## Hida HELMHOLTZ Information & Data Science Academy

HELMHOLTZ INFORMATION & DATA SCIENCE ACADEMY

**REPORT 2021** 



#### I. HIDA IN GENERAL

## II. HELMHOLTZ INFORMATION SCIENCE SCHOOLS

#### III. HIDA ACTIVITIES

- 1. Trainee Network

- Trainee Network \_\_\_\_\_\_
   Talent Scouting & Recruiting
   Education & Training \_\_\_\_\_\_
   Networking Events \_\_\_\_\_\_
   Communication \_\_\_\_\_\_\_
   Finances, Administrative Matters & Personnel \_\_\_\_\_\_
   HIDA Steering Board (HIDA-Ste

## IV. FACTS & FIGURES ON THE HELMHOLTZ INFORMATION & DATA SCIENCE SCHOOLS

1.	DASHH
2.	HDS-LEE
3.	MUDS
4.	HIDSS4Health
5	HEIBRIDS
<u> </u>	
б.	Mardala

V. OUTLOOK

VI. RESPONSIBILITY FOR REPORT / HIDA MAIN CONTACT

& DATA	
	6
	10
	13
	18
	20
	23
	26
eer)	27

#### 28 29 38 49 53 60 64

67

## 68

# I. HIDA IN GENERAL

central component of attracting and promoting excellent young scientists in the field of Information & Data Science is the Helmholtz Information & Data Science Academy (HIDA), founded in 2019. In full extension, HIDA and the Helmholtz Information & Data Science

Schools (HIDSS) will have trained over 280 doctoral researchers until 2025 (incl. associated doctoral researchers as reported by the schools). This makes the program the largest structured postgraduate training program in the field of Information & Data Science in Germany.

HIDA provides networking activities, further education and training programs for Information & Data Science talents of all Helmholtz Centers who combine knowledge in stateof-the-art information processing with scientific knowhow in one of Helmholtz' research programs. HIDA strongly supports the recruitment of the most talented young scientists in the field worldwide for all Helmholtz Centers and their university partners and contributes substantially to the Associations' employer branding efforts.

In 2021, HIDA continued its diverse networking activities and intensified its exchange with the 18 Helmholtz Centers, a growing number of top universities and other partners from science and industry with expertise in the field of Information & Data Science in order to attract and qualify young talents. These efforts aim to promote and intensify collaboration between scientists within the Helmholtz Association and their university partners and to bring additional Information & Data Science expertise to the Helmholtz Centers.

The HIDA education and training initiatives address the shortage of Information & Data Science specialists in both academia and industry and, due to their scale, make a significant contribution to the training of Information & Data Science top talent in Germany.

# II. HELMHOLTZ INFORMATION & DATA SCIENCE SCHOOLS

he HIDA network is the largest postgraduate training program in Information & Data Science in Germany. At the core of this program are the six Helmholtz Information & Data Science Schools, which form a network between 13 Helmholtz Centers and 17 top-tier universities that will have trained over 280 doctoral researchers (incl. associated doctoral researchers) until 2025. The schools

are groundbreaking in the development of new collaborative approaches to evaluating complex, heterogeneous data in the natural sciences with the help of intelligent algorithms - thus enabling modern cutting-edge research. Doctoral researchers are always jointly supervised by one PI from the Helmholtz programs and one PI from the computer science, informatics or mathematics faculties of the schools' partner universities, thus building bridges even beyond the scope of the individual PhD projects.

In 2021, the schools have achieved their ambitious goals in recruitment, training, and publication output in the field of Information & Data Science and have further expanded their activities. Listed below are outstanding highlights of our schools (detailed information on the schools can be found in chapter IV):

- DASHH could win the Hamburg University of Applied Sciences as a new partner that is actively involved in the supervision of ongoing PhD projects. DASHH also partnered up with the Machine Learning in Engineering Initiative (MLE) of the TUHH.
- Throughout the year 2021, 17 Hamburg COVID-19 Lectures were organized by DASHH and a staggering number of almost 1.000 participants took part in the events
- HDS-LEE was invited by the Ministry of Culture and Science of the State of North Rhine-Westphalia to submit a full application of the "BioTwins" draft. This project will be highly benefical for HDS-LEE, as the concept of BioTwins aims to support the training of young data scientists from doctoral researchers to junior research group leaders.
- MUDS doctoral researcher Karin Hrovatin and MUDS associated doctoral researcher Marco Stock received add-on fellowships for interdisciplinary life science by the Joachim Herz Stiftung, an award for talented interdisciplinary young scientists.

- > MUDS was able to recruit 19 additional associated MUDS doctoral researchers.
- > Under the lead of the AWI, and in scientific collaboration with the DLR and the University of Fairbanks, a crew of five researchers including HEIBRiDS PhD researcher Tabea Rettelbach and her supervisor headed to Alaska, USA, in June and July 2021 to conduct an aerial imaging campaign with AWI's Polar-6 aircraft. The goal of the expedition was to acquire very high-resolution images to map and monitor (changing) permafrost landscapes in West Alaska.
- MarDATA's recruitment for the second cohort was a great success: the number of applications have multiplied compared to the first round with a total of 450 applications for 19 projects in 2021 (80 in 2019, 20 projects).
- > HIDSS4Health started an academic partnership with i4health (University College London).
- > HIDSS4Health received additional funding of €1.2m from the Ministry of Science, Research and the Arts of Baden-Württemberg in order to recruit additional doctoral researchers (funding period: 2021 to 2024).

Currently, 252 doctoral researchers are being trained at the schools in Hamburg, Bremerhaven/Kiel, Heidelberg/ Karlsruhe, Berlin, Munich and Cologne/Jülich.



The recruitment at all schools shows a significant increase in received applications over the years.



Fig. 2: Number of received applications across all the schools (2018 - 2021).

Figure 1 shows the total number of recruited doctoral researchers from 2018 up to 2021. The graph highlights thatFigurethe number of recruited doctoral researchers increased1.500steadily from 13 doctoral researchers in 2018 to 252demardoctoral researchers in 2021 (incl. 80 associated doctoral& Dataresearchers, who are funded by third parties or other programs). The recruiting activities are according to plan andThe own nation

Figure 2 summarizes the total number of received applications across all schools from 2018 to 2021. With more than 1.500 applications received in 2021, it is apparent that the demand for training at the interface between Information & Data Science and research domain is high.

The over 1.500 received applications resulted in inter national cohorts of doctoral researchers. Figure 3 shows the 33 countries of origin of the doctoral researchers across all six schools.



Fig. 3: Countries of origin of the doctoral researchers of all six schools



Fig. 4: Gender ratio of all doctoral researchers (excl. associated doctoral researchers).

The special efforts to attract more female candidates resulted in an over all gender ratio of 28% amongst the doctoral researchers over all schools, varying from 20 % (HDS-LEE) up to 39% (MarDATA).

In 2021, 142 publications with contributions from HIDA doctoral researchers were written and published at all schools combined. Details of the publication output of each school are provided in chapter IV. In line with varying subject-specific publication practices, each school has structured and listed its publication data accordingly. The schools' customized curricula offer a wide range of lectures, seminars, workshops, trainings and summer schools. 121 lectures were held at all schools. The lectures are both specific to the subject matter of the schools (e.g. Oliver Jäkel, DKFZ: "Why data sciences will be crucial for modern image guided radiotherapy" at HIDSS4Health) and broader in scope (e.g. Frank Leymann, Uni Stuttgart: "Basics of Quantum Computing and Potentials in Machine Learning" at HEIBRIDS). They also reflect very recent developments as shown by the example of the successful continuation of the "Hamburg COVID 19 Lecture Series" at DASHH with 17 episodes in 2021.

In total, 42 courses took place across all schools in 2021. They aim at providing doctoral researchers with specific skills, exemplified by courses like "Python from 0 to Data Analysis", (HIDSS4Health) or "Machine Learning for Marine Sciences" (MarDATA/HelmholtzAI).

63 recruiting and networking events of various types were organized by the schools, including PhD seminars. This figure also contains events that are part of the schools' regular program, such as MarDATA's Digital Science Mondays.

# **III. HIDA ACTIVITIES**

In 2021, HIDA continued to expand its recruiting, training and networking activities in the field of Information and Data Science. Due to high demand and building on successful formats of 2019 and 2020, we continued and intensified activities such as the Helmholtz Data Science Career Day and launched new activities such as the HIDA Annual Conference or the Lecture Series in cooperation with the Schools and the other Helmholtz Incubator Platforms. HIDA reached an important milestone of its build-up phase by implementing the HIDA Course Catalogue on its website, which is accessible to researchers in the entire Helmholtz Association and its partner institutions.

## HIDA AND THE HELMHOLTZ INCUBATOR PLATFORMS

HIDA is a central component of the Helmholtz Incubator and works in close collaboration with the other Information & Data Science Platforms in order to provide training and networking opportunities for Helmholtz researchers from all centers and programs. The outstanding joint activities were:

- HIDA International Covid Data Challenge

   (28.-29.04.2021): The Incubator platforms Helmholtz AI,
   HIFIS, and HIP, as well as the ELLIS in Genoa and Munich,
   collaborated to contribute the content of the Challenge
   and the supporting program, as well as support.
- > TEACH Talk about Education Across Communities in Helmholtz (07.12.-10.12.2021): Together with the Incubator Platforms HIFIS and HMC HIDA organized the participant-driven event TEACH bringing together 130 Training Coordinators, Instructors and Personnel Developers not only from the Information & Data Science activities but also from from HR departments and central administrations of many centers.
- HIDA Annual Conference (30.11.2021): Workshops from Helmholtz Incubator platforms and partners were in high demand with 124 attendees:
  - » "Disentanglement of bio-related signal from noise for microscopy images" (Tingying Peng, Helmholtz AI Young Investigator Group Leader)
  - » "The next Generation of AI in Medicine" (Shadi Albarqouni, Helmholtz AI Young Investigator Group Leader)
  - » "Solutions for the Ages" (Tobias Huste, HZDR, HIFIS Software Cluster Management)

- » "FAIR enough?! 5 steps to boost the value of your research data" (Dr. Mirl Trösch, GEOMAR, Coordination Training HMC)
- HIP: The Helmholtz Imaging Platform is funding the project "Tackling the segmentation and tracking challenges of growing colonies and microbial diversity (SATOMI)" led by HDS-LEE PI Prof. Wolfgang Wiechert (FZJ), together with Prof. Hanno Scharr (FZJ) (funding period 05.2021-04.2023). The participating doctoral researcher Johannes Seiffarth from HDS-LEE connetcs the school and the Imaging Platform.
- > HIFIS: 4 courses took place in collaboration with HIFIS attracting 92 participants, two at HDS-LEE ("Introduction to Git and GitLab" & "Bring your Own Script and Make It Ready for Publication"), two at HIDSS4Health ("Gitlab" workshop & "Python from 0 to Data Analysis").

#### Helmholtz AI:

- » Niki Kilbertus from Helmholtz AI was invited as a speaker to the HIDA Lectures @MUDS (29.09.2021, 83 participants)
- » In cooperation with HDS-LEE doctoral researchers Helmholtz AI organized the "Hearts Gym" Hackathon (20.-23.07.2021, 38 participants) as a HDS-LEE internal workshop
- » Two doctoral researchers took part in the Helmholtz AI Virtual Conference 2021 (Karel van der Weg, HDS-LEE and Vladyslav Fediukov, MUDS)
   MarDATA cooperated for a workshop on "Machine Learning for Marine Sciences" with Helmholtz AI at DKRZ
- » The HIDA ML Summer School took place jointly with Helmholtz AI (Jul/Sep 2021)

#### THE COVID-19 PANDEMIC IMPACT

In 2021, the COVID-19 pandemic still had a considerable impact on HIDA's and the schools' activities and strongly reduced exchange programs and physical events. In addition, it became apparent that the reduced exchange opportunities were limiting some doctoral researchers. As a result the progress of some projects was delayed.

The pandemic, its prevention and the corona virus as such, however was scientific subject of numerous highly frequented, virtual lectures, seminars, events and publications:

- In April, the HIDA International COVID Data Challenge took place as a virtual event via Zoom and Slack. 63 participants from 15 countries across five time zones competed to improve an algorithm that uses X-ray images to predict the likely course of disease for COVID-19 patients.
- > DASHH initiated a Hamburg COVID-19 Series in collaboration with the Leibniz Science Campus InterACt. The COVID-19 Series is a platform for scientists in the Hamburg metropolitan area to present recent challenges and results from the field of COVID-19 research to increase the visibility of this research fostering new contacts and potential collaborations. In 2021, the Hamburg COVID-19 Lecture Series took place 17 times with 962 participants in total.
- Dr. Andrea Thorn, group leader at the Institute for Nanostructure and Solid State Physics, Universität Hamburg, and leader of the "Coronavirus Structural Task Force" joined DASHH as a new DASHH Scientist.
- Soon after the beginning of the pandemic, researchers from the Heidelberg Institute for Theoretical Studies (HITS) and the KIT developed a web platform for bundling short-term forecasts of COVID-19 cases and deaths in Germany and Poland with contributions by numerous independent modelling teams. Doctoral researcher Daniel Wolffram (HIDSS4Health) was involved in this COVID-19-forecast-hub (for more information please see press release: "More than the sum of its parts: Combining models to improve COVID-19 forecasts", https:// www.h-its.org/2021/09/29/covid-19-forecast-hub/).

- > Doctoral researchers from HDS-LEE, DASHH and HIDSS4Health were involved in six publications on the subject of COVID-19 resp. SARS-CoV-2 (1 HDS-LEE / 3 HIDSS4Health / 2 DASHH, please see chapter IV).
- Marc Horlacher, doctoral researcher at MUDS, addressed the "Computational mapping of the human SARS-CoV-2 protein-RNA interactome" in a virtual poster presentation in March 2021.

The COVID-19 pandemic was not only subject and topic of numerous trainings, events, courses and publications; it also affected the way these activities took place. Although the schools managed to improve their recruiting activities and the number of received applications across all the schools increased strongly; the restrictions brought by the pandemic affected the way the schools and HIDA operated and the doctoral researchers participated in the scheduled activities.

Again in 2021, nearly all of the offered trainings and events were held online and meetings in person were rare opportunities for doctoral researchers to connect and exchange. In Munich, the **MUDS** managed to offer the following two in person events:

- MUDS Networking Event, 07.07.2021, in person event, 48 participants, scientific networking across different institutions within the MUDS community, MUDS questions/ answers session
- MUDS Doctoral Symposium, 28.09.2021, in person event, 50 participants, doctoral researcher & PI talks, poster session, get together

Also, the H<sup>3</sup> Helmholtz Herbst Hackathon (08.-12.09.2021) - a collaboration project with Digital Earth, MarDATA, HDS-LEE, Jülich Challenges, Helmholtz AI@FZJ, HAICORE@ JSC - took place in person in Gummersbach. 46 Helmholtz researchers from different research areas participated and exchanged ideas. Despite the limited opportunities for doctoral researchers to network and connect, the schools and HIDA worked hard in order to make their online trainings and networking opportunities as easily accessible as possible. Switching to virtual formats had the advantage that a broader audience could be addressed. This can be demonstrated by the following examples:

- The Hamburg COVID-19 Lectures organized by DASHH addressed almost 1000 participants in total over the year 2021.
- The HIDA International Covid Data Challenge attracted 63 participants from 15 countries across five time zones.
- The Helmholtz Data Science Career Day is HIDA's virtual international career day for data scientists and took place for the second time in 2021. 600 participants from 75 countries visited the event.
- In January 2021, the DASHH lecture "The Inductive Bias in Machine Learning" by Prof. Ulrike von Luxburg attracted 253 participants.

In the post-pandemic future, HIDA will further expand its own and the schools' virtual activities in recruitment, joint events, and networking, utilizing the experiences made during the COVID-19 pandemic. HIDA will thus continue to raise the profile of Helmholtz as the place to be for international top talents that aim to apply cutting-edge Information & Data Science technologies to solve great societal challenges.

#### **1. TRAINEE NETWORK**

The HIDA Trainee Network is a Helmholtz-wide exchange program and a key element of HIDA: The program promotes one- to three-month- long research stays at different Helmholtz Centers for doctoral and postdoctoral researchers (trainees) whose research has a strong connection to (applied) information or data sciences. With the Trainee Network, HIDA created new forms of collaboration and exchange between scientists of the different Helmholtz Centers. The Trainee Network promotes the mutual transfer of expertise in the field of Information & Data Science in the long term and ensures that methods and algorithms are shared between groups from different Helmholtz Centers and research domains while the hosting labs gain additional recruiting opportunities.

Since its establishment in 2020, the Trainee Network was heavily restrained by the ongoing pandemic, which prohibited travel and physical exchange almost entirely. Only a small fraction of the envisaged numbers of mobility grants could be awarded. However, a total of 45 postdocs and doctoral researchers were awarded the grant up until the end of 2021. So far, 25 trainees participated in the exchange. Three successful applicants had to resign from the exchange due to the ongoing COVID-19 pandemic. The remaining 17 exchanges will take place in 2022.

HIDA is still convinced that the Trainee Network will develop into a great success story once the pandemic and connected travel restraints and planning difficulties fade out. Promising developments at the beginning of 2022 show that the Trainee Netzwork is attractive to Helmholtz researchers: The Trainee Network sub-page of the HIDA website is highly frequented (on average 1000 page views per month since January 2022), the information events for Helmholtz Centers and PIs were very well received (three info events with 186 participants in total), and the number of Helmholtz PIs that requested to be listed as potential host surpassed 50 by now.



