Introducing the DiTraRe:
Digital Transformation of Research
in four research clusters interwoven with four dimensions

Leibniz ScienceCampus
Digital Transformation of Research

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

KIT
Karlsruher Institut für Technologie

**ANNA JACYSZYN** AND **HARALD SACK**
WITH FELIX BACH, MATTHIAS RAZUM AND CHRISTIANE NOE
FIZ Karlsruhe, Leibniz Institute for Information Infrastructure
Anna.Jacyszyn@fiz-Karlsruhe.de

← get in touch!

https://ditrare.de/en
← keep an eye on our new website!

gefördert durch
*Leibniz*
Leibniz-Wettbewerb

DiTraRe

Abstract

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. Within the newly established Leibniz ScienceCampus "Digital Transformation of Research" (DiTraRe) we investigate the effects of the increasing digitalisation of scientific work in four research clusters and develop specific solutions based on use cases from various disciplines. Each of these work packages is approached within four different dimensions.
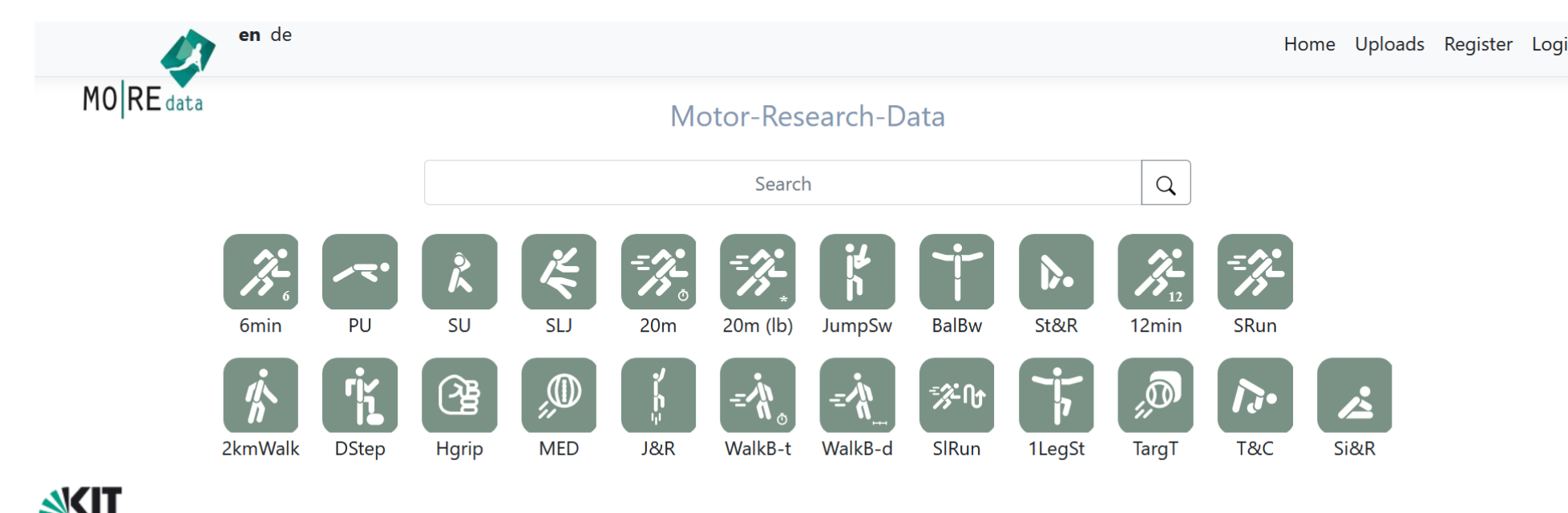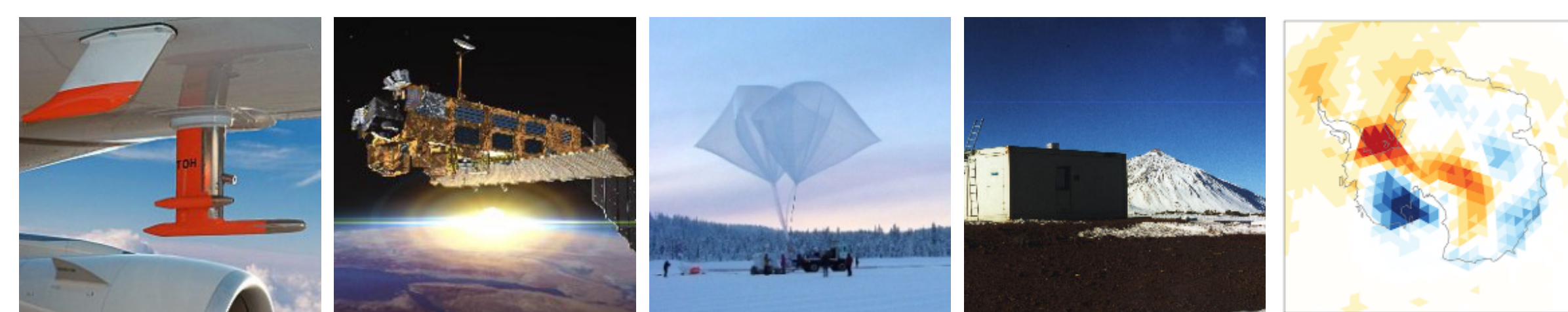
eResearch infrastructure for sports science: motor activity research data
(MO|RE data, KIT IfSS)
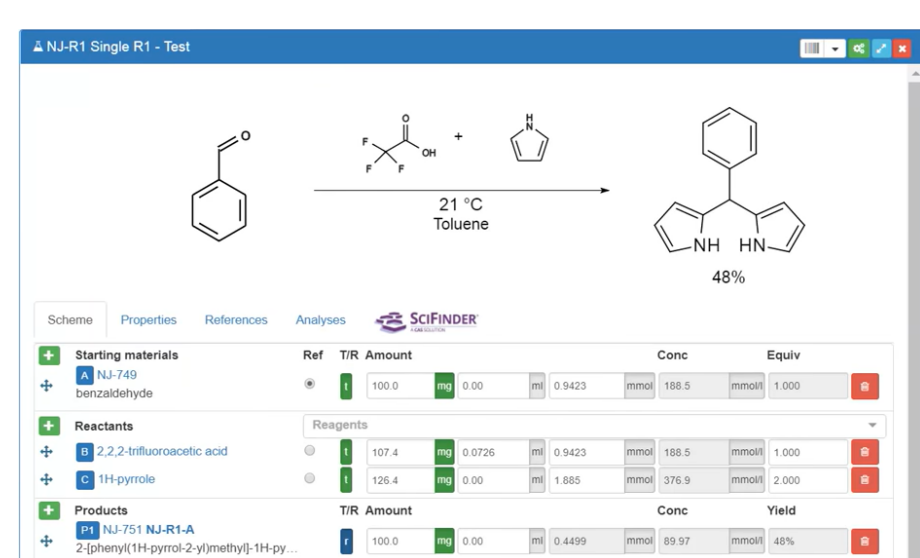
Chemotion Electronic Lab Notebook *(N. Jung KIT IBCS-FMS, Copyright ©2024 Chemotion - KIT. Built with Docusaurus.)*
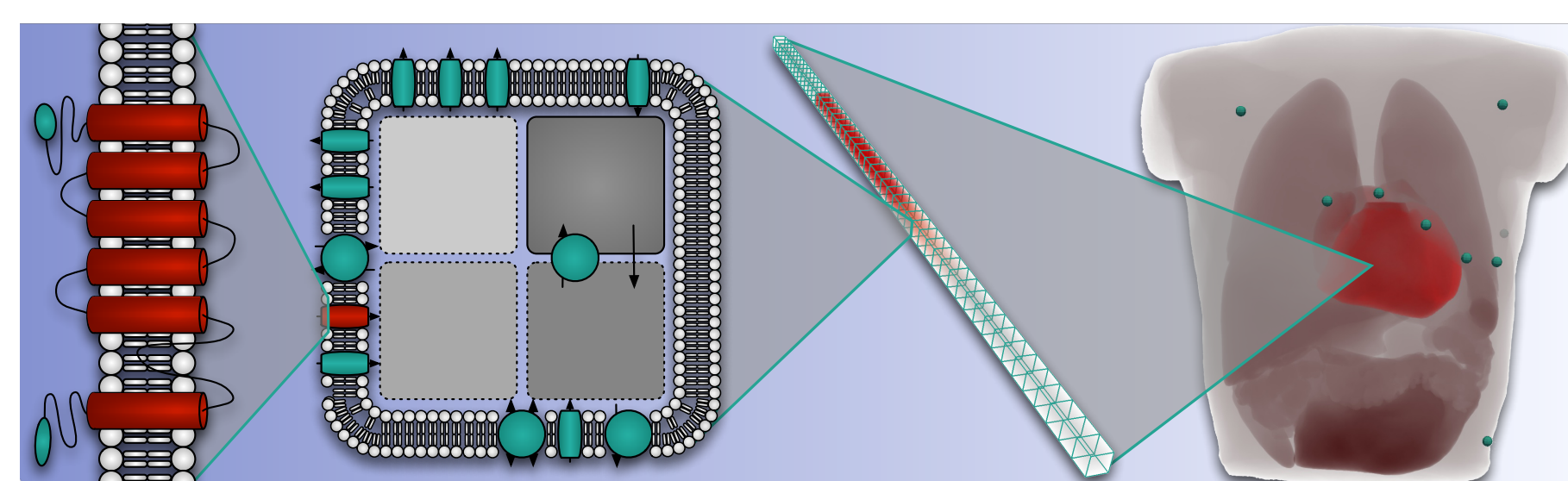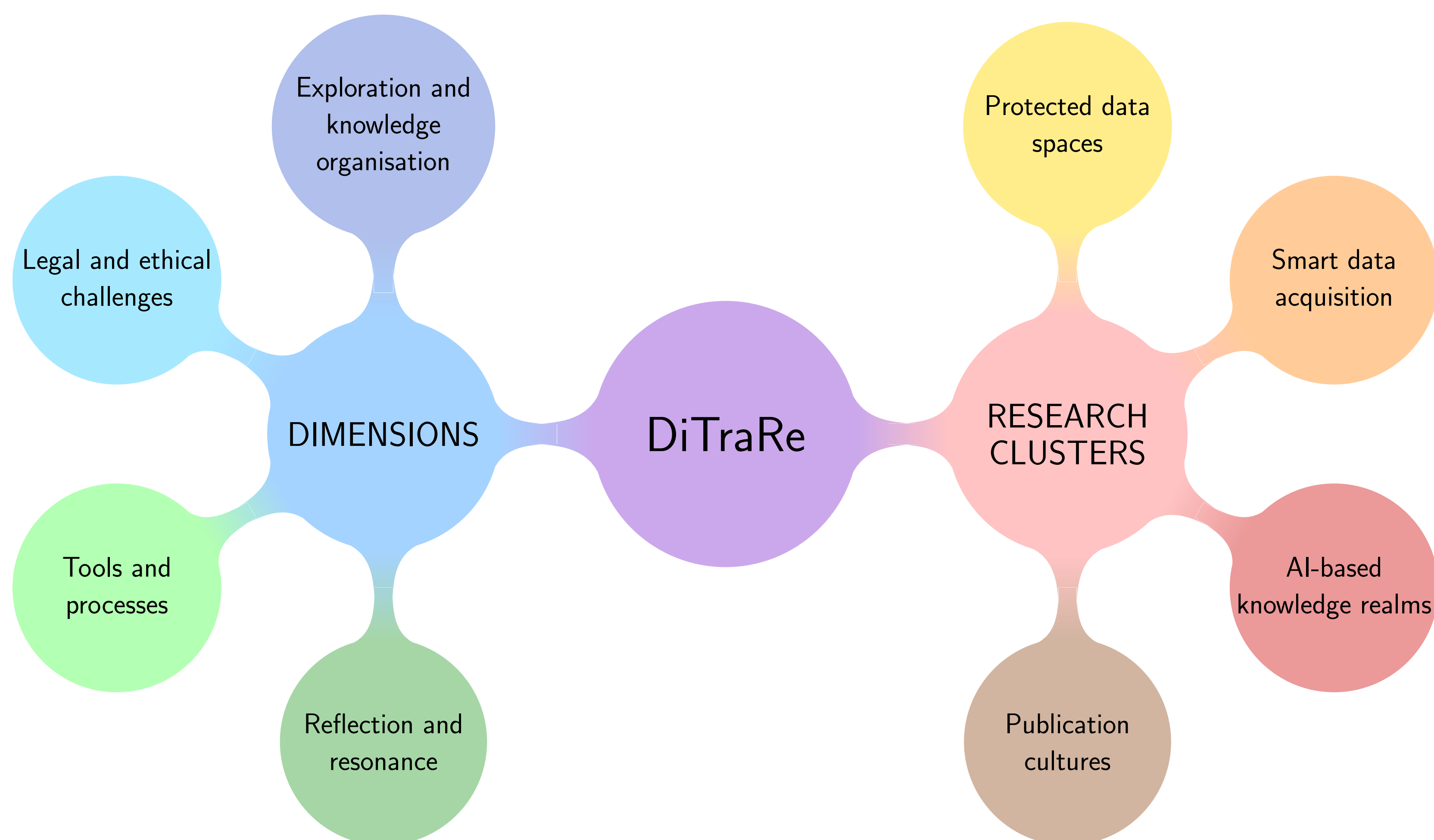

