

C.1 Synergies and academic added value resulting from the network

Both partners are very active members of their respective research fields and nationally and internationally well connected with leading researchers, institutes and projects in the four dimensions of the LSC. FIZ KA provides a unique combination of infrastructural, technical and legal expertise in RDM and digital transformation. KIT's excellent research not only permits the challenging research questions to be addressed, but also provides the scientific use cases that ensure the practical relevance of the LSC and enable the validation of results. The close interaction of both partners with their specific competences forms a unique network that enables a thorough investigation of the topic of the LSC.

C.2 Importance of the subject to the scientific environment

Major societal challenges such as climate change, demographic change and sustainability demand quick solutions. The digitalisation of research is a prerequisite for the necessary acceleration of research processes. The COVID-19 pandemic has accelerated the move towards a culture of data sharing. However, due to the high time pressure, sufficient quality assurance was not always maintained (Besançon, 2021). This sometimes resulted in misinterpretations or inadequate studies which were not critically reflected, especially in the general public. Thus, a comprehensive reflection on the opportunities as well as on the risks of the digital transformation is pivotal. At the same time, the use cases show an enormous

[7] This includes openTA, MoMaf, ReNewRS, NFDI4Chem and NFDI-MatWerk.

need for improved processes, tools, guidelines and best practices. Generating adequate security and privacy awareness is negligent. In all these areas, the LSC will aggregate important insights and provide impulses for the further shaping of the digital transformation.

C.3 Conditions at the scientific environment

Both partners provide important building blocks for the digital transformation on which the LSC can rely upon: systems for archiving and publishing research data (RADAR, RADAR4KIT, bwDataArchive), an HPC environment at KIT's SCC, specialised servers for training large machine learning models, and several virtual research environments. KIT provides Chemotion, a widely used open source ELN. FIZ KA contributes information services, legal expertise and KG with foci on patents, cultural heritage and material sciences.

C.4 Measures for further developing the partnership with the university

On the one hand, it is the expansion of joint teaching and training of young academics. To this end, the LSC offers (i) a transdisciplinary colloquium, (ii) a jointly supervised project course for master students at KIT dedicated to our research topics, (iii) research tasks for interdisciplinary student teams and (iv) bachelor and master thesis topics related to the LCS´s research under joint supervision of KIT and FIZ KA. On the other hand, we will link the infrastructure mentioned in C.3 even more closely in future and expand it further by combining “digital enablers” with application disciplines.

C.5 The governance structures of the Leibniz ScienceCampus being applied for

The governance structure corresponds to the lean setup of the LSC with only two partners and aims at an efficient and transparent decision-making process. The **Board of Directors** (BoD) strategically steers the LSC and ensures its structural anchoring in the two partner institutions. The BoD consists of the President & CEO of FIZ KA and the Vice-President Research of KIT, with the LSC´s spokesperson and its Coordinators as permanent guests. The BoD meets at least twice per year. The **Steering Committee** (SC) executes the strategic objectives, oversees the implementation of the work programme, and is responsible for the LSC’s quality assurance. It consists of the PIs of the dimensions and RCs and is chaired by the spokesperson appointed by the BoD. The SC is supported by an international **Scientific Advisory Board** (SAB) and the **Stakeholder’s Committee** which consists of the SC and the representatives of the use cases. The **Administrative** and **Scientific Coordinators** carry out all operational tasks and follows up on milestones and deliverables. They coordinate the RC and dimensions, partners and communication activities. Both are permanent guests in the SC.

C.6 Measures for further internationalisation

The digital transformation of research is per se international. We are active in international initiatives such as EOSC, Research Data Alliance and Global TA network.[8] We will contribute our results there and, vice versa, take up ideas to further shape our work programme. We will apply for a Dagstuhl seminar to discuss the results of the LSC's work with international experts. The SAB ensures networking with similar projects in other countries.

C.7 The strategic importance of the project to the Leibniz Association

DiTraRe takes up the LA's self-commitment to PFI IV and fills key points of the LA's self-understanding as an enabling structure with life, e.g. for reflection and response in terms of issues, data and methods, for interaction between science and society, for knowledge and technology transfer, and for social and political consultancy (see B.3, B.4, C.8). The LSC underlines the LA´s commitment to Digital Change including Open Science (Leibniz, w.y.). The

[8] https://globalta.technology-assessment.info/ [last visited on July 26, 2022].

joint teaching events with KIT (see C.4) offer young researchers attractive, interdisciplinary fields of education and career paths. Besides strengthening the local and regional visibility, DiTraRe exemplifies a future-oriented collaboration between the research university in the Helmholtz Association and a renowned Leibniz institute of information infrastructure.