Three examples out of 45: Marcel Tiepelt (KIT), Sabine Egerer (Hereon) and Ngai Ham Chan (GFZ Potsdam) successfully completed their HIDA Trainee Exchange in 2021. Reports of their convincing scientific exchanges can be read on the HIDA website. (Photo: M. Tiepelt / S. Egerer / N.H. Chan)

#### **TRAINEE NETWORK EXCHANGES IN 2021**

### **CALL FOR APPLICATIONS FOR THE TRAINEE NETWORK IN 2021**

No.	Doctoral	Home	Host	Project		
	researcher	Center	Center	Title		
1	Anna Aschenbrenner	DZNE	Helmholtz Munich	Construction of a single-cell transcriptome reference for human blood and its application for disease contextualization		
2	Milad Asgarimehr	GFZ	DLR	Hurricane and severe weather tracking: a forecasting model using GNSS Reflectometry measurements based on deep learning		
3	Johannes Boog	UFZ	GFZ	DATA-driven CALibration of Large-Scale Physical Groundwater Flow Models using Meta- Modeling and Visual Analytics (DataCal)		
4	Ngai Ham Chan	GFZ	AWI	Improving the Resolution of Satellite Data for Arctic Permafrost Features Identification and Tracking		
5	Sabine Egerer	HZG	UFZ	Towards a statistical approach for agricultural yield production to evaluate climate change adaptation measures for present and future conditions.		
6	Riccardo Fellegara	DLR	GEOMAR	Real-time visualization and analysis of large-scale marine data		
7	Laura Gassner	KIT	DLR	Machine Learning for Signal Analysis of Coupled Elastic-Acoustic Vibrational Emissions		
8	Flavia Gesualdi	KIT	DKFZ	Exploring dose-uncertainty probability distributions in Analytical Probabilistic Modeling of dose-volume histograms and its impact in radiation optimization		
9	Florian Grünschläger	DKFZ	MDC	Automated single-cell multi-omics pipelines for integrated prognosis, diagnosis and therapy prediction in hematological cancers		
10	Robin Gutsche	FZJ	DKFZ	Development of a multitask deep learning radiomics image network		
11	Johannes Holke	DLR	FZJ	GPU based chemistry simulations in volcanic ash clouds		
12	Paul Jerabek	HZG	KIT	Development of a Digital Workflow linking Atomistic and Mesoscopic Simulations in the Data Infrastructure Kadi4Mat		
13	Franz Kaiser	FZJ	KIT	Identifying critical infrastructures and predicting cascading failures in future power systems		
14	Martin Kühn	DLR	HZI	Data-based optimization of advanced mathematical models for simulating the spread of infectious diseases		
15	Eric Lindberg	MDC	Helmholtz Munich	Cellular and molecular drivers of cardiac remodeling in healthy and failing human hearts		
16	Monika Litvinukova	MDC	Helmholtz Munich	Understanding the roles of polyploidy and differential isoform usage during ageing using single cell transcriptomics.		
17	Shuai Ma	HZDR	DESY	The application of deep learning on slit-scan images processing and transverse emittance prediction		
18	Shammi More	FZJ	KIT	Convolutional Neural Networks for Brain-age Estimation: transfer-learning and rank-consistency		
19	Nikolaos Nikolaou	Helmholtz Munich	HZG	Air temperature modelling and the investigation of adverse temperature effects on cardio-metabolic disease applying various machine learning techniques		
20	Leonardo Rydin	FZJ	DLR	PowerDynSys - Power Dynamical Systems, a python package for dynamical system stability in power-grid systems.		
21	Artem Smirnov	GFZ	FZJ	Global modeling of the ionosphere dynamics using deep learning		
22	Marcel Tiepelt	KIT	DLR	Post-Quantum Authentication and Encryption for Aeronautical Communication Systems		
23	Peter Tillmann	HZB	DLR	Deep learning based cloud detection and segmentation from sky images for irradiance forecast applications.		
24	Josef Vaas	DKFZ	HZI	Development of a high-throughput computing workflow to discover novel large DNA viruses from human sequencing data		
25	Hu Zhao	FZJ	DLR	Python-based Package for Probabilistic Simulation and Uncertainty Quantification for Earth Surface Processes		

The second round of applications (application deadline March 2021) closed with 14 applications. After review by the HIDA Steer committee, 12 of the applications were granted. Eight of the successful applicants were doctoral researchers while four of them were postdocs. In total, 11 of the 18 Helmholtz Centers were involved in this second round of applications as the sending or receiving institution. As sending institutions (Home Center) took part: KIT, FZ Jülich, MDC, GFZ, UFZ, HZDR, DLR and DKFZ. As host institutions (Host Center) took part: DLR, MDC, KIT, GFZ, Helmholtz Munich, HZI, FZ Jülich, DKFZ and DESY.







Fig. 5: The HIDA Trainee Network had two rounds of applications in 2021: The charts provide information about the applications received, the participating home and host centers, and the cross-disciplinary exchange.

For the third round of applications (application deadline September 2021), HIDA received seven applications for the Trainee Network. All of the applicants received funding after review by the HIDA Steer committee. Eight Helmholtz Centers were involved as sending or receiving institution. As sending institutions (Home Center) took part: Hereon, UFZ, DKFZ, FZ Jülich and Helmholtz Munich. As host institutions (Host Center) took part: KIT, DLR, Helmholtz Munich, MDC, UFZ and DKFZ.

#### HELMHOLTZ VISITING RESEARCHER GRANT

The Helmholtz Visiting Researcher Grant was established in 2021. It complements the Trainee Network Grants and enables short-term research stays with a duration of one to three months for researcher from outside Helmholtz who work with a strong link to (applied) data and information science. Postdocs from universities, research organizations and industry worldwide are eligible to apply for this grant. For the first round of applications (application deadline in June of 2021), HIDA received ten applications of which seven were approved by the HIDA steer committee. The successful candidates come from Russia, Italy, Germany, Hong Kong, Ukraine and France and will go to five different Helmholtz Centers (FZ Jülich, Hereon, GFZ, HZI, Helmholtz Munich).

The following exchanges are planned between February and October 2022.

Visiting Researcher's Home Institution	Visiting Researcher's Host Center		
<b>University of Bari Aldo Moro,</b> Dipartimento di Scienze Mediche di Base, Neuroscienze e Organi di Senso (SMBNOS), Italy	Institute of Neuroscience and Medicine, <b>FZ Jülich</b>		
International Max Planck Research School on Earth System Modelling (IMPRS-ESM), Max Planck Institute for Meteorology, Germany	Institute for Advanced Simulation (IAS) Jülich Supercomputing Centre (JSC), <b>FZ Jülich</b>		
<b>National University of Science and Technology MISIS,</b> Moscow Institute of Steel and Alloys, Russia*	Institute of Surface Science, Hereon		
<b>University of Osnabrück,</b> Working Group Remote Sensing and Digital Image Analysis, Institut für Informatik, Germany	Section Hydrology, <b>GFZ</b>		
<b>University of Ostrava,</b> Institute for Research and Applications of Fuzzy Modeling, Ukraine	Biostatistics Project, <b>HZI</b>		
<b>University of Tübingen,</b> Department of Linguistics, Germany	Helmholtz AI, <b>Helmholtz Munich</b>		
Institut Curie - Research Center, Systems Biology of Cell Polarity and Cell Division, France	Institute of Energy and Climate Research - Microstructure and Properties of Materials, <b>FZ Jülich</b>		

\*In view of the BMBF's directive to discontinue or critically review scientific cooperation with Russia against the backdrop of the Russian attack of Ukraine, this research stay cannot take place as planned for the time being (Status February 2022).

#### 2. TALENT SCOUTING & RECRUITING

#### **EMPLOYER BRANDING & RECRUITING**

In 2021, HIDA's employer branding and recruiting activities addressed scientists worldwide. An initial campaign via social media and Google advertising directed users to the HIDA job database more than 10.400 times. HIDA summarizes all vacancies in the field of Information & Data Science at all Helmholtz Centers in this database. A second Google Ads campaign is ongoing.

HIDA organized the Helmholtz Data Science Career Day for the second year in a row. More than 600 data scientist from 75 countries took part in the event (further details on the Helmholtz Data Science Career Day can be found in chapter 4. *Networking Events*). Furthermore, HIDA represented the Helmholtz Association at online career fairs such as HackerX (30.06. & 23.11.2021), raising awareness for the Association's Information & Data Science research and employment opportunities nationally and internationally.

#### INTERNATIONAL EXCHANGE PROGRAMS

HIDA makes data scientists internationally aware of Helmholtz as a leading employer where they can apply their Information & Data Science skills in order to provide solutions to the central questions of our time. International exchange programs with selected partners serve as a tool to introduce talent to Helmholtz and create opportunities for ongoing research collaborations with strategic partners. In 2021, HIDA expanded its exchange program with one strategic partner, Israel, by joining forces with the Israel Data Science Initiative (IDSI). IDSI coordinates the data science research centers at Israel's research universities. 19 data scientists, 10 from Israel and 9 from Germany, took part in this year's exchange program. Researchers from 8 Helmholtz Centers were involved in the exchange (as home centers: DKFZ, KIT, Helmholtz Munich, FZ Jülich; as host centers Hereon, FZ Jülich, DLR, HZDR, CISPA). Despite the ongoing pandemic, most Israeli and German participants were able to travel to their destination country, with only some trips postponed to 2022.

HIDA launched a pilot internship program with students from Princeton University. The entire exchange was virtual. Nevertheless, it convinced organizers, researchers and students alike: Nine Princeton bachelor students with impressive data science skills worked on 10 projects together with experienced data scientists from six Helmholtz Centers

(DLR, FZ Jülich, Hereon, Helmholtz Munich, HZDR and UFZ). For 2022, Princeton plans to send students on trips to various Helmholtz Centers.

### FRIENDS OF HIDA NETWORK

We founded the Friends of HIDA network in late 2019 with the vision to connect HIDA within a worldwide network of institutions to complement education and training in the applied Information & Data Science at Helmholtz. The network adds value to and works synergistically with the Helmholtz Centers, complementing their training offers and introducing talent worldwide to Helmholtz. Within the network, partners collaborate on joint Information & Data Science education and training activities, keep each other abreast and open their programs up to each other's communities where possible. HIDA organizes exchange programs with selected partners.

#### By end of 2021, Friends of HIDA included the following:

- > AI Campus Berlin
- > Al Sweden
- Akademie für Theater und Digitalität
- Alexander von Humboldt Institute for Internet and Society (HIIG)
- BIOMEDAS, Academy of the Translational Alliance of Lower Saxony (TRAIN)
- Data Science Research Center at Ben-Gurion University of the Negev (DSRC@BGU)
- German Academic Exchange Service (DAAD)
- > Israel Data Science Initiative (IDSI)
- Norwegian Artificial Intelligence Research Consortium (NORA)
- > Weizmann Artificial Intelligence Center (WAIC)
- > Y-DATA Yandex School of Data Analysis (YSDA)

Joint activities in 2021 included exchange programs with IDSI and the Akademie für Theater und Digitalität. HIDA also organized a hackathon together with IDSI. For the second year in a row, the DAAD's Postdoc-Net-AI tour that offers talent from around the world a personalized gateway to the German AI community, has also included Helmholtz Centers in its activities. In return, many partners participated in the Helmholtz Data Science Career Day.

#### **3. EDUCATION AND TRAINING**

#### **HIDA COURSE CATALOG**

Together with the relaunch of the HIDA website, the interactive HIDA Course Catalog was published https://www.helmholtz-hida.de/course-catalog. On the one hand, Helmholtz researchers can browse the catalog for Information & Data Science related training offers and events offered at the 18 centers and our schools. On the other hand, any member of the Helmholtz Association can publish and advertise effortlessly Information & Data Science events at his/her center throughout the association and beyond. Visitors can filter by tag, center, date or use the search function. The HIDA Course Catalog is an important milestone towards community-wide sharing and networking Information & Data Science related training opportunities and events.

HIDA HELMHOLTZ Information & Data Science Academy Horizonte erweitern	Lernen & vernetzen	Forschen HIE	A entdecken Jobs
HIDA COURSE CATALOG		- y - 1	+ ADD EVENT
TAGS HEI	MHOLTZ CENTE	RS DATE	mark of
Programming (11) Software (11) HIFIS (10)	and a man	From	
Programming Languages (9) Statistics (8)	O ANI O DEST		
Open Science (7) Open Research Software (6)		SEARCH	
Helmholtz Al (5) Reproducibility (5)			٩
AI (3) Data Analytics (3) Regression (3) MI (2)	O urz O re	TYPE	
Open Research Data (2) Small Datasets (2)	O DAR	hybrid	on site
Testing (2) Unsupervised Learning (2)		online	A AND THE REAL PROPERTY AND
Algorithms (1) Big Data (1) Data Cleansing (1)	O Internate Musico	RESEAR	CH AREAS
Data Visualization (1) Database (1)	hanno	Aeronautics,	Space Earth and Environ-
Deep Learning (1) HEIBRIDS (1) HIDA (1)		and Transport	t ment
HIDSS4Health (1) HPC (1) MUDS (1)		Information &	Data 🔲 Matter
Quantum Technologies (1)	ALL OF THE R.	Science	
TITLE \$	DATE 🗘	ADMISSION	FORMAT
Enführung in Dutkon	7.14 0000	Heimholtz & partner	<b>0</b>
	7 Mar 2022	only	Course
↓ Introduction to Scalable Deep Learning	14 - 18 Mar 2022	Public	Course
Reproducibility for Matter Research - Re- producibility in Research by R4E	17 Mar 2022	Helmholtz & partner only	Workshop
Reproducibility for Matter Research - Foun- dation of Research Software Publication	17 Mar 2022	Heimholtz & partner only	Workshop
↓ Data Science Jobs for Health @ Helmholtz	17 Mar 2022	Public	Career Day
Introduction to Shall + Git + Gitl ab	19 21 Mar 2022	Helmholtz & partner	Course
<ul> <li>Introduction to onen + ort + offLab</li> </ul>	10 - 2 I Wal 2022	only	Course
↓ Helmholtz GPU Hackathon 2022	21 - 31 Mar 2022	Public	Hackathon
			_
<ul> <li>Data Carpentry: Ecology with Python</li> </ul>	29 - 30 Mar 2022	Public	Course
Foundations of Research Software Publica-	29 - 30 Mar 2022	Heimholtz & partner	Warkshop
tion	27 . 30 Widi 2022	only	**********
Helmboltz Al Virtual Roadshore	30 Mar 2022	Helmholtz & partner	Discussion
	30 Wai 2022	only	Discussion

#### **HIDA COURSE FUNDING**

HIDA Course Funding https://www.helmholtz-hida.de/en/ learn-connect/course-funding/ enables dedicated funding & development of interdisciplinary and community-oriented training opportunities in cooperation with Helmholtz institutions, as well as national and international partners. HIDA thus stimulates the exchange of methods as well as the transfer of knowledge and research data sets in the field of Information & Data Science between all 18 Helmholtz Centers. In order to give researchers from all centers access, HIDA co-finances training formats within the Helmholtz Association. Requirements for granting HIDA Course Funding are the advertisement of such events throughout the association and eventually beyond. Events have to be suitable to support the 18 Helmholtz Centers in developing and recruiting young talent and to make the Helmholtz Association visible as an attractive employer in the field of Information & Data Science.

Within this framework, HIDA has tested several different formats:

#### LIVE ONLINE WORKSHOPS

Teaching programming, ML, data management, analysis, visualization to Helmholtz researchers. So far, approx. 450 participants from all Helmholtz Centers took part.

- > Instructor-Training with The Carpentries (Jun 2021)
- Nature Research Workshops on Open Data with DASHH (Jun 2021)
- R for Reproducible Scientific Analysis with The Carpentries (Jul 2021)
- Introduction to R with Core Facilities Statistical Consulting at Helmholtz Munich (Jul 2021)
- HIDA ML Summer School with LMU, MCML, Helmholtz AI, MUDS (Jul/Sep 2021)
- Programming with Python with HEIBRiDS and The Carpentries (Aug 2021)
- Graphics with R with Core Facilities Statistical Consulting at Helmholtz Munich (Oct 2021)
- Programming with R with HEIBRiDS and The Carpentries (Oct 2021)
- Programming with Python with The Carpentries (Dec 2021)

#### NETWORKING & TRAINING ON SITE: H<sup>3</sup> HELMHOLTZ HERBST HACKATHON (08.-12.09.2021)

A collaboration project with Digital Earth, MarDATA, HDS-LEE, Jülich Challenges, Helmholtz AI@FZJ, HAICORE@JSC. 46 young Helmholtz researchers from different research areas worked together. To solve five data challenges - a collection of exciting datasets and research questions from various domains, which were put together by a team from FZ Jülich and released as Jülich Challenges Platform - the participants split up in nine interdisciplinary teams. For solving the challenges, participants could use the computational power of JUWELS booster, the flagship of the Jülich Supercomputing Centre, which is ranked No. 8 in the current Top500 list. For many researchers, it was their first contact with a supercomputer. Keynotes and evening poster sessions enabled more networking and enriched discussions.





Real world experience for the participants: The Helmholtz Herbst Hackathon was one of the few events in 2021 that could take place onsite. (Photo: Ramona Kloß /HDS-LEE)

### TRAIN THE TRAINER WORKSHOPS

As a member organization in the international project **The Carpentries,** which is dedicated to teach computer skills to researchers, HIDA benefits from train-the trainer workshops. In these workshops, selected Helmholtz-affiliates will learn how to train and on-board new Carpentries instructors. Having Carpentries trainers "inhouse" will allow HIDA to directly finetune instructor trainings according to the needs within the Helmholtz community and further expand workshop offers for all Helmholtz researchers. Seats for theses workshops are evenly distributed among the centers.

#### COMMUNITY BUILDING: TEACH - TALK ABOUT EDUCATION ACROSS COMMUNITIES IN HELMHOLTZ (07.12.-10.12.2021)

Talent management in Helmholtz is dedicated to personal and professional development for researchers of all qualification levels, ranging from undergraduate students to established scientists at all Helmholtz Centers. The development and management of training programs and courses for doctoral researchers, Postdocs or research groups requires skills and resources. Together with Helmholtz Federated IT Services (HIFIS), Helmholtz Metadata Collaboration (HMC), and the Helmholtz Open Science Office, HIDA initiated and implemented the participant-driven event TEACH. TEACH brought together 130 Training Coordinators, Instructors and Personnel Developers from Information & Data Science Schools and platforms as well as from HR departments and central administrations of many centers. Their goal was to exchange experience and best practices, collaborate and share resources regarding recruiting, talent management, the whole education life cycle, and specific training programs. The mobilization and activation of teaching and learning potentials as well as this specific community building across the association and the facilitation of open exchange within Helmholtz generated positive feedback.



The Gathertown platform offered a real conference atmosphere for TEACH - with virtual booths and meeting points.

#### **4. NETWORKING EVENTS**

## CAREER TALK: WOMEN IN DATA SCIENCE (14.04.2021)

On 14.04.2021, HIDA organized the international virtual online Career Talk "How to become a Data Scientist? Career Paths of Women in Data Science" with over 180 participants as a pre-event of the Gender Summit 2021. Experienced female researchers discussed biographies and careers together with young scientists from the field of Information & Data Science. The five speakers were: Narges Ahmidi (Helmholtz Munich), Ninja Marnau (CISPA), Hela Mehrtens (GEOMAR), Xiaoxiang Zhu (DLR), Nicole Ludwig, University of Tübingen.