Computerized modeling of the human heart *(O.Dössel, A.Loewe; Copyright ©2017, Elsevier)*

Climate research: atmospheric trace gases and remote sensing *(KIT IMK-ASF)*
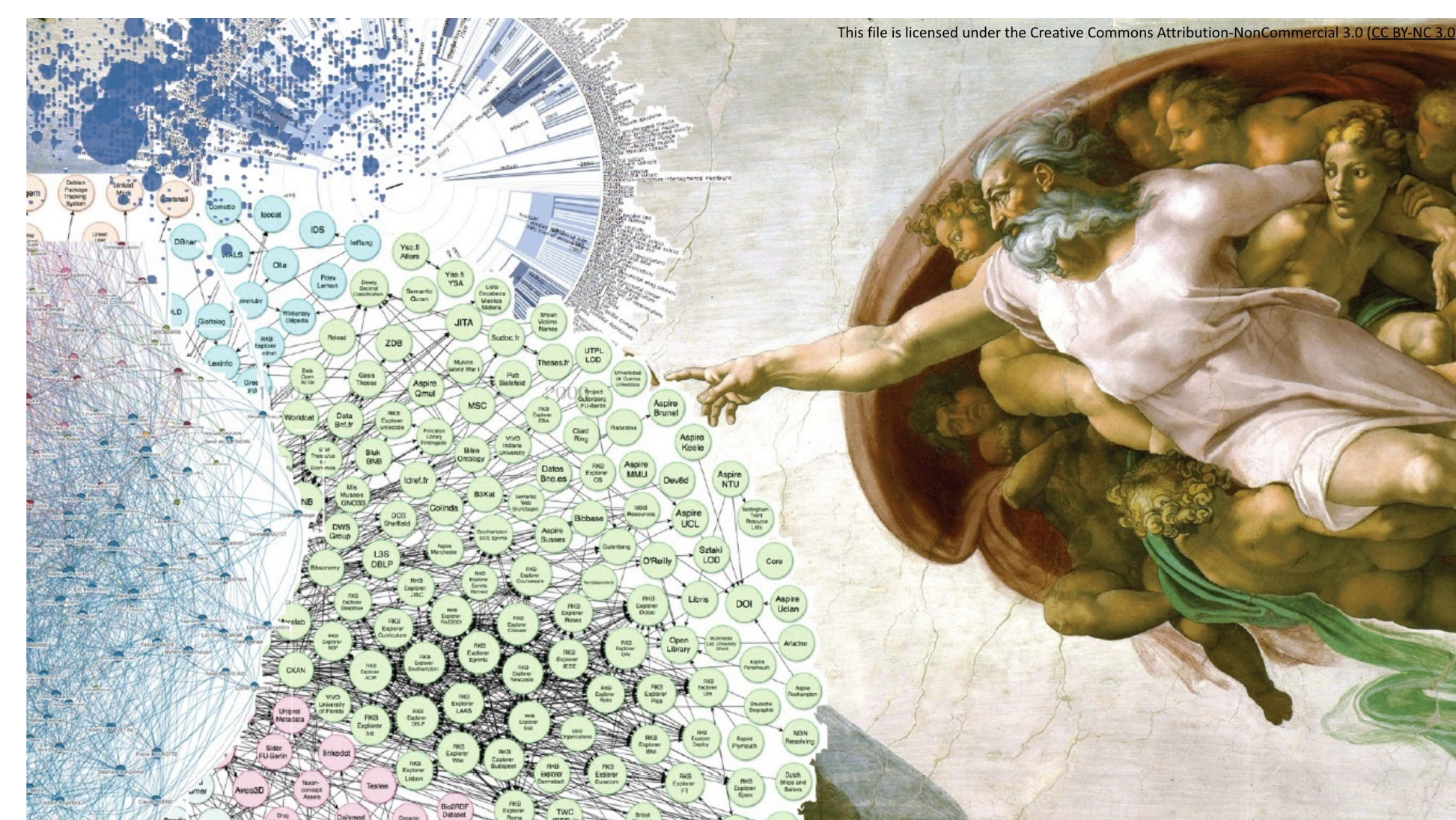
Dimensions and research clusters