C.8 Communication concept

The LSC is committed to Open Science and publishes its results transparently and comprehensively. The communication concept is aimed at three target groups: academia, policymakers, and the public. This threefold approach is accompanied by a website and a social media strategy. **Academia** is addressed by scientific communication means (see B.3.6.). **Policymakers** are mainly addressed by D-LE, focusing on consultation and commentary on current legislation. D-RR and the RC “Publication cultures” cover the communication toward **the public**. At the interface between science and the public, we organise targeted dialogue formats with citizens and regular blog posts including guest contributions by science journalists.

C.9 Intellectual property protection and knowledge transfer

The LSC aims to publish as many research results as possible in Open Access or Open Source. Both partners see the open availability of research results as one of the central prerequisites for their transfer and thus for the impact of DiTraRe. Dissemination of the results is ensured through (i) use case partners, (ii) the compendium, (iii) scientific conferences and (iv) colloquia and courses for master students.

Bibliography

Bello, M., Galindo-Rueda , F.: Charting the digital transformation of science: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2). *OECD Science, Technology and Industry Working Papers,* No. 2020/03, OECD Publishing, Paris, 2020, 93 p., https://doi.org/10.1787/1b06c47c-en.

Besançon, L., Peiffer-Smadja, N., Segalas, C. et al.: Open science saves lives: lessons from the COVID-19 pandemic. *BMC Med Res Methodol,* vol. 21(117), 2021, https://doi.org/10.1186/s12874-021-01304-y.

Bieringer, L. et al.: Mental Models of Adversarial Machine Learning. *arXiv:2105.03726 [cs.CR] (accepted at SOUPS 2022),* https://doi.org/10.48550/arXiv.2105.03726.

BMBF: Sample agreements for research and development cooperation - Guidelines for cooperation between the academic sector and industry. 2019, 64 p., https://www.bmwk.de/Redaktion/EN/Publikationen/sample-agreements-for-research-and-development-cooperation.html.

Bodenreider, O.: Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearbook of medical informatics*, 17(01), 2008, pp. 67-79.

Boenisch, F., Battis, V., Buchmann, N., Poikela, M.: “I Never Thought About Securing My Machine Learning Systems”: A Study of Security and Privacy Awareness of Machine Learning Practitioners. In: *MuC '21: Mensch und Computer 2021*, 2021, pp. 520–546, https://doi.org/10.1145/3473856.3473869.

Bogner, A.; Decker, M.; Nentwich, M.; Scherz, C. (Eds.): Digitalisierung und die Zukunft der Demokratie. Beiträge aus der Technikfolgenabschätzung. Nomos Verlag, Baden-Baden, 2022, 288 p., ISBN 978-3-8487-8531-5.

Bonn, A. et al.: Weißbuch. Citizen Science-Strategie 2030 für Deutschland. Version 5.8.2021. Entwurf zur öffentlichen Konsultation, 2021, https://www.citizen-science-weissbuch.de/sites/default/files/downloads/2021-08-05_weissbuch_citizen_science_strategie_2030_entwurf.pdf.

Böschen, S. Grunwald, A., Krings, B.J., Rösch, C. (Eds.): Technikfolgenabschätzung. Handbuch für Wissenschaft und Praxis. Nomos Verlag, Baden-Baden, 2022, 498 p., ISBN: 978-3-8487-6070-1.

Borgwardt A.: Digitalisierung in der Wissenschaft. Friedrich-Ebert-Stiftung, 2018, 70 p., https://library.fes.de/pdf-files/studienfoerderung/14620.pdf.

BMBF: Digitale Zukunft: Lernen. Forschen. Wissen. *Bundesministerium für Bildung und Forschung*, 2019, 39 p., https://bmbf-prod.bmbfcluster.de/upload_filestore/pub/BMBF_Digitalstrategie.pdf.

Chiou, T.: Copyright lessons on Machine Learning: what impact on algorithmic art?. *JIPITEC*, 10(3), 2020, pp. 398-411 (para 1), urn:nbn:de:0009-29-50250.

EU Com/Directorate-General for Research and Innovation: Open Science and Intellectual Property Rights. How can they better interact? State of the art and reflections. 2022, 139 p., https://ec.europa.eu/info/publications/open-science-and-intellectual-property-rights_en.