#### BIG-DATA.AI SUMMIT (21.-22.04.2021)

From April 21-22, 2021, HIDA participated in the Big-DATA. AI Summit (#BAS21) of the digital association Bitkom. HIDA was partner of the event with a digital partner area and successfully placed a digital speaker slot in the conference program. The goal was to make Helmholtz widely known as a top employer and HIDA and the six Information & Data Science Schools for talent acquisition in the field of Information & Data Science. The international program included 70 talks, panels and workshops with more than 120 speakers and over 3.000 participants. In the digital partner area, visitors learned about the six Information & Data Science Schools as well as HIDA events and courses, job opportunities and funding and exchange programs. Bitkom integrated HIDA on the event website, into event mailings and Bitkom's social media communication. More than 80 participants visited the partnership area during the event. The Helmholtz Association was represented in the conference by Prof. Lena Maier-Hein, Head of the Department Computer Assisted Medical Interventions at the DKFZ, in the topic area "Digital Health, Medical Research, Diagnostics and Therapy" with: "Comparative validation of AI algorithms: common pitfalls and new concepts". She presented both HIDA and HIP in the area of Artificial Intelligence/Data Science and Imaging. The slot was very well attended with 203 participants.

## HIDA INTERNATIONAL COVID DATA CHALLENGE (28.-29.04.2021)

In April, the HIDA International COVID Data Challenge took place as a virtual event via Zoom and Slack. 63 participants from 15 countries across five time zones competed to improve an algorithm that uses X-ray images to predict the likely course of disease for COVID patients.

The competition brought together Helmholtz data scientists from 11 centers for cross-platform and cross-disciplinary collaboration, and provided young data scientists from 21 other institutions in Germany and abroad with an attractive insight into working on grand challenges at Helmholtz. Finally, the winning teams' solutions contributed to the improved applicability of the algorithm in clinical research. HIDA's partners from academia and industry helped connect data scientists in the community and establish new contacts for scientific collaboration: the Incubator platforms Helmholtz AI, HIFIS, and HIP, as well as the ELLIS in Genoa and Munich, collaborated to contribute the content of the Challenge and the supporting program, as well as support. Experienced data scientists from the Israel Data Science Initiative advised teams in finding solutions. HAICORE@ KIT and NVIDIA Israel provided computing resources and support. The Italian partners Centro Diagnostico Italiano and BRACCO Imaging provided clinical data and X-ray images and funded prizes for winning teams and participants - computing resources from Amazon Web Services and technical literature worth €8.000.





Positive feedback on Twitter: International teams were pleased to have taken part in the Challenge.

## HELMHOLTZ DATA SCIENCE CAREER DAY (22.9.2021)

The Helmholtz Data Science Career Day is HIDA's virtual international career day for data scientists and took place for the second time in 2021. 600 participants from 75 countries visited the event. 22 Helmholtz centers and international partner institutions jointly presented more than 150 current job openings and career paths for data scientists in six research fields at the career day: Energy, Earth and Environment, Health, Aeronautics, Space and Transport, Matter and Information. 1.409 people from 125 countries had registered for the event and viewed the center and program information provided on the virtual event platform (hopin).

In November, 167 quests attended the first HIDA Annual program information provided on the virtual event platform Conference via the event platform hopin. The meeting connected Helmholtz scientists (especially PhD researchers and postdocs) in Information and Data Science from The conference program included eight panels and two all Helmholtz institutions through networking events keynotes by data scientists from Helmholtz Centers as well and workshops, promoting exchange on Information & as an interview with Helmholtz president Otmar D. Wiestler. Data Science projects and methods. The conference HIDA more than doubled the number of contacts and converprogram included a keynote by Neil Chue Hong, Director sations between centers and potential applicants compared and Principal Investigator of the Software Sustainability Institute, University of Edinburgh, an interview with Otmar to last year. The conference program provided insight into the exciting research at Helmholtz and the career paths D. Wiestler, a presentation on HIDA's services, seven taken by data scientists at the centers. Experts such as workshops, a networking session, and the presentation Achim Streit (KIT), Sabine Attinger (UFZ) and Marie Piraud of 21 Scientific Telegram videos with related exchanges. (Helmholtz Munich) spoke in keynotes and panels with Workshops from Helmholtz Incubator platforms and partinteractive O&A's about their career paths and opportuniners were in particularly high demand with 124 attendees, ties at Helmholtz. providing up-to-date knowledge on topics such as sustainable software development, the next generation of AI in medicine, writing scientific blogs, and more. In 112 networking sessions, organized as speed-dating video chats, data scientists met scientists from different Helmholtz Centers every five minutes.



A total of 600 visitors from 75 countries attended the Helmholtz Data Science Career Day 2021. 455 networking meetings took place. The career day provided opportunities for exchange in speed-dating video chats. Here, Information & Data Science talents met scientists from the Helmholtz Centers every five minutes to learn more about specific career opportunities. Afterwards, the 22 centers and partners presented themselves with Expo booths and were able to deepen conversations with participants in a (group) video conference format.

### HIDA ANNUAL CONFERENCE (30.11.2021)



Keynote speaker Neil Chue Hong, University of Edinburgh.

#### **5. COMMUNICATION**

Three experts from HIDA Steer, Stephan Frickenhaus (GEOMAR), Uwe Konrad (HZDR) and Achim Streit (KIT), reviewed the submitted virtual "Scientific Telegram" videos. During the conference, HIDA awarded six winners out of the 21 video submissions:

- Ist places Lorenz Lamm (MUDS, Helmholtz Munich): "Membrane analysis for Cryo-ET"; Jan Sodoge (UFZ): "Combining supervised and unsupervised machine learning to assess drought impacts from newspaper articles"
- 2nd places: Yvonne Jenniges (MarDATA, AWI): "Machine learning to define and characterize 3D ocean regions"; Edouard Fouché (KIT): "Bandit-based battery management with cost/privacy tradeoffs"
- 3rd places: Shammi More (FZJ): "Brain-age prediction: a systematic comparison of machine learning workflows"; Jonas Kasmanas (UFZ): "The human gut resistome exposed to different diets"

#### **HIDA LECTURES**

In 2021, HIDA established the Helmholtz Information & Data Science Academy Lectures (HIDA Lectures), a series of events organized by HIDA together with the six Helmholtz Information & Data Science Schools. Throughout the year, (international) data scientists have spoken about their current research and exchanged ideas with participating scientists. The HIDA Lectures support the exchange of information between the Schools and serve to promote Helmholtz as an employer and to attract talent in the field of Information & Data Science. Thematically, the virtual HIDA Lectures cover all Helmholtz research areas of the Schools - Energy; Earth and Environment; Health; Information; Matter; and Aeronautics, Space and Transport. The format is open to the public and the interested international Information & Data Science community and especially PhD researchers of the Helmholtz Association are invited.

At the beginning of each lecture, HIDA and HIDA's activities are presented to the audience as well as the respective School. In addition, the chairs of the schools introduced the speakers of the lectures and moderated the Q&A sessions as well as the "Meet the Speaker" exchange with the participants. The format started very successfully and had an average of about 80 participants.

## HIDA IMPLEMENTED THE FOLLOWING HIDA LECTURES:

15.04.2021: HIDA Lectures @DASHH with Anatole von Lilienfeld (Universität Wien), 147 participants

16.06.2021: HIDA Lectures @HEIBRIDS with Alan Akbik (Humboldt-Universität zu Berlin), 91 participants

29.09.2021: HIDA Lectures @MUDS with Niki Kilbertus (Helmholtz AI), 83 participants

20.10.2021: HIDA Lectures @HEIBRiDS with Dennis Shasha (Courant Institute of New York University), 30 participants

#### 07.12.2021:

HIDA Lectures @HDS-LEE with Steven Brunton (University of Washington), 83 participants



#### WEBSITE

In 2021, a relaunch of the HIDA website was undertaken and in November 2021, the site went live with a new design. The goal was to develop the website into an attractive platform that promotes and makes visible cross-community education & training offers in the field of Information & Data Science as well as career opportunities at all Helmholtz Centers.

A key tool for this is the integration of the HIDA Course Catalog, which now provides a dedicated subpage for users to intuitively search for Information & Data Science course offerings and events from across the Helmholtz community.

Another important component of the HIDA website is the HIDA Data Science Job Board, which presents all Information & Data Science job opportunities at the Helmholtz Centers. Through it, HIDA aims to make Information & Data Science talent aware of career opportunities at the Helmholtz Centers. In 2021, the job board recorded over 32.000 views. A majority of the hits on the job board resulted in views of the detailed job postings on the Centers' websites.

The new website also gave articles a prominent position in the "Newsroom". Divided into HIDA News and Centers News, users can get a comprehensive picture of the activities as well as research in the field of Information & Data Science in the Helmholtz Association. A series of portraits on the individual Helmholtz Information & Data Science Schools offers in-depth insights into the special features of the schools. These reports are flanked by interviews with personalities from state governments about the importance of the Schools for the promotion of young scientists and Information & Data Science in Hamburg (Rolf Greve), Baden-Württemberg (Simone Schwanitz) and Bavaria (Manfred Wolter).

The content partnerships with magazines from various Helmholtz Centers (DLR Magazine, KIT Magazine, effzett Jülich, infact HZI, Helmholtz Klimainitiative and Helmholtz.de) were continued in 2021, and in the future, there will be more opportunities on the new Helmholtz.de website to additionally tease HIDA news articles.

In 2021, the HIDA website saw a steady increase in user interest. The number of visitors to www.hida-helmholtz.de 2021 increased by around 50% compared to the previous year. Of this traffic, 85% was generated organically. The



HIDA offerings and the HIDA Data Science Job Board again accounted for the largest share of traffic in 2021. Incidentally, the average time spent on the HIDA website in 2021 was almost 2 minutes. This shows that an interested target group is reached via the approach on the website, where they can obtain information that is relevant to them. The increase in the download rate of info-documents by over 25% compared to 2020 also proves this.



In 2021, visitors to the site came from 120 countries, with most accessing the site from Germany and the U.S.A.

Website visitors in 2021 came from 120 countries: 35% from Germany, 36% from the USA, 10% from the UK, 6% from Russia, 1% from China and 12% from other countries. This year's website traffic also shows that, with around 15.000 outbound referrals, many visitors were redirected from the HIDA site to the site of a Helmholtz Center or School - an increase of over 50% compared to the previous year. The website thus successfully functions as a distribution platform and contributes to the consolidation of the Helmholtz-wide information offering and the networking of the Centers.



Fig. 6: HIDA website: Proportion of visitors by country.

Most outbound links remain within the Helmholtz Association and lead to Centers and Schools. The top 10 overview of outbound referrals in 2021 (number of unique link clicks and percentage increase compared to 2020):

FZ Jülich 1.952 (+ 177%)
 DLR 1.252 (+ 63 %)
 HIFIS 951 (+ 100 %)
 helmholtz.de 748 (+ 76 %)
 Jobs DKFZ 387 (+ 316 %)
 MDC Berlin 375 (+ 36 %)
 Helmholtz Munich 358 (+ 100%)
 DESY 358 (+ 26 %)
 HEIBRIDS 298 (+ 7 %)
 GFZ 249 (+ 1%)

#### SOCIAL MEDIA

In 2021, HIDA continued to use various social media channels to provide information about its own offerings and Information & Data Science offerings from the Helmholtz network and to highlight events. HIDA's social media channels have become well established, as the user numbers show, and are growing steadily: The HIDA Twitter channel @HIDAdigital grew to over 2.900 followers in 2021; posts reached over 1.8 million tweet impressions in 2021. In total, the HIDA Twitter profile was viewed over 54.000 times last year.

The HIDA LinkedIn profile saw a doubling of followers in 2021, now reaching over 1.200 people, with over 200.000 post impressions.

The HIDA newsletter, which provides information about HIDA offerings, debuting in 2020, has over 900 subscribers by the end of 2021. In addition, special newsletters are sent out occasionally to inform about special events or calls for proposals.



Twitter and LinkedIn are HIDA's social media channels with a steadily growing reach.



#### Dear Friends of HIDA,

It promises to be an exciting data science autumn: join us, network and land your job at our 25 Helmholtz Centers and partners during the second Data Science Career Day on September 22. If you would like to know what's happening in the Helmholtz data science universe, mark November 30 in your agenda and attend our HIDA Ann Those who would like to advance their skills can sign up for a workshop in R and shouldn't miss the next HIDA Lecture: Young Investigator Niki Kilbertus from Helmholtz AI in Munich will be speaking at MUDS on September 29. Would you like to get to know (other) Helmholtz Centers? Apply now for the ing Program, which is for researchers from other institutions, or ioin the HIDA Trainee Network as a Helmholtz scientist. Postdoc Noai Han Chan from GFZ did it and took a Data Science look at permafrost soils with the help of AWI colleagues. Linking domain science and data science in graduate education is precisely what makes the Helmholtz Data Science Schools so appealing. The concept also convinces politicians: The state of Baden-Württemberg is providing over 1,2 million euros of addition funding for HIDSS4Health for 2021 to 2024. By the way: If you are still looking for a way into data science: The HDS-LEE is currently advertise doctoral positions again - spread the word! May the data be with you!

#### Featured Story

"Heimholtz has created an outstanding initiative for advancing the careers of young scientists in the data sciences by setting up its information & Data Science Schools, This is exactly the right way forward."

#### "This is an Ideal Way of Promoting Networks"

The state of Baden-Württemberg is a major funder of HIDSS4Health and is providing over 1,2 million euros of additional funding for the school for 2021 to 2024. But what was it about the HIDSS4Health concept that impressed policymakers? We spoke with Dr. Simone Schwanitz, Senior Ministry Official at the Baden-Württemberg Ministry of Science, to learn more.

Read the interview >

Training & Events



The HIDA newsletter reaches over 900 readers.

#### 7. HIDA STEERING BOARD (HIDA-STEER)

There have been changes in HIDA-Steer in 2021. Due to<br/>the resignation of Veit Hagenmeyer (representative:<br/>Research Field Energy), Klaus Maier-Hein (representative:<br/>HIDSS4Health) and Frank Jenko (representative: MUDS)<br/>from HIDA-Steer replacements were necessary. On<br/>September 15-16, 2021, the Helmholtz Association's General<br/>Assembly officially appointed the following Helmholtz<br/>experts to help shape the overall direction of HIDA:Two virtual HIDA-Steer Meetings took place (08.03.2021 and<br/>08.09.2021).<br/>In 2021, HIDA organized two virtual HIDA Liaison Officer and<br/>School Coordinator Meetings (06.05.2021 and 10.11.2021).

- > Manuel Dahmen, Forschungszentrum Jülich (new representative: Research Field Energy)
- Daniel Hübschmann, German Cancer Research Center (DKFZ) (new representative: HIDSS4Health)
- Julia Schnabel, Helmholtz Munich (new representative: MUDS)

With these new colleagues, HIDA-Steer is now complete and includes the following members:



» Prof. Dr. Sabine Attinger, UFZ





» Prof. Dr. Stephan Frickenhaus, AWI



» Dr. Uwe Konrad, HZDR





» Prof. Dr. Uwe Ohler, MDC



» Prof. Dr. Julia Anne Schnabel, Helmholtz Munich



» Prof. Dr. Wolfgang Wiechert, FZ Jülich



» Dr.-Ing. Manuel Dahmen, FZ Jülich

» Dr. Dr. Daniel Hübschmann, DKFZ

» Prof. Dr. Alice McHardy, HZI

» Prof. Dr. Nina Rohringer, DESY

» Prof. Dr. Achim Streit, KIT (SCC)

» Prof. Dr. Xiaoxiang Zhu, DLR

#### DASHH Data Science in Hamburg HELMHOLTZ Graduate School for the Structure of Matter

# IV. FACTS & FIGURES ON THE HELMHOLTZ INFORMATION & DATA SCIENCE SCHOOLS

#### 1. DATA SCIENCE IN HAMBURG -HELMHOLTZ GRADUATE SCHOOL FOR THE STRUCTURE OF MATTER (DASHH)

he Data Science in Hamburg - Helmholtz Graduate School for the Structure of Matter (DASHH) aims to educate the future generation of data and information scientists that will tackle tomorrow's scientific challenges coming along with large-scale experiments on an interdisciplinary level.

A total of 25 doctoral researchers is currently enrolled with In DASHH, the Deutsches Elektronen-Synchrotron (DESY) coa dissertation project at DASHH of which 19 are men and six operates with world-leading large-scale research facilities are women. 13 doctoral researchers are international, 12 and other key-research institutions and universities from are German. The international PhD researchers come from Hamburg and other northern German states to provide inno-Italy, Romania, Belarus, China, Iran, Syria, Russia, India, and vative training and cooperation for doctoral researchers in Colombia. the field of Information & Data Science. With its approach, DASHH is pioneering the development of new collaborative Additionally, three doctoral researchers at DESY, all male approaches to evaluate complex, heterogeneous data using (two international researchers), are associated with DASHH intelligent algorithms.

Research Areas: DASHH research covers data challenges in highly interdisciplinary topics from a wide range of fields such as Particle Physics, Structural Biology, Materials Science, Ultrafast X-Ray Science, Accelerator Science, Computer Science, and Applied Mathematics.

Partners of DASHH: Deutsches Elektronen-Synchrotron DESY, Hamburg University of Technology (TUHH), Universität Hamburg (UHH), Helmut Schmidt University Hamburg (HSU), Helmholtz Center hereon (formerly Helmholtz Center Geesthacht), Helmholtz Center for Infectious Diseases (HZI), Max Planck Institute for the Structure and Dynamics of Matter (MPSD), European XFEL (EuXFEL), Hamburg University of Applied Sciences (HAW).

#### **APPLICANT SITUATION AND RECRUITMENT**

Eight new PhD positions were awarded in each of 2019 and 2020, and nine new PhD positions were awarded in 2021. Applicants came from all over the world: In the Call for Applications 2019, 208 people from 35 countries applied; in 2020, 118 people from 27 countries applied, and finally in 2021, 126 people from 33 countries applied. The gender ratio was similar in all application rounds with 27.5% female applicants in 2019, 22% female applicants in 2019, and 27% female applicants in 2020 respectively.

#### **EVENTS AND NETWORKING**

While there were six events in 2019 (four networking events and two lectures), 29 events were carried out in 2020 (11 networking events and 18 lectures). DASHH initiated a Hamburg COVID-19 Series in collaboration with the Leibniz Science Campus InterACt. The COVID-19 Series is a platform for scientists in the Hamburg metropolitan area to present recent challenges and results from the field of COVID-19 research to increase the visibility of this research fostering new contacts and potential collaborations.