► **Protected data spaces: Sensitive data in sports science** *Institute of Sport and Sport Science, Karlsruhe Institue of Technology (KIT IfSS)*
KIT-IfSS research would strongly benefit from linking health with physical fitness data. An overarching concept for the secure handling of sensitive data is lacking.

► **Smart data acquisition: Chemotion Electronic Lab Notebook** *Institute of Biological and Chemical Systems, Karlsruhe Institute of Technology (KIT IBCS)*
Chemistry labs in academia have limited use of lab automation and device integration. Concerns include i.e. dependencies on software and technologies not under control of scientists.

► **AI-based knowledge realms: Artificial Intelligence in Biomedical Engineering** *Institute of Biomedical Engineering, Karlsruhe Institute of Technology (KIT IBT)*
KIT-IBT develops computer models of the human heart to predict cardiovascular diseases. Simulated data are often essential to overcome issues of data privacy and existing bias, but raise questions of explainability of AI decisions and trust.

► **Publication cultures: Publication of large datasets** *Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology (KIT IMK)*
KIT-IMK generates and analyses very large datasets in climate simulations. Publication and re-use of those data is currently very inefficient.

Exploration and knowledge organisation – Information Service Engineering

The *Information Service Engineering Group* **led by Prof. Dr. Harald Sack at FIZ Karlsruhe** is responsible for the dimension "Exploration and knowledge organisation" in the DiTraRe project.

Within the use case sensitive data in sports science our plan is to develop a knowledge graph which will enable sport scientists to easily analyse data and make predictions. We will work with chemists on the Chemotion Electronic Lab Notebook and novel methods of data acquisition. The use case Artificial Intelligence in Biomedical Engineering will profit from our support concerning introducing large language models into their research. We will work with climate researchers on publication of large datasets where we will employ AI techniques for an unchallenging organisation of very large amounts of research data.

A warp or a weft? Interweaving tasks

		RESEARCH CLUSTERS		
DIMENSIONS	**Protected data spaces**	**Smart data acquisition**	**AI-based knowledge realms**	**Publication cultures**
Exploration and knowledge organization	representation of (partially) protected information in open KG and interweaving protected with non-critical data	application of novel methods of data acquisition, analysis and interpretation, and their evaluation	explainability and explanatory components based on symbolic knowledge representation (explainable AI)	massively parallel authoring and quality assurance of large KG
Legal and ethical challenges	data protection and pseudonymisation/anonymisation of different categories of sensitive data, conceptualisation of re-use models	IP legislation and licences for cooperatively created data and data resulting machine-based analyses	potential of synthetic training data for AI systems with regard to the challenges of data protection	impact and analysis of data laws, policies and data strategies, particularly with respect to Open Science
Tools and processes	requirements for secure storage of personal data with regard to integrity, confidentiality, authenticity and availability	automation and virtualisation of laboratory research ("smart lab"), design and implementation of next-generation ELN	necessary infrastructure to generate, store and disseminate (synthetic) training data, with emphasis on interoperability and re-use	publication and exploration of extensive datasets, on new publication formats, e.g. including actionable source code
Reflection and resonance	ethical issues in the handling of sensitive data, e.g. necessary negotiation processes and trade-offs between transparency and data protection	appropriate involvement of society in virtual spaces, e.g. define the conditions for acceptance of data from a stakeholders' perspective	design and communication of decision-making processes in view of the uncertainty of knowledge, e.g. knowledge gained by AI methods	necessity of a cultural change with regard to the needs of research and society for mutual communication, especially in the light of Open Science