EU Com/Directorate-General for Research and Innovation/Corcho, O., Eriksson, M., Kurowski, K., et al.: EOSC interoperability framework: report from the EOSC Executive Board Working Groups FAIR and Architecture. Publications Office, 2021, 60 p., https://data.europa.eu/doi/10.2777/620649.

Fathalla, S., Vahdati, S., Auer, S., Lange, C.: Towards a knowledge graph representing research findings by semantifying survey articles. In: *International Conference on Theory and Practice of Digital Libraries*, Springer, Cham., 2017, pp. 315-327.

Globocnik, J., Richter, H.: Rechte an Daten und Datenzugangsrechte. In: *Chibanguza, K., Kuß, Ch., Steege, H. (Eds.): Künstliche Intelligenz - Recht und Praxis automatisierter und autonomer Systeme*, Nomos, Baden-Baden 2022, pp. 93-110.

Graber-Soudry, O., Minssen, T., Nilsson, D., Corrales, M., Wested, J., Illien, B.: Legal Interoperability and the FAIR Data Principles. Zenodo, 2021, https://doi.org/10.5281/zenodo.4471312.

Grunwald, A.: The Responsibility of Researchers and Engineers: Codes of Ethics for Emerging Technologies 2022. In: *Laas, K. (Ed.) Codes of Ethics and Ethical Guidelines*, Springer International Publishing, 2022, pp. 243–258.

Hey, T., Tansley, R., Tolle, K., Gray, J.: The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research, 2009, 252 p., ISBN: 978-0-9825442-0-4.

Hoffmann, J., Begoña, O., Demystifying The Role Of Data Interoperability In The Access And Sharing Debate. *JIPITEC,* 11(3), 2020, pp. 252-273 (para 1), urn:nbn:de:0009-29-51874.

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G.D., Gutierrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., Ngomo, A.C.N.: Knowledge graphs. *Synthesis Lectures on Data, Semantics, and Knowledge*, 12(2), 2021, pp.1-257.

Holl-Supra, S., Rank, S., Wiesenfeldt, S.: Digitalisation Strategy of the Helmholtz Association. Bonn, 2019, 36 p., https://www.helmholtz.de/fileadmin/user_upload/04_mediathek/perspektiven/18_Helmholtz_Strategiebroschuere_EN_Web.pdf.

Hosseini, M., Wieczorek, M., Gordijn, B.: Ethical Issues in Social Science Research Employing Big Data. In: *Science and Engineering Ethics,* 28(29), 2022, 21 p., https://doi.org/10.1007/s11948-022-00380-7.

Huijstee, M. van, Boheemen, P. van, Das, D., Nierling, L., Jahnel, J., Karaboga, M., Fatun, M.: Tackling deepfakes in European policy. *European Parliament Think Tank,* 2021.

Katerbow, M., Kümmel, Ch., Crispin, J., Kerremans, D.: Digital Turn in Science and Humanities. *Position paper of Deutsche Forschungsgemeinschaft*, Bonn, 2020, 16 p., https://doi.org/10.5281/zenodo.4191345.

Koalitionsvertrag 2021- 2025, published by Sozialdemokratische Partei Deutschlands, Bündnis 90 / Die Grünen and Freien Demokraten, Berlin, 2021, 144 p., https://www.spd.de/fileadmin/Dokumente/Koalitionsvertrag/Koalitionsvertrag_2021-2025.pdf.

Leibniz: Leibniz Strategic Forum on Digital Change. Leibniz Association, Berlin, without year, 13 p., https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/Bilder_und_Downloads/%C3%9Cber_uns/Organisation/Dokumente/WhitePaper.pdf.

Lin, D., Crabtree, J., Dillo, I. et al.: The TRUST Principles for digital repositories. *Sci Data,* 7(144), 2020, https://doi.org/10.1038/s41597-020-0486-7.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., Yarkoni, T.: How open science helps researchers succeed. *eLife*, 5, e16800, 2016, 19 p., https://doi.org/10.7554/eLife.16800.

Naiakshina, A. et al.: Why Do Developers Get Password Storage Wrong?: A Qualitative Usability Study. In: *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 311–328, https://doi.org/10.1145/3133956.3134082.

OECD: The Digitalisation of Science, Technology and Innovation: Key Developments and Policies. *OECD Publishing*, Paris, 2020, 120 p., https://doi.org/10.1787/b9e4a2c0-en.