DASHH also established a monthly Data Science Colloquium where renowned scientists discuss recent Data Science challenges. The Colloquium is jointly organized by DASHH, the Center for Data and Computing in Natural Sciences (CDCS), the clusters of excellence CUI - Advanced Imaging of Matter and Quantum Universe of the Universität Hamburg (UHH), the Department of Informatics of the UHH, the Institute of Mathematics of the Hamburg University of Technology and ahoi.digital - the Alliance of Hamburg Universities for Computer Science and open to all Helmholtz Information & Data Science School researchers. Monthly Get Together events with all DASHH doctoral researchers and associated doctoral researchers were established at the end of 2020 and were organized as either online or on-site events throughout 2021.

In 2021, 63 events were carried out (22 networking events, 34 lectures, 7 courses).

DASHH organized nine lectures in form of PhD seminars one of which was a joint event with the graduate school HELIOS (Helmholtz-Lund International Graduate School). Furthermore, DASHH (co-)organized 7 courses (in 38 course units) where one of the highlights was the Research Data for Open Science course by the Nature Research Academies. The Hamburg COVID-19 Lecture Series took place 17 times with on average 56 participants.

The Data Science Colloquium took place 8 times with on average 149 participants and 69 to 85% external attendees. In May 2021, DASHH organized a Match Making Workshop to facilitate networking between DASHH PIs and DASHH Scientists. The workshop was opened to established DASHH PIs as well as DASHH PIs aspirants. It was very successful with 43 participants (4 new researchers interested in DASHH) and led to many successful project proposals.

#### » 9 Courses

#### DASHH & IMPRS-UFAST courses

- > 19.04.-23.04.2021, Hans Fangohr, Henning Glawe, Heiko Appel: "Introduction to Programming with Python for Computational Science" (31 participants)
- > 10.05.-12.05.2021 & 17.05.-21.05.2021 & 24.05.-31.05.2021, Heiko Appel, Henning Glawe, Hans Fangohr, Nicolas Tancogne-Dejean: "Numerical Methods and Practical Skills for Computational Physics" (33 participants)
- > 06.09.-10.09.2021, Martin Lueders, Micael Oliveira, Nicolas Tancogne-Dejean, Heiko Appel: "Introduction to the Octopus Code - Basics" (16 participants)
- > 20.09.-24.09.2021, Martin Lueders, Micael Oliveira, Angel Rubio, Sebastian Ohlmann, Nicolas Tancogne-Dejean, Heiko Appel: "Introduction to the Octopus Code - Advanced Topics" (15 participants)

#### DASHH & PEP Courses

- > 21./22.01.2021, Bernd Klein: "Introduction into Machine Learning with Python" (14 participants)
- > 10./11.05.2021, Dr. Matthias Mayer: "Time Management for Doctoral Candidates" (12 participants)
- > 14.09.2021, Dr. Alexander Schiller: "Good Scientific Practice" (18 participants)

#### **Research Data for Open Science Courses**

 > 21.06.2021, Nature Research Academies: "Data Publishing" & "Allowing Reuse and Gaining Credit" & 22.06.2021, "Practical Applications of the FAIR Principles" & "Sharing Sensitive Research Data" (85 participants)

#### » 7 Data Science Colloquium/Lectures

- > 14.01.2021, Prof. Ulrike von Luxburg: "The Inductive Bias in Machine Learning" (253 participants)
- > 22.04.2021, Prof. Frank Krauss: "Introducing JUNE

   an Open-source Epidemiological Simulation" (108
   participants)
- 20.05.2021, Prof. Weinan E.: "A Mathematical Perspective of Machine Learning" (219 participants)
- > 24.06.2021, Prof. Michael Bronstein: "Geometric Deep Learning: From Euclid to Drug Design" (170 participants)
- > 08.07.2021, Prof. Ullrich Köthe: "Invertible Neural Networks and their Applications in the Sciences" (116 participants)
- > 21.10.2021, Dr. Daniel Ratner: "Machine Learning for X-Ray and cryoEM at SLAC" (105 participants)
- > 10.12.2021, Prof. Holger Gohlke: "Enzyme Function
   Much to Unterstand, Optimize, and Discover" (72 participants)

#### » 1 Data Science Colloquium / HIDA Lecture @ DASHH

 > 15.04.2021, Prof. Anatole von Lilienfeld: "Quantum Machine Learning in Chemical Compound Space" (147 participants)

#### » 17 Hamburg COVID-19 Lecture Series

- > 13.01.21, Dr. Andrea Zaliani: "Rapid Identification of Ligands and Inhibitors of Viral Proteins Using Biochemical Screening of Fragment, Lead and Drug-like Compounds" (67 participants)
- > 27.01.21, Dr. Stephan Menzel: "Characterization of SARS-CoV-2 Specific Antibody Response by Flow Cytometry" (50 participants)
- > 10.02.21, Dr. Andrea Thorn: "The Coronavirus Structural Taskforce" (87 participants)

- > 24.02.21, Prof. Reidun Twarock: "Mathematical & Structural Virology: Coronavirus Assembly as a Novel Drug Target" (53 participants)
- > 10.03.21, Prof. Nicole Fischer: "SARS-CoV-2 Genome Sequencing: The Impact of Viral Mutants at Local, National and International Level" (111 participants)
- > 24.03.21, Dr. Ayan Paul: "On Transmission Dynamics of COVID-19: A Physicist's Perspective" (77 participants)
- O7.04.21, Prof. Chris Meier: "Viral Polymerases as Intervention Targets using Nucleotide Drugs" (86 participants)
- > 21.04.21, Dr. Charlotte Uetrecht: "Flying Viruses from Biophysical to Structural Characterization" (78 participants)
- > 05.05.21, Dr. Katrin Schöning-Stierand, Dr. Christiane Ehrt: "Exploring SARS-CoV-2 Targets with the ProteinsPlus Web Server" (54 participants)
- > 19.05.21, Prof. César Muñoz-Fontela: "An Update on COVID-19 Epidemiology from the WHO R&D Blueprint Team" (67 participants)
- > 16.06.21, Dr. Ricardo Strauss, Kaja Kristensen: "Exploring the Use of Web Searches for Risk Communication During COVID-19 in Germany" (21 participants)
- > 30.06.21, Prof. Gülsah Gabriel: "Sex Differences in COVID-19 Pathogenesis" (29 participants)
- > 28.07.21, Prof. Giulia Rosetti: "A Blueprint for High Affinity SARS-CoV-2 Mpro Inhibitors from Activity-Based Compound Library Screening Guided by Analysis of Protein Dynamics" (25 participants)
- > 11.08.21, Dr. Muhamed Amin: "How Molecular Modeling Could Help in the COVID19 Crisis?" (32 participants)
- > 25.08.21, Dr. Dr. Madeleine Altfeld-Bunders: "3D Organoids for Infectious Disease Research" (41 participants)
- > 08.09.21, Dr. Irene Fernandez-Cuesta: "Development of a New COVID-19 Rapid Test" (40 participants)
- > 22.09.21, Prof. Fischer, Prof. Grundhoff: "Monitoring the Emergence and Origin of SARS-CoV-2 Mutations -Genetic Inter- and Intrahost SARS-CoV-2 Diversity in Immunocompetent and Immunocompromised Patients" (44 participants)

#### » 9 PhD Seminar Lectures

- O5.02.21, Prof. Patrick Huber: "Liquids in Nanopores: From the Structure and Dynamics of Confined Matter to the Design of Functional Materials" (31 participants)
- > 12.02.21, Dr. Sadia Bari: "X-ray Spectroscopy of Gasphase Biomolecules" (38 participants)
- > 30.04.21, Prof. Ralf Röhlsberger: "Phase Retrieval in Synchrotron based Nuclear Resonant Scattering" (40 participants)
- > 21.05.21, Prof. Mathias Trabs: "Introduction to Stochastic Differential Equations" (39 participants)
- O2.07.21, Prof. Christian Schroer: "Improved Modeling and Resulting Challenges for the Solution of the Phase Problem in Near-field Xray Microscopy" (47 participants)
- O6.11.21, Prof. Elisabetta Gallo: "The Higgs Boson and the bbh Channel" (37 participants)
- > 10.12.21, Prof. Gregor Kasieczka: "Unsupervised Learning as a New Tool in Particle Physics" (48 participants)
- 17.12.21, Dr. Sebastian Günther: "Protein Dynamics under Non-Ambient Conditions studied by X-Ray Crystallography" (34 participants)
- > 11.06.21 (DASHH meets HELIOS lectures), Dr. Irene Fernandez Cuesta: "Single Molecule Detection on Nanodevices for Biomedical Applications" (39 participants)

#### » 1 MatchMaking Workshop

 28.05.21, with Henry Chapman, Steve Aplin, Daniel Ruprecht, Robin Santra, Christian Schwanenberger, Berit Zeller-Plumhoff, Imke Greving, Jan Baumbach, Kerstin Tackmann, Katharina Kubicek, Andreas Stierle, Walid Maalej, Jochen Küpper, Nihat Ay, Andrea Thorn, (43 participants)

#### » 11 PhD Get Together - Networking Events

(3 in cooperation with the Center for Data and Computing in Natural Sciences, CDCS)

#### » 10 PhD Seminars

- > 05.02.21, Lars Dammann: "Correlating High Resolution X-ray Diffraction with Molecular Dynamics Simulation Data" (31 participants)
- > 12.02.21, Amir Kotobi: "Dynamic Structure Investigation of Biomolecules with Pattern Recognition Algorithms and X-ray Experiments" (38 participants)
- 30.04.21, Ankita Negi: "Phase Retrieval in Synchrotronbased Nuclear Resonant Scattering" (40 participants)

- > 21.05.21, Vladislav Sukharnikov: "Generalized Second Quantization and Stochastic Description of Open Quantum Systems" (39 participants)
- > 02.07.21, Johannes Dora: "Improved Modeling and Resulting Challenges for the Solution of the Phase Problem in Near-Field Xray Microscopy" (47 participants)
- > 06.11.21, Maryam Bayat Makou: "Search for the B-associated Standard Model Higgs Boson with Neural Networks at CMS" (37 participants)
- > 10.12.21, Sebastian Bieringer: "Amplifying Calorimeter Simulations with Deep Neural Networks" (48 participants)
- > 17.12.21, Thorben Poburski: "MD Simulation for Temperature-Dependent Protein Dynamics Analysis" (34 participants)
- > 11.06.21 (PhD Seminar DASHH meets HELIOS), Franziska Esmek: "Single Molecule Detection on Nanodevices for Biomedical Applications" (39 participants)
- > 02.03.21 (HIDSS Exchange networking), Nils Margenberg: "Deep Neural Network Multigrid Solver" (26 participants)

#### PUBLICATIONS

DASHH doctoral researchers are highlighted.

#### » 9 First-author Publications

**Grünhagen, A.,** Branlard, J., Eichler, A., **Martino, G.,** Fey, G., and Tropmann-Frick, M. (2021). Fault Analysis of the Beam Acceleration Control System at the European XFEL using Data Mining. Paper presented at: Asian Test Symposium (ATS) (Ehime, Japan, 2021).

Martino, G., Gruenhagen, A., Branlard, J., Eichler, A., Fey, G., and Schlarb, H. (2021). Comparative Evaluation of Semi-Supervised Anomaly Detection Algorithms on High-Integrity Digital Systems. Paper presented at: 2021 24th Euromicro Conference on Digital System Design (DSD).

**Saleh, Y.,** Sanjay, V., Iske, A., Yachmenev, A., and Küpper, J. (2021). Active Learning of Potential-Energy Surfaces of Weakly Bound Complexes with Regression-Tree Ensembles. Journal of Chemical Physics 155, 144109.

Weber, T., Riebisch, M., Borras, K., Jansen, K., and Krücker,
D. (2021). Modelling for Quantum Error Mitigation. Paper
presented at: 2021 IEEE 18th International Conference on
Software Architecture Companion (ICSA-C).

Schütte, M., Eichler, A., Lamb, T., Rybnikov, V., Schlarb, H., and Wilksen, T. (2021). Subsystem Level Data Acquisition for the Optical Synchronization System at European XFEL. Paper presented at: IPAC2021 (Campinas, SP, Brazil). (DASHH Associate PhD researcher)

**Schütte, M.,** Eichler, A., Schlarb, H., Lichtenberg, G., and Werner, H. (2021). Decentralized Output Feedback Control using Sparsity Invariance with Application to Synchronization at European XFEL. Paper presented at: 60th IEEE Conference on Decision and Control (CDC) (Austin, Texas, USA). (DASHH Associate PhD researcher)

**Günther, S.,** Reinke, P.Y.A., Fernandez-Garcia, Y., Lieske, J., Lane, T.J., Ginn, H.M., Koua, F.H.M., **Ehrt, C.,** Ewert, W., Oberthuer, D., et al. (2021). X-Ray Screening Identifies Active Site and Allosteric Inhibitors of SARS-CoV-2 Main Protease. Science 372, 642-646. (DASHH PostDoc)

**Kan, Y.,** Kärtner, F., Le Borne, S., Ruprecht, D., and Zemke, J.-P. (2020). Parallel Computation of Inverse Compton Scattering Radiation Spectra based on Lienard-Wiechert Potentials, under review.

**Dammann, L.,** Meißner, R. H., Seeck, O., Huber, P., Imbibition of Water by Nanoporous Silica (MCM-41), Album of Porous Media - Structure and Dynamics, Springer, accepted.

#### » 22 Co-author Publications

Ai, X., **Mania, G.,** Gray, H.M., Kuhn, M., and Styles, N. (2021). A GPU-Based Kalman Filter for Track Fitting. Computing and Software for Big Science 5.

Plambeck, S., **Martino, G.**, and Fey, G. (2021). Metrics for the Evaluation of Approximate Sequential Streaming Circuits. Paper presented at: 2021 24th Euromicro Conference on Digital System Design (DSD).

Sirunyan, A.M., Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Del Valle, A.E., Fruhwirth, R., Jeitler, M., et al. (2021). Search for Charged Higgs Bosons Produced in Vector Boson Fusion Processes and Decaying into Vector Boson Pairs in Proton-Proton Collisions at  $\sqrt{s}$ =13 TeV. European Physical Journal C Particles and Fields 81, 723. (DASHH co-author: **Rübenach, J.**) Sirunyan, A.M., Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., et al. (2021). Search for Lepton-Flavor Violating Decays of the Higgs Boson in the  $\mu\tau$ and et Final States in Proton-Proton Collisions at  $\sqrt{s}$ =13 TeV. Physical Review D 104. (DASHH co-author: **Rübenach, J.**) Tonon, N., Aarup Petersen, H., Aldaya Martin, M., Asmuss, P., Baxter, S., **Bayatmakou, M.,** Behnke, O., Bermúdez Martínez, A., Bhattacharya, S., Bin Anuar, A.A., et al. (2021). Probing Effective Field Theory Operators in the Associated Production of Top Quarks with a Z Boson in Multilepton Final States at  $\sqrt{s}$ =13 TeV. Journal of High Energy Physics 2021. (DASHH co-authors: **Bayatmakou, M., Rübenach, J.**)

Sirunyan, A.M., Tumasyan, A., Adam, W., Andrejkovic, J.W.,
Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del
Valle, A., Frühwirth, R., Jeitler, M., et al. (2021). Constraints
on Anomalous Higgs Boson Couplings to Vector Bosons and
Fermions in its Production and Decay Using the Four-Lepton
Final State. Physical Review D 104. (DASHH co-author: **Rübenach, J.**)
Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T.,
Chatterjee, S., Dragicevic, M., Del Valle, A.E., Fruhwirth, R.,
Deitler, M., Krammer, N., et al. (2021). Combined Searches for the
Production of Supersymmetric Top Quark Partners in ProtonProton Collisions at √s=13 TeV. European Physics Journal C
Particles and Fields 81, 970. (DASHH co-author: **Rübenach, J.**)

Sirunyan, A.M., Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., et al. (2021). Measurements of Higgs Boson Production Cross Sections and Couplings in the Diphoton Decay Channel at  $\sqrt{s}=13$  TeV. Journal of High Energy Physics 2021. (DASHH co-author: **Rübenach, J.**) Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., Krammer, N., et al. (2021). Search for New Particles in Events with Energetic Jets and Large Missing Transverse Momentum in Proton-Proton Collisions at  $\sqrt{s}=13$  TeV. Journal of High Energy co-authors: **Bayatmakou, M., Rübenach, J.**)

Sirunyan, A.M., Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., et al. (2021). Search for W' Bosons Decaying to a Top and a Bottom Quark at √s=13 TeV in the Hadronic Final State. Physics Letters B 820. (DASHH co-author: **Rübenach, J.**)

Sirunyan, A.M., Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., et al. (2021). Search for Resonant and Nonresonant New Phenomena in High-Mass Dilepton Final States at √s=13 TeV. Journal of High Energy Physics 2021. (DASHH co-author: **Rübenach, J.**)

Sirunyan, A.M., Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Fruhwirth, R., Jeitler, M., et al. (2021). Observation of a New Excited Beauty Strange Baryon Decaying to  $\pm_b^ \pi^+ \pi^-$ . Physical Review Letters 126, 252003. (DASHH co-author: **Rübenach, J.**)

Sirunyan, A.M., Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Valle, A.E.D.,
Fruhwirth, R., Jeitler, M., et al. (2021). Precision Luminosity
Measurement in Proton-Proton Collisions at√s=13 TeV in 2015 and 2016 at CMS. European Physics Journal C Particles and
Fields 81, 800. (DASHH co-author: Rübenach, J.)
Proton-Proton Collisions at √s=13 TeV. Physical Review D 104. (DASHH co-author: Rübenach, J.)
Proton-Proton Collisions at √s=13 TeV. Physical Review D 104. (DASHH co-author: Rübenach, J.)

Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., Krammer, N., et al. (2021). Search for Long-Lived Particles Decaying in the CMS End Cap Muon Detectors in Proton-Proton Collisions at  $\sqrt{s}$ =13 TeV. Physical Review Letters 127. (DASHH co-authors: **Bayatmakou, M., Rübenach, J.**)

Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., Krammer, N., et al. (2021). Observation of tW Production in the Single-Lepton Channel in pp Collisions at  $\sqrt{s}$ =13 TeV. Journal of High Energy Physics 2021. (DASHH co-author: **Rübenach, J.**)

Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., Krammer, N., et al. (2021). Measurement of Differential ( $\overline{tt}$ ) Production Cross Sections in the Full Kinematic Range using Lepton + Jets Events from Proton-Proton Collisions at  $\sqrt{s}$ =13 TeV. Physical Review D 104. (DASHH co-author: **Rübenach, J.**) Measurements of the Electroweak Diboson Production Cross CONFERENCE CONTRIBUTIONS Sections in Proton-Proton Collisions at √s=5.02 TeV Using Leptonic Decays. Physical Review Letters 127, 191801. (DASHH co-author: Rübenach, J.)

Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., Krammer, N., et al. (2021). Search for a Heavy Higgs Boson Decaying into Two Lighter Higgs Bosons in the ttbb Final State at 13 TeV. Journal of High Energy Physics 2021. (DASHH co-author: **Rübenach, J.**)

Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., > Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., Krammer, N., et al. (2021). Study of Z Boson Plus Jets Events Using Variables Sensitive to Double-Parton Scattering in pp Collisions at 13 TeV. Journal of High Energy Physics 2021. (DASHH co-author: **Rübenach, J.**)

Tumasyan, A., Adam, W., Andrejkovic, J.W., Bergauer, T., Chatterjee, S., Dragicevic, M., Escalante Del Valle, A., Frühwirth, R., Jeitler, M., Krammer, N., et al. (2021). Measurements of the pp $\rightarrow$ W<sup>+</sup>± yy and pp $\rightarrow$ Zyy Cross Sections at  $\sqrt{s}$ =13 TeV and  $\rightarrow$ Limits on Anomalous Quartic Gauge Couplings. Journal of High Energy Physics 2021. (DASHH co-author: **Rübenach, J.**)

Fields, J.K., Kihn, K., Birkedal, G.S., Klontz, E.H., Sjöström, K., Günther, S., Beadenkopf, R., Forsberg, G., Liberg, D., Snyder, G.A., et al. (2021). Molecular Basis of Selective Cytokine Signaling Inhibition by Antibodies Targeting a Shared Receptor. Frontiers in Immunology 12. (DASHH PostDoc)

Graef, J., Ehrt, C., Diedrich, K., Poppinga, M., Ritter, N., and Rarey, M. (2021). Searching Geometric Patterns in Protein Binding Sites and Their Application to Data Mining in Protein Kinase Structures. Journal of Medicinal Chemistry. (DASHH PostDoc)

#### » 1 Preprint

Srinivasan, V., Brognaro, H., Prabhu, P.R., de Souza, E.E., Günther, S., Reinke, P.Y.A., Lane, T.J., Ginn, H., Han, H., Ewert, W., et al. (2021). SARS-CoV-2 Papain-Like Protease PLpro in Complex with Natural Compounds Reveal Allosteric Sites for Antiviral Drug Design, bioRxiv 2021.11.17.468943. (DASHH co-author: **Größler, M.**)

#### » 12 International

#### Posters

- 28.01.2021, DESY/XFEL Users' Meeting, Poster, Superfluorescence of multilevel compact system: exact solution and stochastic modelling, Vladislav Sukharnikov (DASHH PhD researcher)
- 24.05.-28.05.2021, 12th International Particle Accelerator Conference, Poster, Subsystem Level Data Acquisition for the Optical Synchronization System at European XFEL, Maximilian Schütte (DASHH Associate PhD researcher)
- 24.08.-03.09.2021, Warsaw summer school for quantum physics and chemistry, Poster, Spectral and active learning for (ro-)vibrational calculations of weakly-bound molecules, Yahya Saleh (DASHH PhD researcher)
- 01.12.2021, 20th International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Poster, A GPU-based Kalman Filter for Track Fitting, Georgiana Mania (DASHH PhD researcher)

#### Talks

- 22.03.2021, QSA 1<sup>st</sup> Workshop on Quantum Software Architecture, Talk, Modelling for Quantum Error Mitigation, Tom Weber (DASHH PhD researcher)
- 08.07.2021, ML4jets, Talk, Amplifying Statistics with Generative Models, Sebastian Guido Bieringer (DASHH PhD researcher)
- 08.07.2021, 6<sup>th</sup> ECCOMAS Young Investigators Conference, Talk, Deep neural networks for geometric multigrid methods, Nils Margenberg (DASHH PhD researcher)
- 23.07.2021, LASPHYS 2021, Talk, Superfluorescence of Compact System: Exact Solution and Stochastic Treatment, Vladislav Sukharnikov (DASHH PhD researcher)
- 23.07.2021, LASPHYS 2021, Talk, Probabilistic Stochastic Methodology for the Light-Matter Interaction, Stasis Chuchurka (DASHH PhD researcher)
- 02.09.2021, Euromicro Conference on Digital System Design (DSD), Talk, Comparative Evaluation of Semi-Supervised Anomaly Detection Algorithms on High-Integrity Digital Systems, Gianluca Martino (DASHH PhD researcher)
- 23.11.2021, Asian Test Symposium (ATS), Talk, Fault Analysis of the Beam Acceleration Control System at the European XFEL using Data Mining, Arne Grünhagen (DASHH PhD researcher)

> 16.12.2021, 60th IEEE Conference on Decision and Control (CDC), Talk, Decentralized Output Feedback Control Using 16.03.2021, DPG Spring Conference 2021, Talk, Search for Sparsity Invariance with Application to Synchronization at heavy Higgs bosons decaying to top guark pairs using the European XFEL, Maximilian Schütte (DASHH Associate PhD CMS experiment, Jonas Rübenach (DASHH PhD researcher) researcher) 17.03.2021, DPG Spring Meeting, Talk, Combination of

#### » 20 National

#### Posters

- > 15.03-19.03.2021, Molecular dynamics in the gas-phase conference (MD-GAS), Poster, Dynamic structure investigation of biomolecules with pattern recognition algorithms and X-ray experiments, Amir Kotobi (DASHH PhD researcher)
- > 08.09.-12.09.2021, H3: Helmholtz Herbst Hackathon, Poster, Spectral learning for (ro-)vibrational calculations of weakly-bound molecules, Yahya Saleh (DASHH PhD researcher)
- > 08.09.-12.09.2021, H3: Helmholtz Herbst Hackathon, Poster, Dynamic structure investigation of biomolecules with pattern recognition algorithms and X-ray experiments, Amir Kotobi (DAHHH PhD researcher)
- 31.05.2021, Machine learning in engineering (MLE) at Hamburg University of Technology, Talk, Dynamic structure investigation with pattern recognition algo-> 08.09.-12.09.2021, H3: Helmholtz Herbst Hackathon, rithms and X-ray experiments, Amir Kotobi (DASHH PhD Poster, Towards X-ray Free Electron Laser simulation in researcher), Video: https://webcast.tu-harburg.de/ proteins, Carlos Ortiz Mahecha (DASHH PhD researcher) Mediasite/Play/ef2f74fff5f847e78d11b208346ad0ba1d
- > 08.09.-12.09.2021, H3: Helmholtz Herbst Hackathon, 07.07.2021, Machine learning for quantum X conference, Poster, Single-Shot Structured Light 3D Scanning with Invited Talk, Active and spectral learning for quantum Convolutional Neural Networks Trained on Synthetic dynamics of weakly-bound molecular complexes, Yahva Fractals, Ke Li (DASHH PhD researcher) Saleh (DASHH PhD researcher)
- > 13.09.-17.09.2021, BiGmax Summer School on Big-Data-27.07.2021, Atomic Summer Camp, Talk, Probabilistic Driven Materials Science 2021, Poster, Dynamic structure stochastic methodology for superfluorescence, Stasis investigation of biomolecules with pattern recognition Chuchurka (DASHH PhD researcher) algorithms and X-ray experiments, Amir Kotobi (DASHH 29.07.2021, Atomic Summer Camp 2021, Talk, Generalized PhD researcher) second quantization of open quantum systems, Vladislav > 30.11.2021, HIDA Virtual Annual Conference, Research Sukharnikov (DASHH PhD researcher)
- 23.09.2021, Deutsche CMS conference (DCMS), Talk, Telegram, A multi-purpose framework for efficient parallelized execution of charged particle tracking, Georgiana Search for B-Associated Standard Model Higgs Production Mania (DASHH PhD researcher) in Decay Channel with Two Tau Leptons, Maryam Bayat Makou (DASHH PhD researcher)
- > 30.11.2021, HIDA Virtual Annual Conference, Research Telegram, Dynamic structure investigation of biomole-30.09.2021, Helmholtz MT ARD ST3 Workshop, Speed cules with supervised and unsupervised machine learning, Talk, Decentralized Control Approach for the Optical Amir Kotobi (DASHH PhD researcher) Synchronization System at the European XFEL, Maximilian Schütte (DASHH Associate PhD researcher)

#### Talks

- Higgs boson measurements using Simplified Template Cross Sections at the ATLAS experiment, Joshua Clercx (DASHH Associate PhD researcher)
- 18.03.2021, DPG conference, Talk, A search for resonances decaying into a Higgs boson and a new particle X with the ATLAS detector, Nicola de Biase (DASHH Associate PhD researcher)
- > 23.03.2021, Quantum Universe Day, Talk, Search for heavy Higgs bosons decaving to top guark pairs using the CMS experiment, Jonas Rübenach (DASHH PhD researcher)
- 12.05.2021, Bunsen-Tagung, Contributing Talk, Applications of machine learning in quantum simulations of hydrogen-bond dynamics, Yahya Saleh (DASHH PhD researcher)

> 23.11.2021, Terascale14, Talk, Search for heavy Higgs bosons decaying to top quark pairs using the CMS experiment, Jonas Rübenach (DASHH PhD researcher)

#### FURTHER NEWS FROM DASHH

#### » Supervision of Bachelor and Master Theses **Bachelor Theses**

> Performance Study on Gpu Offloading Techniques Using the Matrix Inverse Algorithm, Yannik Koenneker, co-supervised by Georgiana Mania (DASHH PhD researcher), status: in progress

#### Master Theses

- > Solution of the Vibrational Schrödinger Equation Using Neural Networks, Jannik Eggers, co-supervised by Yahya Saleh (DASHH PhD researcher), status: in progress
- > Grammar-Based Generation of Systemverilog Assertions Using Signals' Context and Traces, Julian Engelskirchen, co-supervised by Gianluca Martino (DASHH PhD researcher), status: finished
- > Fault Analysis of the Beam Acceleration Control System at European XFEL Using Data Mining, Arne Grünhagen, co-supervised by Gianluca Martino (DASHH PhD researcher), status: finished

#### » New Partners

By the end of 2020, DASHH could win the Hamburg University of Applied Sciences as a new partner that is actively involved in the supervision of two ongoing PhD projects and is involved in three new project proposals advertised in 2021. In 2021, five new senior researchers and 10 new young researchers joined DASHH as DASHH PIs and DASHH Scientists respectively. The new DASHH PIs are

- > Prof. Arwen Pearson, Group Leader Experimental Biophysics, UHH;
- > Prof. Jan Baumbach, Head of the Research Group Computational Systems Biology, UHH;
- > Prof. Walid Maalej, Head of the Department of Informatics, UHH;
- > Prof. Holger Sondermann, Head of the Structural Microbiology group at DESY and researcher at the Center for Structural Systems Biology;
- > Prof. Nihat Ay, Head of the Institute for Data Science Foundations, TUHH.

In addition, Dr. Andrea Thorn, group leader at the Institute for Nanostructure and Solid State Physics, Universität Hamburg, and leader of the "Coronavirus Structural Task Force" joined DASHH as DASHH Scientist. All new members (except for Prof. Sondermann who joined at the end of 2021) successfully submitted PhD project proposal which were accepted by the DASHH Topic Evaluation and Selection Committee for the DASHH Call for Applications 2021 and are expected to supervise new DASHH PhD researchers in 2022. In 2021, DASHH partnered up with the Machine Learning in Engineering Initiative (MLE) of the TUHH. MLE's objective is to pool competencies in machine learning and transfer knowledge to economy and industry.

#### » Communication

DASHH is present and active on LinkedIn (since 2019) and since 2021 on Twitter: @DASHHPhD. To extend visibility, DASHH representatives attended to the Virtual Fair "Research in Germany" at the UHH MIN booth.

DASHH was present with a virtual booth on the Helmholtz Data Science Career Day and participated in the matchmaking session with DASHH PIs and the representatives of doctoral researchers

DASHH has an entry in the DAAD PhDGermany Database since summer 2021

#### » PhD Researcher Activities

In March 2021, DASHH doctoral researcher Ankita published a blog post titled 'The "Phase" Problem of My Life as a PhD Researcher' on Science Chatter Hamburg (Blogs and Podcasts from Researchers at Science City Hamburg Bahrenfeld).

In April 2021, DASHH was featured in an article published on the HIDA website under the title "Revolutionizing Science with Big Data" together with an interview with Rolf Greve, Senate Director at the Ministry for Science and Research at the city of Hamburg, on the potential of interdisciplinary and the importance of the Data Science School for Hamburg. (https://www.helmholtz.de/en/newsroom/article/ revolutionizing-science-with-big-data/)

On the 2021 HIDA Annual Conference, two DASHH PhD researchers participated in the Science Telegrams session and presented their research:

- > Georgiana Mania: "A Multi-Purpose Framework for Efficient Parallelized Execution of Charged Particle Tracking "
- Free-Electron Laser Experiments."

> Amir Kotobi: "Dynamic Protein Pattern Recognition in Ke Li works on the DESY/UHH PhD project "Mixed Reality User Interfaces for Operating Particle Accelerators and Laser Facilities". After doing her master thesis on "Single In May 2021, DASHH PhD researcher Amir Kotobi held a talk Shot Structured Light 3D Scanning with Convolution Neural Networks And Synthetic Fractals" at the 3D sensing group in the 'Train Your Engineering Network' lecture series organized by the Machine Learning in Engineering Initiative of at the Fraunhofer Institute of Applied Optics and Precision the Technical University of Hamburg with the title "Dynamic Engineering in Jena, Germany, Ke Li started to work on her Structure Investigation of Biomolecules With Pattern DASHH PhD project in July 2021. At DASHH, she works at Recognition Algorithms and X-Ray Experiments". the accelerator control system group at DESY and the human-computer interaction (HCI) group at the UHH to create » DASHH PI News mixed reality (MR) applications to operate modern accelera-In April 2021, the research by DASHH PI Alke Meents is feators and X-ray free-electron laser facilities. She is currently tured in an article on the DESY News Page titled "Desy X-Ray building MR laser safety goggles for the laser technology Light Source Identifies Promising Candidates for COVID group at DESY where laser lab users use a video-see-through Drugs". https://www.desy.de/news/news\_search/index\_ head-mounted display (VST-HMD) as an eye protector for eng.html?openDirectAnchor=2044&two\_columns=0 full-band laser protection. In her free time, she enjoys color-pencil drawing and composing music. She is also a skilled In June 2021, the research by DASHH PI Patrick Huber who clarinet player.

heads the research group "High-Resolution X-Ray Analytics of Materials" at DESY and the Institute for Materials and X-ray Physics at TU Hamburg, is featured in an article on the DESY News Page titled "Laser pulses make nanomaterials sound". https://www.desy.de/news/news\_search/index\_eng. html?openDirectAnchor=2090

DASHH PI and DESY physicist Jens Osterhoff received the 2021 Bjørn H. Wiik Prize in recognition of his outstanding contributions to the field of plasma acceleration.

DASHH PI Prof. Martin Müller, Helmholtz Center Hereon, has been appointed head of the Materials Research Division at Hereon

DASHH PI Sadia Bari, young research group leader at DESY, has been appointed professor at the University of Groningen in the Netherland in DESY's first joint professorship agreement with an international university.

#### » New DASHH PhD Researchers in 2021

As in the previous years, DASHH was proud to admit very talented young researchers with impressive CVs as PhD researchers in 2021. The short portrayal of PhD researcher Ke Li stands representatively for the excellent background of DASHH PhD researchers.



Silicon sensor of the CMS Fast Beam Conditions Monitor, connected via needles to precision multimetres for testing - by DASHH doctoral researcher Jonas Rübenach, first price at CMS Deutschland Photo Contest 2020

## HDSLEE HELMHOLTZ SCHOOL FOR DATA SCIENCE IN LIFE LEARTH LENERGY

#### 2. HELMHOLTZ SCHOOL FOR DATA SCIENCE IN LIFE, EARTH AND ENERGY (HDS-LEE)

he HDS-LEE structured doctoral program aims at excellent graduates of mathematics, computer science, natural sciences and engineering from all over the world. The doctoral researchers at HDS-LEE are trained in all essential areas of Information & Data Sciences as well

as in communication and other key gualifications. The training components of the program are strengthened by individually tailored training measures, e.g. at the Jülich Supercomputing Center (JSC).

#### **APPLICANT SITUATION AND RECRUITMENT**

While 13 positions were filled in 2019 and eight in 2020, four new PhD positions were filled in 2021. In 2019, a total of 330 people from 70 nations applied; in the following year, 90 applications came in from 28 nations. In 2021, 435 scientists from 40 nations applied to eleven open positions advertised through the recruiting portal at FZJ. In all years, the high proportion of applicants from Iran, India and China should be emphasized.

The current 25 PhD positions at HDS-LEE are held by people from 14 nations, twelve of them are Germans. Five women are among the doctoral researchers. Sixteen other doctoral researchers are associate members, including four women. Of the associated researchers, thirteen belong to FZ Jülich, three to RWTH Aachen University, one to University of Cologne and one to University Bonn. Two associated doctoral researchers finished their theses, successfully defended their dissertations and received their doctoral degrees in 2021.

#### EVENTS AND NETWORKING

While there were only two HDS-LEE events in 2019 (one networking event and one lecture), in 2020 significantly more events took place, namely 41 (9 networking events, 16 lectures, 16 courses). In 2021, 50 events took place (27 lectures, 9 networking events, 14 courses).

#### » 9 Networking Events

The weekly virtual coffee breaks (since April 15, 2020) was well attended and used as platform to get feedback on the current situation of the doctoral researchers. The coffee break was initiated in 2020 to enable an informal exchange between the doctoral researchers and to have the possibility to have regular contact to the coordinator team.

In March 2021, a monthly **virtual women lunch** was initiated to enable the exchange between female doctoral researchers. They network, learn about programs for female scientists and get to know other female experts in their field.