Orwat, C., Bareis, J., Folberth, A., Jahnel, J., Wadephul, C.: Risikoregulierung von künstlicher Intelligenz und automatisierten Entscheidungen. In: *Hoeren, T., Pinelli, S. (Eds.) Künstliche Intelligenz - Ethik und Recht*, Verlag C.H.Beck, 2022, pp. 255-287.

Reinheimer, B. et al.: An investigation of phishing awareness and education over time: When and how to best remind users. *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, 2020, pp. 259-284, https://www.usenix.org/conference/soups2020/presentation/reinheimer.

RfII: Enhancing Research Data Management: Performance through Diversity. Recommendations regarding structures, processes, and financing for research data management in Germany. Rat für Informationsinfrastrukturen, Göttingen 2016, 90 p., urn:nbn:de:101:1-20161214992.

RfII: Digital competencies – urgently needed! – Recommendations on career and training prospects for the scientific labour market. Rat für Informationsinfrastrukturen, Göttingen 2019, 56 p., urn:nbn:de:101:1-201910251023399932679 9.

RfII: Nutzung und Verwertung von Daten im wissenschaftlichen Raum − Empfehlungen zur Ausgestaltung von Datendiensten an der Schnittstelle zwischen Wissenschaft und Wirtschaft. Rat für Informationsinfrastrukturen, Göttingen 2021, 120 p., urn:nbn:de:101:1-2020052673.

Richter, H.: Open Science and Public Sector Information – Reconsidering the exemption for educational and research establishments under the Directive on re-use of public sector information, *JIPITEC, 9(1),* 2018, pp. 51-74 (para 1), urn:nbn:de:0009-29-46794.

Sack, H.: Hybride Künstliche Intelligenz in der automatisierten Inhaltserschließung. In: *Franke-Maier, M., Kasprzik, A., Ledl, A., Schürmann, H. (Eds.): Qualität in der Inhaltserschließung*, De Gruyter Saur, Berlin, Boston, 2021, pp. 387-406. https://doi.org/10.1515/9783110691597-019.

Schomburg S., Leggewie C., Lobin H. und Puschmann C. (Eds.): Digitale Wissenschaft, Stand und Entwicklung digital vernetzter Forschung in Deutschland. *Beiträge der Tagung 2., ergänzte Fassung*. hbz, Köln, 2011, 232 p., https://hbz.opus.hbz-nrw.de/opus45-hbz/frontdoor/deliver/index/docId/206/file/PDFA_Tagung_Digitale_Wissenschaft_hbz_2011_7.pdf.

Schrape, J.-F.: Die Digitale Transformation. Transcript Verlag, Bielefeld, 2021, 264 p., ISBN: 978-3-8252-5580-0.

Smith, B., Ashburner, M., Rosse, C. et al.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol,* 25, 2007, pp.1251–1255, https://doi.org/10.1038/nbt1346.

Specht-Riemenschneider, L., Wehde A.: Forschungsdatenzugang. Rahmenbedingungen, Prinzipien und Leitlinien für einen privilegierten Zugang zu Daten für Forschung und Wissenschaft. *ZGI,* 2022, pp. 3-11.

Surblytė-Namavičienė, G.: Competition and Regulation in the Data Economy. Does Artificial Intelligence Demand a New Balance? Edward Elgar, Cheltenham, 2020, 296 p., ISBN: 978-1-78811-664-0.

Tuladhar A., et al.: An Analysis of the Role of Situated Learning in Starting a Security Culture in a Software Company. In: *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, 2021, pp. 617-632, https://www.usenix.org/conference/soups2021/presentation/tuladhar.

Volkamer, M., Renaud, K.: Mental models–general introduction and review of their application to human-centred security. *Number Theory and Cryptography*, Springer, Berlin, Heidelberg, 2013, pp. 255-280.

Volkamer, N., Sasse, M.A., Boehm, F.: Analysing Simulated Phishing Campaigns for Staff. In: *Computer Security. ESORICS 2020. Lecture Notes in Computer Science*, vol. 12580, Springer, Cham, 2020, pp. 312–328, https://doi.org/10.1007/978-3-030-66504-3_19.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J.: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 2016, pp. 1-9.

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D. and Zhu, J.: Explainable AI: A brief survey on history, research areas, approaches and challenges. In: *CCF international conference on natural language processing and Chinese computing*, Springer, Cham., 2019, pp. 563-574.