Monthly Doctoral Seminars: Discussion groups on video presentations (13.01.2021, 17.02.2021, 24.03.2021, 26.05.2021): The discussion group meetings initiated in 2020 were continued in 2021 as a compensation for the retreat, since that could not take place on-site in 2020 and 2021. In these informal meetings, a doctoral researcher presents his work and answers questions. The aim is to exchange knowledge and build cooperation between the doctoral researchers (16-25 participants)

As the **annual internal retreat** could not take place on-site in 2021, it was held as an online doctoral seminar on two days (10.-11.06.2021). The annual internal retreat is designed as a platform where doctoral researchers inform each other about their latest research results and open issues. It is a forum for knowledge transfer and information exchange. The doctoral researchers give presentations on current research topics (both results and open issues). (61 participants)

To foster the group networking, an **internal workshop** was initiated, which was composed of two parts:

- > Workshop part 1 Introduction to Workflow Tools: MLflow, DVC (16.07.2021, 38 participants)
- > Workshop part 2 Hackathon: Hearts Gym (20.-23.07.2021, 38 participants)

Part 1 was organized by a group of three HDS-LEE doctoral researchers. They planned the topic, content, and hands- on 5 EU Regional School Courses (AICES, RWTH) session themselves. This workshop "Introduction to Workflow Tools" gave a glimpse into the topic and an overview of the The EU Regional School is open to graduates from universiworkflow tools to show how MLOps can save time and conties in Germany, Belgium, and the Netherlands. There are 5 tribute an improved quality of data analysis. This workshop courses per term. It is a series of three-hour "short coursfocuses on scientific use cases to help to decide if and in es" on introductory and advanced various topics in comwhich cases a tool could potentially support a Information putational science, and are attended by a cross-section of graduates in computational science (fluid mechanics, com-& Data Science project. In the hands-on session, it is taught how to tack models, results, and data, as well as tricks of the puter science, molecular dynamics, contact mechanics, reduced-order modeling, etc.). Organized by AICES, RWTH and trade to enable multistage analysis pipelines with MLFlow. Part 2 was jointly organized together with the Helmholtz AI JARA-CSD (missing course sessions were postponed due to Team at FZJ. The idea to teach an agent plaving the heart Corona situation), https://blog.rwth-aachen.de/irtg-mip/ game came from a doctoral researcher. The aim was to learn eu-regional-school-videos-2020 reinforcement learning in a card game. Reinforcement learning is used to teach an agent optimal strategies (policy) to 25.03.2021, Prof. Irene Gamba Ph.D. - Non-linear achieve maximum reward in a Markov decision process by Boltzmann type models in Collisional Theory using a simple card game: Hearts. In Hearts, four players > 11.05.2021, Prof. Dr. Andreas Fichtner - Probabilistic Fullplay against each other and have to avoid winning hands that waveform Inversion > 20.05.2021, Steve Lionel - Modern Fortran: Features for include hearts or the queen of spades. These players can be High-Performance Computing modelled as reinforcement learning agents. The Helmholtz 29.09.2021, Prof. Tan Bui-Tanh Ph.D. - On Unifying AI team at FZJ created a multi-agent environment to train Randomized Methods for Inverse Problems agents on playing the Hearts game, which also includes a 06.12.2021, Prof. Dr. Martin Frank - Uncertainty client-server architecture to remotely evaluate local agents. In this server architecture, the agents can battle against Quantification for Hyperbolic Conservation Laws each other to evaluate which group has taught the best cards 2 Lectures on Data Science, Methods & Applications player. The overall aim is to do some Collaborative Coding (HDS-LFF) within the HDS-LEE doctoral researchers community as a networking event. The Lectures on Data Science, Methods & Applications aim

In HDS-LEE three transferable skill courses are mandatory: "Management of a doctoral project", "Scientific writing", and "Academic Presentation". After having completed the last of these courses "Academic Presentation", the idea arose to practice giving presentations in the group of the doctoral researchers. Therefore, the Presentation Club was initiated as an offer to doctoral researchers to practice giving presentations and giving feedback. (21.10.2021, 17 participants) On December 6, 2021, an on-site poster session was planned, which could unfortunately could not take place. Instead of that, a doctoral seminar was held with presentations of doctoral researchers and a general assembly of the PhDs, where they elected a representative. The poster session was postponed to April 2022. (54 participants)

In 2021, the coordinator played a key role in the planning and implementation of the Helmholtz Herbst Hackathon. Four HDS-LEE doctoral researchers attended and presented their work in the poster session.

#### » 14 Courses

to teach and train the doctoral researchers in all essential elements of Information & Data Science. The doctoral researchers are introduced to the respective disciplines and Information & Data Science methods, get an insight and apply what they have learned in hands-on sessions.https:// www.hds-lee.de/events/lectures-on-data-science/

- > 07.-08.06.2021, Dr. N. Ben Rached, Dr. C. Ben Hammouda-Uncertainty Quantification in a nutshell -The Bayesian framework for inverse problems (40 participants)
- > 16.-17.09.2021 Jun.-Prof. Dirk Witthaut Time Series Analysis (35 participants)

#### 2 HIFIS Courses

- > 20.-21.04.2021, Introduction to Git and GitLab (about 20 participants)
- 22.-23.04.2021, HIFIS Course: Bring your Own Script and Make It Ready for Publication (about 20 participants)

#### 4 Transferable Skill Courses (JuDocs, FZJ)

Transferable skills courses aim to foster the professional skills and personal development essential to the successful completion of doctoral projects. Transferable Skill Courses are offered to the entire group of HDS-LEE doctoral researchers and are organized via JuDocs, a center for transferable skills training for doctoral researchers at FZJ.

- > 22.-23. April 2021, Doing Science/ Management of a Doctoral Project (12 participants)
- > 25.-26.11.2021, Scientific Writing (12 participants)
- > 09.-10.03.2021 and 27.-28.03.2021, Academic Presentation (12 participants each)

#### Mental Health Course (Dragonfly, MH)

Managing Upward is a workshop that introduces to strategies and techniques that will improve the ability to set expectations, establish and renegotiate healthy boundaries, and anticipate and address issues that arise. This course was conducted by Dragonfly, Mental Health, jointly with HIDSS4Health.

18.06.2021, Managing upward Course - Optimizing Student-Supervisor Communication to Support Wellbeing (15 participants)

#### » 27 Lectures HDS-LEE Lecture Series

In the "HDS-LEE Seminar Series", Information & Data Science experts from the application areas Life, Earth and Energy are invited to present the current state of their research to interested researchers. The goal is that the doctoral researchers get to know the experts in the individual fields and have the opportunity to interact and discuss with them.

- > 07.01.2021, Prof. Ira Assent: Data Science Concepts and Methods to Learn in a Digital World (34 participants)
- 26.01.2021, Prof. Dr. Holger Gohlke: Simulation and Data Science for Structure-based Computational Biotechnology
- > 07.12.2021, Prof. Dr. Steven L. Brunton: Machine Learning for Scientific Discovery, with Examples in Fluid Mechanics (HIDA Lecture @HDS-LEE)

#### SSD Seminar Series (SSD, JARA-CSD)

Experts are invited to present their research in a seminar series, operating on a semester-based schedule. The aim is to invite speakers relevant to the SSD, but also potentially of interest to a wider audience. Such seminars provide opportunities for learning about state-of-the-art research, and for interaction and discussion with top experts. The seminar series invites up to 20 speakers per year. A combination of national and international guest speakers is anticipated.

- > 11.01.2021, Prof. Dr. Bruno Sudret Surrogate Models for Forward and Inverse Uncertainty Quantification
- > 18.01.2021, Prof. Dr. Simona Perotto Topology
   Optimization: From the Macro- to the Micro-scale
- 25.01.2021, Prof. James Sprittles Ph.D. Noisy Interfacial Nanoflows
- O1.02.2021, Prof. Connor Coley Ph.D. Chemical Discovery and AI-Assisted Chemical Synthesis
- > 08.02.2021, Prof. Dr. Johann Sebastian Trimpe Eventtriggered Learning
- > 26.04.2021, Jan Christian Hueckelheim, Ph.D. Automatic Differentiation in High Performance Computing and Machine Learning
- > 03.05.2021, Prof. Dr. Anne K. Reinarz Bayesian Inverse Problems for Tsunami Simulation
- > 10.05.2021, Markus Ihmsen, Ph.D. Virtual Testing with Particle-based Methods in the Industry
- O7.06.2021, Dr.-Ing. Andre Weiner Computational Fluid Dynamics and Machine Learning with OpenFOAM and PyTorch
- > 14.06.2021, Prof. Dr. Florian Wagner An Overview of Joint Inversion Approaches in Geophysical Imaging of the Subsurface
- > 21.06.2021, Prof. Dr. Katerina Stankova Improving Cancer Treatment Through Game Theory
- > 28.06.2021, Prof. Dr. Victor Zavala Tejeda The Euler Characteristic: A General Topological Descriptor for Complex Data
- > 05.07.2021, Dr. Martin Eigel An Adaptive Tensor Reconstruction for Bayesian Inversion
- > 12.07.2021, Dr. Micheal F. Herbst Accelerating the Discovery of Tomorrow's Materials by Robust and Errorcontrolled Electronic-structure Simulations
- > 08.11.2021, Dr. Federica Ferraro and Ludovico Nista, M.Sc. - The Role of High-Performance Computing for Energy Transition
- > 29.11.2021, Prof. Spyros Pandis, Ph.D. Atmospheric Nanoparticles, Air Quality and Climate Change

- > 06.12.2021, Thomas Hösgen, M.Sc. High-Performance Computing for the Optimization of Modern Gas Turbine Cooling Systems
- > 13.12.2021, Dr. Michel Make Spline-Based Methods for Fluid-Structure Interaction
- 20.12.2021, Prof. Masayuki Yano Towards Reliable and Automated Model Reduction of Parametrized Nonlinear PDEs: Error Estimation, Adaptivity, and Application to Aerodynamics
- > 10.01.2021, Dr. Patrick Simon Stumpf Digital Twins in Biomedical Research - Towards a Systems-Level Model of Human Disease
- > 17.01.2021 Dr. Daniel Caviedes-Voullième Performance Portability for Earth System Modeling: Who, Why, How?
- 20.01.2021, Prof. Christian Rohde Compressible Two-Phase Flow across Scale using Sharp and Diffuse Interface Ideas
- > 24.01.2021, Prof. Susanne Buiter Geodynamic Experiments on the Formation of Plate Margins during Continental Break-Up
- 31.01.2021, Binbin Lin, M. Sc. and Dr. Daniel Utt and Setareh Medghalchi, M. Sc. - Computational and Experimental Methods in Material Science

#### PUBLICATIONS

In 2021, 23 first author publications and 14 co-authored publications were published as peer-reviewed journal articles, preprints, or conference proceedings.

HDS-LEE doctoral researchers gave 31 talks at conferences and they were involved in six talks given by collaborating scientist of their projects. One talk was awarded at a conference.

HDS-LEE doctoral researchers presented their project 19-times in posters at conferences and were involved in the work of eight other posters presented at conferences. They received two poster prizes and awards.

HDS-LEE doctoral researchers are **highlighted**, associated doctoral researchers are marked with asterisk. First author publications of HDS-LEE doctoral researchers are *italic*.

#### •

#### » 2 Doctoral Theses

\*Gorjão, L. R. Stochastic timeseries analysis in electric power systems and paleo-climate data. Dissertation, University of Cologne, 2021

\*Zhao, H. Gaussian processes for sensitivity analysis, Bayesian inference, and uncertainty quantification in landslide research. RWTH Aachen University. 2021

#### » 37 Peer-reviewed Journal Articles, Preprints

Aljawad, H., \*Rüttgers, M., Lintermann, A., Schröder, W., & Lee, K. C. (2021). Effects of the Nasal Cavity Complexity on the Pharyngeal Airway Fluid Mechanics: A Computational Study. Journal of Digital Imaging, 34(5), 1120-1133.

\*Beumer, L., Niemeyer, I. Data Science Meets Nuclear – What Data Analytics, Computational Intelligence and Machine Learning Can Contribute to Nuclear Waste Management and Nuclear Verification. Conference proceedings, INMM & ESARDA Joint Annual Meeting 2021.

Betancourt, C., \*Stomberg, T., Roscher, R., Schultz, M. G., Stadtler, S. AQ-Bench: a benchmark dataset for machine learning on global air quality metrics; Earth System Science Data; volume 13; pages 3013-3033; year 2021. **Boledi, L.,** Terschanski, B., Elgeti, S., & Kowalski, J. (2021). A level-set based space-time finite element approach to the modelling of solidification and melting processes. arXiv preprint 2021.

**Cramer, E.,** Mitsos, A., Tempone, R., & Dahmen, M. (2021). Principal Component Density Estimation for Scenario Generation Using Normalizing Flows. 1-15. arXiv preprint 2021.

**Cramer, E.,** Gorjão, L. R., Mitsos, A., Schäfer, B., Witthaut, D., & Dahmen, M. (2021). Validation Methods for Energy Time Series Scenarios from Deep Generative Models. 1-16. arXiv preprint 2021.

**Doncevic, D. T.,** Schweidtmann, A. M., Vaupel, Y., Schäfer, P., Caspari, A., & Mitsos, A. (2020). Deterministic Global Nonlinear Model Predictive Control with Neural Networks Embedded. IFAC-PapersOnLine, 53(2), 5273-5278.

Felton, K. C., \*Rittig, J. G., & Lapkin, A. A. (2021). Summit: Benchmarking Machine Learning Methods for Reaction Optimisation. Chemistry-Methods, 1(2), 116-122.

Felton, K. C., \*Rittig, J. G., & Lapkin, A. A. (2021). Summit: Benchmarking Machine Learning Methods for Reaction Optimisation. Chemistry-Methods, 1(2), 116-122.

\**Gerloff, C.,* Konrad, K., Bzdok, D., Büsing, C., & Reindl, V. (2021). Interacting brains revisited: A cross-brain network neuroscience perspective. Biorxiv preprint 2021.

**Grimm, V.,** Heinlein, A., Klawonn, A., Lanser, M., & Weber, J. (2021). Estimating the time-dependent contact rate of SIR and SEIR models in mathematical epidemiology using physics-informed neural networks. ETNA - Electronic Transactions on Numerical Analysis, 56, 1-27.

\**Gorjão, L. R.,* Witthaut, D., Lehnertz, K., & Lind, P. G. (2021). Arbitrary-order finite-time corrections for the Kramers-Moyal operator. Entropy, 23(5), 1-15.

\**Gorjão, L. R.,* Schäfer, B., Witthaut, D., & Beck, C. (2021). Spatio-temporal complexity of power-grid frequency fluctuations. New Journal of Physics, 23(7), 073016.

\**Gorjão, L.R.,* Vanfretti, L., Witthaut, D., Beck, C., Schäfer, B. Phase and amplitude synchronisation in power-grid frequency fluctuations in the Nordic Grid. arXiv preprint. 2021. \**Gorjão, L. R.,* Hassan, G., Kurths, J., Witthaut, D. MFDFA: Efficient Multifractal Detrended Fluctuation Analysis in Python. arXiv preprint. 2021

Guo, Y., Dietrich, F., Bertalan, T., **Doncevic, D. T.,** Dahmen, M., Kevrekidis, I. G., & Li, Q. (2021). Personalized Algorithm Generation: A Case Study in Meta-Learning ODE Integrators. arXiv preprint. 2021, 1-34.

Hammouda, C. Ben, Rached, N. Ben, Tempone, R., & Wiechert, S. (2021). Efficient Importance Sampling via Stochastic Optimal Control for Stochastic Reaction Networks. arXiv preprint 2021.

\**Helleckes, L. M.*, Osthege, M., Wiechert, W., von Lieres, E., & Oldiges, M. (2021). Bayesian calibration, process modeling and uncertainty quantification in biotechnology. arXiv preprint 2021.

König, A., **Siska, M.,** Schweidtmann, A. M., \*Rittig, J. G., Viell, J., Mitsos, A., & Dahmen, M. (2021). Designing production-optimal alternative fuels for conventional, flexible-fuel, and ultra-high efficiency engines. Chemical Engineering Science, 237, 116562.

Kruppa, J. A., Reindl, V., \*Gerloff, C., Oberwelland Weiss, E., Prinz, J., Herpertz-Dahlmann, B., Konrad, K., & Schulte-Rüther, M. (2021). Brain and motor synchrony in children and adolescents with ASD — a fNIRS hyperscanning study. Social Cognitive and Affective Neuroscience, Interpersonal Synchrony Special Issue, 103-116.

Kowalski, J., \*Zhao, H., Cai, Y. Topographic uncertainty in avalance simulations, International Snow Science Workshop Proceedings 2018.

*Kruse, J.,* Schäfer, B., & Witthaut, D. (2021). Revealing drivers and risks for power grid frequency stability with explainable AI. Patterns, 2(11), 100365.

*Kruse, J.,* Schafer, B., & Witthaut, D. (2021). Exploring deterministic frequency deviations with explainable AI. 2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 133-139.

*Neubacher, C.,* Witthaut, D., Wohland, J. Multi-decadal offshore wind power variability can be mitigated through optimized European allocation. Advances in Geosciences 2021, 54, 205-215.

Reindl, V., Wass, S., Leong, V., Scharke, W., Wistuba, S., Wirth, C. L., Konrad, K., & \*Gerloff, C. (2021). Synchrony of mind and body are distinct in mother-child dyads. BioRxiv Preprint 2021.

Reindl, V., Schippers, A., Tenbrock, K., Job, A., \*Gerloff, C., Lohaus, A., Heinrichs, N., & Konrad, K. (2021). Caregiving quality modulates neuroendocrine and immunological markers in young children in foster care who have experienced early adversity. Journal of Child Psychology and Psychiatry 2021.

**M. Röhrig-Zöllner,** J. Thies and A. Basermann. Performance of low-rank approximations in tensor-train format (TT-SVD) for large dense tensors. arXiv preprint 2021.

Rother, D., & \*Malzacher, S. (2021). Computer-aided enzymatic retrosynthesis. Nature Catalysis, 4(2), 92-93.

**Ruzaeva, K.,** Nöh, K. and Berkels, B. Polar Space Based Shape Averaging for Star-shaped Biological Objects. Eurographics Workshop on Visual Computing for Biology and Medicine, virtual, Germany, 2021.

\*Rüttgers, M., Waldmann, M., Schröder, W., & Lintermann, A. (2021). Machine-Learning-Based Control of Perturbed and Heated Channel Flows. In H. Jagode, H. Anzt, H. Ltaief, & P. Luszczek (Eds.), High Performance Computing (pp. 7-22). Springer International Publishing.

*Simson, A.,* Löwe, H., and Kowalski, J.: Elements of future snowpack modeling – Part 2: A modular and extendable Eulerian-Lagrangian numerical scheme for coupled transport, phase changes and settling processes, The Cryosphere, 15, 5423-5445, 2021.

\*Stomberg, T., Weber, I., Schmitt, M., Roscher, R. Junglenet: Using explainable machine learning to gain new insights into the appearance of wilderness in satellite imagery, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume V-3-2021 XXIV ISPRS Congress (2021 edition).

Thies, J., *Röhrig-Zöllner, M.*, Basermann, A. (R)SE challenges in HPC. RSE-HPC-2021 Workshop der Supercomputing 2021 (SC'21), arXiv preprint 2021.

Theorell, A., *Jadebeck, J. F.*, Nöh, K., & Stelling, J. (2021). PolyRound: polytope rounding for random sampling in metabolic networks. Bioinformatics, 38(2), 566-567. \*Zhao, H., Kowalski, J. Topographic uncertainty quantification for flow-like landslide models via stochastic simulations, Natural Hazards and Earth System Sciences, 2020.

\*Zhao, H., Amann, F., Kowalski, J. Emulator-based global sensitivity analysis for flow-like landslide run-out models, Landslides (2021).

\*Zimmer, M., Pesch, T., & Benigni, A. (2021). Time-series analysis and forecasting of power consumption using gaussian process regression. SEST 2021 – 4th International Conference on Smart Energy Systems and Technologies, 10-15.

#### » 64 Conference Proceedings

#### Talks

Betancourt, C., Stadtler, S., \*Stomberg, T., \*Edrich, A.-K., \*Patnala, A., Roscher, R., Kowalski, J., & Schultz, M. (2021). Global fine resolution mapping of ozone metrics through explainable machine learning. vEGU21 Conference.

**Boledi, L.,** Terschanski, B., Elgeti, S. and Kowalski, J. A Space-Time FE Level-set method for convection coupled phase-change processes. VI ECCOMAS Young Investigators Conference YIC2021. Valencia, 7-9 July 2021.

Boxberg, M. S.. Audehm, J., Becker, F., **Boledi, L.,** Burgmann, B., Chen, Q., Friend, P., Haberberger, N., Heinen, D., Nghe, C. T., **Simson, A.,** Stelzig, M., Kowalski, J. 2021: TRIPLE – Ice Data Hub, Model-based Mission Support and Forefield Reconnaissance System. 81st Conference of the German Geophysical Society, online (Kiel).

**Cramer, E.,** Mitsos, A., Tempone, R., & Dahmen, M. A Principal Component Normalizing Flow for Manifold Density Estimation. SIAM Conference on Computational Science and Engineering 2021.

**Cramer, E.,** M. A Principal Component Normalizing Flow for Modeling Renewable Electricity Generation. 4th Machine Learning and Al in Bio(Chemical) Engineering Conference 2021. **Awarded talk** 

**Cramer, E.,** M. A Principal Component Normalizing Flow for Modeling Renewable Electricity Generation. Annual meeting of American Institute of Chemical Engineers 2021. **Doncevic, D.T.,** Schweidtmann, A., Kerkenhoff, T., Dahmen, M. and Mitsos, A. Guaranteed Accuracy of Machine Learning-Based Surrogate Models Using Bilevel Optimization in AIChE Annual meeting 2021.

\*Dwaipayan, C. Can satellite images provide supervision for cloud systems characterization. Helmholtz information data science academy virtual conference, 2021.

\**Edrich, A.-K.* Machine learning based prediction of shallow landslides. AGU (virtual) 2021.

\**Edrich, A.-K.* Machine learning based prediction of shallow landslides. MBD Colloquium, RWTH Aachen 2021.

Fleitmann, L. H. J.; Ackermann, P.; \*Rittig, J. G.; Schweidtmann, A. M.; König, A.; Dahmen, M.; Mitsos, A.; Bardow, A.; Leonhard, K.. Fuel design by combined machine learning and predictive thermodynamics in 9. Internationale Konferenz – Fuel Science- From Production to Propulsion Aachen digital 2021-06-22 - 2021-06-24.

\**Gerloff, C.,* Konrad, K., Bzdok, D., Buesing, C. and Reindl, V. Network based inference and prediction for interacting brains. Virtual conference: fNIRS 2021.

\**Germscheid, S.,* Dahmen, M., Mitsos, A. Assessing the Demand Response Potential of Power-Intensive Processes by Stochastic Scheduling Optimization, AIChE Annual Meeting 2021, Boston, 2021.

\**Gorjão, L. R.* Phase and Amplitude Synchronisation in Power-Grid Frequency Fluctuations. SIAM virtual Conference on Applications of Dynamical Systems (DS21), 2021.

*Hassan, J.,* von Lieres, E. Hybrid process modeling combining mechanistic equations with machine learning Conference: 16th International PhD Seminar on Chromatographic Separation Science, 2021.

*Kruse, J.* Exploring deterministic frequency deviations with explainable AI. Conference "IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)" 2021.

*Kruse, J.* Erklärbare KI für mehr Transparenz im Stromnetz. Conference "Future Energies Science Match 2021". \**Malzacher, S.* Synthesis planning and data management in biocatalysis. Webinar of the European Society of Applied Biocatalysis (ESAB) 2021.

\**Malzacher, S.* Biocathub, data management platform for biocatalysis. 2nd EnzymeML Workshop 2021.

Müller, C.; \*Helleckes, L.; Griesbach, T.; Waffenschmidt, V.; Osthege, M.; Wiechert, W.; Oldiges, M. From strain to screening: Automated workflows for identifying the best signal peptide for target protein secretion via split GFP assay in German Conference on Synthetic Biology, GCSB 2021, virtual, Germany, 2021.

Müller, C.; \*Helleckes, L.; Waffenschmidt, V.; Wiechert, W.; Oldiges, M. Automated signal peptide screening workflow for secretion of heterologous proteins in C. glutamicum in ECCE 13 und ECAB 6, virtual, Germany, 2021.

**Röhrig-Zöllner, M.** Performance of high-order SVD approximation: reading the data twice is enough, SIAM Conference on Applied Linear Algebra (LA21), 2021.

**Röhrig-Zöllner, M.** and Holke, J. Testing HPC C++ software with GoogleTest: adjusting the test framework for distributed-parallel tests using MPI, SeptembRSE'21, 2021.

\**Rittig, J. G.;* Ritzert, M.; Schweidtmann, A. M.; Winkler, S.; Grohe, M.; Mitsos, A.; Dahmen, M.: Computer-Aided Design of Molecules with High Octane Rating and Octane Sensitivity through Graph-Based Machine Learning, 9th International Conference Fuel Science-From Production to Propulsion, Aachen 2021.

\**Rittig, J. G.;* Schweidtmann, A.r M.; König, A.; Dahmen, M.; Grohe, M.; Mitsos, A.: Deep End-To-End Learning on Molecular Graphs for Physico-Chemical Property Prediction using Graph Neural Networks : 8th – 9th July 2020 ; 3rd International Conference on Machine Learning an Al in (bio) Chemical Engineering.

\**Rittig, J. G.,* Schweidtmann, A. M., König, A., Dahmen, M., Grohe, M., Mitsos, A. Deep End-To-End Learning on Molecular Graphs for Physico-Chemical Property Prediction using Graph Neural Networks. In: 3rd International Conference on Machine Learning and AI in (bio)Chemical Engineering Cambridge online 2020. \**Rüttgers, M.* Analysis of medical and simulation data for improved treatment in rhinology. JLESC workshop 2021.

\**Rüttgers, M.* Machine-Learning-Based Control of Perturbed and Heated Channel Flows. ISC HPC 2021 conference 2021.

*\*Rüttgers, M.* Shape Optimization for Respiratory Flows through Reinforcement Learning. SCONA 2022 conference 2021.

**Ruzaeva, K.,** Nöh, K., Berkels, B. Polar space based shape averaging for star-shaped biological objects. 11th EG Workshop on Visual Computing for Biology and Medicine, Paris, France, 2021.

**Terhag, F.** Probabilistic Volume Prediction from Heart MRIs. Wissensaustausch Workshop Maschinelles Lernen. 2021.

*Wiechert, S.* Efficient Importance Sampling via Optimal Control for Stochastic Reaction Networks. MCM 2021 -Mannheim (13th International Conference).

**Wolff, D.,** Elgeti, S., Behr, M. Algorithmic Differentiation in a Modern C++17 Continuum Mechanics Simulation Code. WCCM-ECCOMAS Virtual Congress (Paris, Frankreich) 2021.

**Wolff, D.,** Posada, A., Elgeti, S. Invertible Neural Networks as Reduced Models for Inverse Problems in Production Engineering Applications 6th ECCOMAS Young Investigators Conference (Valencia, Spanien) 2021.

**Wolff, D.,** Posada, A., Elgeti, S. Invertible Neural Networks as Reduced Models for Inverse Shape Design Problems. Mechanistic Machine Learning and Digital Twins for Computational Science, Engineering & Technology (San Diego, USA) 2021.

Yildiz, A., Baselt, I., \*Edrich, A.-K., Fischer, J.-T., Mergili, M., \*Zhao, H., & Kowalski, J. (2021). Emulation techniques for rapid flow-like geohazards: a case study-based performance analysis. EGU General Assembly 2021, vEGU21, Online.

\*Zimmer, M., Pesch, T. and Benign, A. Time-Series Analysis and Forecasting of Power Consumption using Gaussian Process Regression. 4th International Conference on Smart Energy Systems and Technologies (SEST) 2021, Vaasa, Finland.

#### Posters

\*Beumer, L., Berkels, B., Niemeyer, I. Verification from Space – Building Transparency and Confidence through Earth Observation Big Data. SCIENCE, PEACE, SECURITY '21, Aachen, Germany.

Beyβ, M., **Jadebeck, J. F.,** Paul, R. D., Theorell, A., Wiechert, W., Nöh, K. Shifting the Limits of Bayesian Inference in Metabolic Modeling: The HPC Sampling Library HOPS Metabolic Engineering 14 virtual, Hawaii, USA, 2021.

Boxberg, M. S., Baader, F., **Boledi, L.,** Chen, Q., Dachwald, B., Francke, G., Kerch, J., Plesa, A.-C., **Simson, A.**, and Kowalski, J.: Concepts to utilize planetary analogue studies for icy moon exploration missions, EGU General Assembly 2021, online.

Jazib H., E. von Lieres. Hybrid process modelling of single component chromatography breakthrough curves using various machine learning methods. 1st Helmholtz Herbst Hackathon, Gummersbach (Germany), 2022-09-15 - 2022-09-18.

\*Helleckes, L.; Müller, C.; Puchta, D.; Czech, H.; Wiechert, W.; Oldiges, M. Automated high-throughput strain characterisation: From the frozen working cell bank to the final product assay, in ECCE 13 und ECAB 6. 2021: virtual, Germany. **"Best Poster Award"** 

\*Helleckes, L., Osthege, M.; Küsters, K.; Wagner, C.; Wiechert, W.; Oldiges, M. Bayesian Optimisation Meets Robotic Workflows: Data-Efficient Phenotyping of Catalytically Active Inclusion Bodies, in German Conference on Synthetic Biology. 2021: virtual, Germany.

\*Helleckes, L.; Osthege, M.; Wiechert, W.; Oldiges, M. Data-efficient phenotyping of strain collections by Bayesian inference of key performance indicators, in Himmelfahrtstagung on Bioprocess Engineering 2021 - New Bioprocesses, New Bioproducts. 2021: virtual, Germany. "Poster Prize"

Jadebeck, J. F. ; Paul, R. D. ; Wiechert, W.; Nöh, K. Automated and distributed parameter space sampling for inverse problems in systems biology, in German Conference on Synthetic Biology. 2021: virtual, Germany.

**Jadebeck, J. F.;** Theorell, A.; Wiechert, W. ; Nöh, K. Netflixing Constraint-Based Sampling with a Microservice Architecture: Robust and Effective Tools for Genome-Scale Challenges, in 7th Conference on Constraint-Based Reconstruction and Analysis. 2021: virtual, Germany.

Köhler, C. A., Ulianych, D., Gerkin, R. C., Davison, A. P., Grün, S., Denker, M. Facilitating the sharing of data analysis results through in-depth provenance capture. 5th HBP Student Conference on Interdisciplinary Brain Research, online, 2021.

Köhler, C. A., Ulianych, D., Gerkin, R. C., Davison, A. P., Grün, S., Denker, M. Capturing detailed provenance information in the analysis of electrophysiology data. 14th Göttingen Meeting of the German Neuroscience Society, online, 2021.

Küsters, K., Wagner, C., \*Helleckes, L., Saborowski, R., Wiechert, W. Accelerated production and screening of catalytically active inclusion body libraries via automated workflows in German Conference on Synthetic Biology, GCSB 2021, virtual, Germany, 13 Sep 2021 - 17 Sep 2021.

\*Malzacher, S.; Warmelink, P.; Lohmann, J.; Pleiss, J.; Rother, D., BioCatHub, a platform for FAIR data acquisition in biocatalysis, in Biotrans 2021. 2021: virtual, Germany.

\*Malzacher, S. BioCatHub, a graphical user interface for standardized data acquisition in biocatalysis. Process Net 2021.

\*Malzacher, S. BioCatHub, a research data management platform for biocatalysis based on the FAIR data principles. MECP-conference 2021.

Neubacher, C. Multi-decadal offshore wind power variability can be mitigated through optimized European allocation. EGU General Assembly 2020.

Osthege, M., \*Helleckes, L., Wiechert, W., Oldiges, M. Efficient screening of strain collections with Bayesian inference and Thompson sampling in Himmelfahrtstagung on Bioprocess Engineering 2021 - New Bioprocesses, New Bioproducts, virtual, Germany, 10 May 2021 - 12 May 2021.

Osthege, M., \*Helleckes, L., Müller, C. Tenhaef, N., Hemmerich, J., Wiechert, W., Oldiges, M. Quantitative autonomous experimental - are we there yet? in German Conference on Synthetic Biology, GCSB 2021, virtual, Germany, 13 Sep 2021 - 17 Sep 2021.

\*Patakchi Yousefi, K. Deep Learning of Model- and Reanalysis- Based Data Mismatches: A case study for Learning from the Precipitation Mismatches over Europe. Helmholtz Herbst Hackathon 2021.

Puchta, D., \*Helleckes, L., Müller, C., Czech, H., Wiechert, W., Oldiges, M. A wholistic approach to autonomous strain characterization: Automation of cryo- and pre-cultures in high-throughput phenotyping in German Conference on Synthetic Biology, GCSB 2021, virtual, Germany, 13 Sep 2021 - 17 Sep 2021.

\*Rittig, J., Schweidtmann, A.M., König, A., Ackermann, P., Dahmen, M., Grohe, M., Mitsos, A. Graph Neural Networks for Prediction of Fuel Ignition Quality In: Fuel Science -From Production to Propulsion Aachen online 2020-06-23 - 2020-06-25.

Ruzaeva, K., Küsters, K., Oldiges, M., Berkels, B., Nöh, K. A hybrid approach for automated characterization of CatIBs production in a biotechnological screening system. BioImage Informatics 2021, Institut Pasteur virtual, France, 2021.

Samadi, M. A training strategy for hybrid models to break the curse of dimensionality: An application to COVID-19 data. 4th Machine Learning and AI in Bio(Chemical) Engineering, Cambridge (Online) 2021.

Simson, A. and Kowalski, J. Constrain melting robots ambient environment based on in-situ data and process models. AGU Fall Meeting 2021, hybrid (New Orleans).

\*Seiffarth, J., Nöh, K. SegUI: Creating high-quality image annotation data sets in microbial bioimaging in BioImage Informatics 2021, Institut Pasteur, virtual, France, 29 Nov 2021 - 1 Dec 2021

Sternfeld, F., Jadebeck, J. F., Wiechert, W., Nöh, K. 13C-Metabolic flux analysis with Bayesian statistics: Effects of early stopping isotope labeling experiments. Quantitative Modelling of Cell Metabolism - International Conference & Workshop, virtual (Denmark), 2021-09-20 - 2021-09-22

van der Weg, K. J., Gohlke, H. Designing 3D Protein Descriptors for Use in Convolutional Neural Networks in Helmholtz AI Virtual Conference 2021.

#### FURTHER NEWS FROM HDS-LEE

#### » Awards

Johannes Seiffarth has been selected by the award jury of the projects, graduate schools and study programs. Computer Science Department of RWTH Aachen University for the andrena award for his outstanding master thesis entitled Together with the project partners, the HDS-LEE PI Julia "Instance Segmentation and Tracking in Microbial Live-Cell Kowalski successfully obtained the "KI Lighthouse Project" Imaging: Boosting Automated Pipelines." The thesis deals of the German Federal Ministry for the Environment. The aim with the automated analysis of large amounts of image data of the three-year project with the title "KI:STE- AI Strategy using state-of-the-art machine learning methods and thus for Earth System Data" is to analyze and process data on the addresses one of the biggest bottlenecks in microfluidic sinstate of nature and the environment with the help of artificial gle-cell analysis. His preliminary work in his master's thesis intelligence and to make it available to the public. The "AI is the basis for his doctoral dissertation in HDS-LEE. The prize, Lighthouses for Environment, Climate, Nature and Resources" endowed with 750 euros, was presented during the "Day of funding initiative supports projects that use artificial intel-Computer Science" event on December 03, 2021. There was a ligence to address environmental challenges. Participating digital congratulation from the founding company and a video in KI:STE are RWTH Aachen University and project partners greeting to all prizewinners. from Forschungszentrum Jülich, the University of Cologne, Rheinische Friedrich-Wilhelms-Universität Bonn, 52° North L. Helleckes received the InnovationPlus Award from Initiative for Geospatial Open Source Software GmbH, and Forschungszentrum Jülich for rtlive.de together with Michael Ambrosys GmbH. Five PhD researchers from the KI:STE proj-Osthege. With the InnovationPlus Award, Forschungszentrum ect have already started working on their projects and thus Jülich recognizes the personal efforts of individual employgained access to the education program at HDS-LEE. They will ees who have made a contribution to innovation at Jülich. in turn strengthen the HDS-LEE community.

L. R. Gorjão was awarded a PostDoc position in Japan by the Prof. Alexander Mitsos successfully applied for fund-JSPS (Japan Society for the Promotion of Science), for 4 ing by the **DFG priority program** with the title "Machine months, for a total of ~1.85M¥ (~14.5k€). Learning in Chemical Engineering - Knowledge Meets Data: Interpretability, Extrapolation, Reliability, Trust". His doctoral researcher of this program, Jan Rittig, was admitted to HDS-» Scientists at HDS-LEE Have Acquired New Large-scale LEE in 2021.

## Al Proiects:

HDS-LEE PI Björn Usadel and his research group at the Jülich The project "Tackling the segmentation and tracking chal-Institute of Bio- and Geosciences are involved in the recently lenges of growing colonies and microbial diversity (SATOMI)" initiated research consortium "DataPLANT" (NFDI4Plants). led by HDS-LEE PI Prof. Wolfgang Wiechert (FZJ), together "DataPLANT", which is funded by the German Research with Prof. Hanno Scharr (FZJ) is funded by the Helmholtz Imaging Platform (HIP) from 05.2021-04.2023. The involved Foundation (DFG) over a period of five years and in a volume doctoral researcher Johannes Seiffarth was associated in of 10 million euros, aims to develop a service and data infrastructure that will enable modern plant research to collect HDS-LEE in 2021 and provide large amounts of data in compliance with FAIR The sketch "BioTwins: Digital twins for the life sciences" was

data principles. submitted to NRW MKW with Prof. Alexander Mitsos in the Two HDS-LEE PIs Giulia Rossetti (FZ Jülich) and Andreas lead. BioTwins is connected to the data science and life part Schuppert (RWTH Aachen) have joined forces together with of HDS-LEE including several PIs from HDS-LEE. At the end of their colleague Dorit Merhof (RWTH Aachen) to initiate the 2021, the involved HDS-LEE PIs Prof. Alexander Mitsos, Prof. Simulation Data Lab "Digital Patient". In this Simulation Data Wolfgang Wiechert, Prof. Holger Gohlke Prof. Axel Klawonn, Lab (SDL), which is funded over a period of ten years, a perand Dr. Achim Basermann were invited to submit a full applisonalized model of a patient is to be created that can be cation of the BioTwins draft. If it works out, the BioTwins projused for diagnostics, prediction of disease progression and ect could start in August 2022 for 4 years and would mean therapy results. The SDL "Digital Patient" is located within an incredible benefit for HDS-LEE as the concept of BioTwins

the research consortium NHR4CES (NHR for Computational Engineering Science). Here, RWTH and TU Darmstadt cooperate to promote computational engineering science in joint aims to support the training of young data scientists from doctoral researchers to junior research group leaders.

#### » Communication/Marketing

HDS-LEE has a **website** (www.hds-lee.de) and a **Twitter account** with 175 national and international followers and about 5 posts per month. HDS-LEE uses **advertising services**. In 2021, the services of academic positions were used with posts on LinkedIn, Twitter, Facebook and Instagram. Additionally their social media campaign addressed specifically appropriate people on Facebook, Instagram, and Facebook Audience Network with a reach of 38.161 people who saw the advertisement. To address female scientists, the Diversity and Inclusion Recruitment of sciencejobs.org was used.

HDS-LEE used the **job-advertising mailing list** (Hds-leenewsletter@fz-juelich.de) to advertise the open positions that were published in 2021 for the second cohort. Fourteen people are subscribed to this mailing list.

The **event-advertising mailing list** for scientist interested in talks and workshops organized by the school (Hds-leeevent@fz-juelich.de) has 31 subscriptions in 2021.

HDS-LEE was present with a virtual booth at the Helmholtz Virtual Data Science Career Day on September 26, 2021 and participated in the matching session with HDS-LEE PIs. HDS-LEE PI Prof. Julia Kowalski presented HDS-LEE in the session "A PhD - Saving Lives and the Earth with Data Science".

HDS-LEE presented itself at the DAAD Networking Tour in Artificial Intelligence (PostDoc Net AI) on April 20, 2021. A selected group of outstanding young international researchers visited the talks of several HDS-LEE PIs, who gave an overview of their AI research field and presented their university and research center structure. In the matching session, they met the PIs and had the opportunity to talk to them one-onone. The main goal of the PostDoc Net AI is to connect these young international researchers to the German research community to set the base for new collaborations, extend their network, and to initiate concrete career opportunities.

### MUNICH SCHOOL FOR DATA SCIENCE HELMHOLTZ | TUM | LMU

#### 3. MUNICH SCHOOL FOR DATA SCIENCE @ HELMHOLTZ, TUM & LMU (MUDS)

MUDS trains data scientists in the Munich Metropolitan Region. Here, the Technical University and the Ludwig Maximilian University in Munich as well as Helmholtz Munich and the German Aerospace Center (DLR) have joined forces with the Max Planck Institute for Plasma Physics to form an internationally visible and highly attractive research network. The MUDS also cooperates with the Leibniz Computing Centre (LRZ) and the Max Planck Computing & Data Facility (MPCDF) and works with Roche Penzberg to promote application-oriented doctoral projects in biomedicine.

#### APPLICANT SITUATION AND RECRUITMENT

In 2021, MUDS received 393 applications in two calls (28% female, 58% international). A total of 25 PhD positions were filled (8 female, 18 international), 7 of them funded by MUDS (1 female, 4 international), and 19 as associated PhD researchers (7 female, 14 international). 1 position (male, international) is funded in a collaboration with Roche.

#### EVENTS AND NETWORKING

#### » 3 Networking Events

- MUDS Networking Event, 07.07.2021, in person event, 48 participants, scientific networking across different institutions within the MUDS community, MUDS questions/ answers session
- MUDS Doctoral Symposium, 28.09.2021, in person event, 50 participants, doctoral researcher & PI talks, poster session, get together
- MUDS Welcome Event (class 2021), 11.11.2021, 26 participants, Welcome, introduction to PhD administrative matters, Q&A, joint lunch

#### » 2 Recruiting Events

- MUDS Recruiting Event 2020/2021, 08.03. 09.03.2021, virtual, 62 participants annual recruiting of doctoral researchers
- MUDS Recruiting Event 2020/2021, 20.-22.07.2021, 54 participants virtual, annual recruiting of doctoral researchers

#### » 4 Lectures

- Data Science Block Course spring, 02.03. 25.03.2021
   (18 x half day + QA session + exam day), virtual, 14
   participants, 2 week online block course covering data
   science-related topics
- Data Science Block Course fall, 08.11. 02.12.2021 (18 x half day + QA session + exam day), virtual, 24 participants, 2 week online block course covering data science-related topics
- Life Science Prime Course, 11.-12.05., 19.-21.05., 25.-27.05.2021 (8 full days), virtual, 17 participants, course on basic life science/biology concepts and experimental methods for doctoral researchers with a more theoretical/computational background
- HIDA Lecture @MUDS, 29.09.2021, Niki Kilbertus (Helmholtz AI München) "(How) can we find true cause effect relationships in data? — An introduction to the instrumental variable setting"

#### » 5 Transferable Skills Courses

- Presentations Skills Course 1, 13.04., 20.-21.04.21, virtual, 12 participants, course on how to present research
- > Presentations Skills Course 2, 22.06., 25.06., 29.06.21, virtual, 12 participants, course on how to present research
- Scientific Writing Course, 17.-18.05.2021, virtual, 5 participants, course on how to communicate research clearly and effectively in a research article.(shared course with an LMU PhD program)
- Stress Release Workshop 1, 22.02.21 + 01.03.21 (each 1,5 h), virtual, 12 participants, course to learn strategies to help dealing with stress and anxiety and increase concentration
- Stress Release Workshop 2, 05.03.21 + 12.03.21 (each 1,5 h) (each 1,5 h), virtual, 13 participants, course to learn strategies to help dealing with stress and anxiety and increase concentration

#### » 1 Seminar Series MUDS Seminar Series / Doctoral

Researchers Progress Reports, on average biweekly (started 05.08.20 - ongoing), virtual, talks of guest speakers invited by doctoral researchers and progress reports by MUDS doctoral researchers

#### » 2 Co-organized by MUDS

- Dynamic Earth Net Challenge, 01.03.2021 01.06.2021: http://www.classic.grss-ieee.org/earthvision2021/challenge.html
- Responsible Research Symposium, 25.-26.03.2021: https://www.responsibleresearch.graduatecenter.unimuenchen.de/event2021/index.html

#### PUBLICATIONS

 $\ensuremath{\mathsf{MUDS}}$  doctoral researchers are  $\ensuremath{\mathsf{highlighted}}$  , associated doctoral researchers are marked with asterisk.

#### » 14 First-author Peer-reviewed Journal Articles and Conference Proceedings

Boniolo F., **Dorigatti E.**, Ohnmacht A.J., Saur D., Schubert B., Menden M.P. (2021). Artificial intelligence in early drug discovery enabling precision medicine. Expert Opinion on Drug Discovery, 1-17.

**Geisler S.**, Schmidt T., Şirin H., Zügner D., Bojchevski A., Günnemann S. (2021). Robustness of Graph Neural Networks at Scale. Thirty-Fifth Conference on Neural Information Processing Systems, 2021.

**Geisler S.**, Zügner D., Bojchevski A., Günnemann S. (2021). Attacking Graph Neural Networks at Scale. DLG workshop @ AAAI

\*Hetzel L., Fischer D.S., Günnemann S., Theis F.J. (2021). Graph representation learning for single-cell biology. Current Opinion in Systems Biology Vol 28, Dec 2021, 100347 https://www.sciencedirect.com/science/article/pii/ S2452310021000329

**Hrovatin K.**, Fischer D.S., Theis F.J. (2021). Toward modeling metabolic state from single-cell transcriptomics. Molecular Metabolism, in press, available online 14 Nov 2021 DOI: https://doi.org/10.1016/j.molmet.2021.101396.

Kondmann L., Toker A., Saha S., Schölkopf B., Leal-Taixé L., Zhu XX. (2021). Spatial Context Awareness for Unsupervised Change Detection in Optical Satellite Images. IEEE Transactions on Geoscience and Remote Sensing DOI: 10.1109/TGRS.2021.3130842.

Kondmann L., Toker A., Rußwurm M., Camero A., Peressuti D., Milcinski G., Mathieu P.P., Longépé N., Davis T., Marchisio G., Leal-Taixé L., Zhu XX. (2021). DENETHOR: The DynamicEarthNET dataset for Harmonized, inter-Operable, analysis-Ready, daily crop monitoring from space. In Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track. (This is more of a journal paper actually but technically counts as a conference.)

**Kondmann L.**, Taubenböck H., Zhu XX. (2021). Blinded by the Light: Monitoring Local Economic Development Over Time With Nightlight Emissions. In IEEE International Geoscience and Remote Sensing Symposium (IGARSS). (Conference paper)

**Kondmann L.**, Zhu XX. (2021). Under the Radar-Auditing Fairness in ML for Humanitarian Mapping. In 2nd Datadriven Humanitarian Mapping Workshop at KDD. (Conference paper)

**Kondmann L., Toker A.,** Marc Rußwurm, Andrés Camero, Devis Peressutti, Grega Milcinski, Nicolas Longépé, Pierre-Philippe Mathieu, Timothy Davis, Giovanni Marchisio, Laura Leal-Taixé, and Xiao Xiang Zhu.

**Matos F.**, Menkovski V., Pau A., Marceca G., Jenko F. and the TCV Team (2021). Plasma confinement mode classification using a sequence-to-sequence neural network with attention. Nuclear Fusion, Vol 61, No.4 046019.

Rath K., Albert C.G., Bischl B., von Toussaint U. (2021). Symplectic Gaussian process regression of maps in Hamiltonian systems. Chaos 2021, 31, 053121. https://doi. org/10.1063/5.0048129.

**Rath K.,** Albert C.G., Bischl B., von Toussaint U. (2021) Orbit Classification and Sensitivity Analysis in Dynamical Systems Using Surrogate Models. Phys. Sci. Forum 2021, 3, 5. https://doi.org/10.3390/psf2021003005. Toker A., Zhou Q., Maximov M., Leal-Taixe L. (2021). Coming Down to Earth: Satellite-to-Street View Synthesis for Geo-Localization. Conference on Computer Vision and Pattern Recognition (CVPR). Also featured on CVPR daily magazine as BEST of CVPR https://www.rsipvision.com/ ComputerVisionNews-2021July/26/

#### » 4 Co-author Peer-reviewed Journal Articles and Conference Proceedings

Lee J., **Feng J.**, Humt M., Müller M.G., Triebel R. (2021). Trust Your Robots! Predictive Uncertainty Estimation of Neural Networks with Sparse Gaussian Processes. In: Conference on Robot Learning (CoRL). Proceedings of Machine Learning Research (PMLR). 5th Conference on Robot Learning (CoRL), London, United Kingdom. ISSN 2640-3498.

Saha S., **Kondmann L.**, Song Q., Zhu XX. (2021). Change Detection in Hyperdimensional Images using Untrained Models. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, in press.

Saha S., **Kondmann L.**, Zhu XX. (2021). Deep No Learning Approach for Unsupervised Change Detection in Hyperspectral Images. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Conference Paper).

Stadler M., Charpentier B., **Geisler S.,** Zügner D., Günnemann S. (2021). Graph Posterior Network: Bayesian Predictive Uncertainty for Node Classification. Thirty-Fifth Conference on Neural Information Processing Systems, 2021.

#### » Conference Proceedings

#### 4 Conference Talks

Artigues V., Jenko F., & JET Contributors (2021). Multi-class disruption prediction at JET using a shapelet based neuralnetwork. Talk presented at DPG-Tagung der Sektion Materie und Kosmos (SMuK). Virtual. 2021-08-30 - 2021-09-03. http://hdl.handle.net/21.11116/0000-0009-1DF7-1.

**Wagner N.,** Celik M., Gagneur J. (2021). Enhanced SpliceMap and RNA-seq from clinically accessible tissues improves outlier prediction for non-accessible tissues. Talk at European Human Genetics Conference 2021. **Wagner N.,** Celik M., Hölzlwimmer F., Gagneur J. (2021). Enhanced SpliceMap and RNA-Seq from Clinically Accessible Tissues Improves Outlier Prediction for Non-Accessible Tissues. 49th European Mathematical Genetics Meeting (EMGM) 2021.

Yudin Y., Lakhlili J., Luk Oo., von Toussaint U., Coster DP. (2021). Gaussian Process Surrogate Models for Uncertainty Quantification in Multiscale Fusion. International Conference on Computational Science 2021, Krakow, Poland.

#### 6 Poster Presentations

**Bergmann M.**, Fischer, R., Jenko F. (2021). Introduction of quasilinear transport models to the Integrated Data Analysis framework. Poster presented at DPG-Tagung der Sektion Materie und Kosmos (SMuK), Virtual.<u>http://hdl.</u> handle.net/21.11116/0000-0009-1E16-E.

\*Fediukov, V. (2021). Data Science for Terramechanics. Helmholtz AI virtual conference.

\*Oleshko S., Marsico A. (2021). Computational mapping of the human SARS-CoV-2 protein-RNA interactome. Network Biology Cold Spring Harbor Laboratory, 16-19 March 2021 (virtual).

\*Oleshko S. (2021). Integration of multi-omics data using graph neural networks to identify and contextualize biomarker genes for neuropsychiatric disorders. EMBO | EMBL Symposium: Multiomics to Mechanisms: Challenges in Data Integration. 15-17 Sep. 2021

\*Oleshko S. (2021). Integration of multi-omics data using graph neural networks to identify and contextualize biomarker genes for psychiatric disorders. 23rd EMBL PhD Symposium: The Big Picture: Zooming into Life. 16-17 Dec 2021.

\*Stock M., Scialdone A. (2021). Self-supervised completion of gene regulatory networks using graph autoencoders. German Stem Cell Network conference 2021.

#### FURTHER NEWS FROM MUDS

#### » Software License

The Software JessEV, developed and published by MUDS doctoral researcher Emilio Dorigatti, (Dorigatti E, Schubert B (2020) Joint epitope selection and spacer design for stringof-beads vaccines. Bioinformatics, Vol. 36, i643-i650) has been licensed for non-academic users in 2021.

#### » Industry Track

The scientific and legal basis for three new MUDS PhD projects involving industry collaborations were aligned in 2021, which are planned to be filled with candidates in 2022. This adds up to the ongoing 4 industry collaborations within MUDS to a total of 7.

#### » Additional Funding

The recruitment of additional 19 associated MUDS PhD researchers demonstrates the added benefit for the Information & Data Science network in the Munich community.

#### » Special Networking Events

MUDS was proud to host an in-person Doctoral Symposium in September at the Technical University Munich.

#### » Awards and Prizes

MUDS doctoral researcher Karin Hrovatin and MUDS associated doctoral researcher Marco Stock were awarded add-on fellowships for interdisciplinary life science by the Joachim Herz Stiftung, an award for talented interdisciplinary young scientists

MUDS doctoral researcher Aysim Toker's work was featured at the Conference on Computer Vision and Pattern Recognition (CVPR) in the CVPR daily magazine as "best of CVPR" https://www.rsipvision.com/ComputerVisionNews-2021 July/26/

#### » Special Lectures

MUDS was proud to host Dr. Niki Kilbertus of Helmholtz AI in a talk about "(How) can we find true cause effect relationships in data? - An introduction to the instrumental variable setting", in the HIDA lecture@MUDS. Steering board member Göran Kauermann (LMU) was chairing the session.

#### » Publicity

MUDS was highlighted on the HIDA website in an article by Anja Dilk: https://www.helmholtz-hida.de/hida-news/ fuer-die-forschung-brennen-promovieren-an-der-muds/

Portraits of MUDS doctoral researcher Karin Hrovatin and Laura Martens were published on the HIDA website: Karin Hrovatin, 'Hope for diabetics': https://www.helmholtz-hida. de/en/hida-news/hope-for-diabetics/ and Laura Martens, 'Healing by gene therapy?': https://www.helmholtz-hida. de/en/hida-news/healing-by-gene-therapy/



Researchers with perspectives: PhD candidates at the MUDS in Munich. (Photo: Julia Schlehe/Munich School for Data Science)



#### **4. HELMHOLTZ INFORMATION &** DATA SCIENCE SCHOOL FOR HEALTH (HIDSS4HEALTH)

The goal of HIDSS4Health is to train and promote the best » 21 Networking Events young talent at the interface between Information & Data Science and health. HIDSS4Health offers a structured doc-23.04.2021, Selection Event 2021, Online (65 participants): toral program with close integration into an interdisciplinary Selection Event with talks from the applicants, short project network of leading scientists from both the data and life pitches and personal interviews. sciences

Research Areas: Imaging & diagnostics, surgery & intervention 4.0 and models for personalized medicine.

Partners of HIDSS4Health: Karlsruhe Institute of Technology, German Cancer Research Center Heidelberg, and Heidelberg University

#### APPLICANT SITUATION AND RECRUITMENT

There are currently 32 PhD positions filled at HIDSS4Health, 11 of which were awarded in 2019, nine in 2020 and 12 in 2021. In 2021, of the School received 172 applications, of which 144 came from abroad. Of the 2021 accepted doctoral researchers, 75% (9) are German and 17% (2) female.

Of the 32 doctoral researchers in total, 78% (25) are German, and the proportion of women is 34% (11). Two additional female researchers and four additional male researchers are associated with HIDSS4Health; they are from KIT (2), DKFZ (3), and Heidelberg University (1).

### EVENTS AND NETWORKING

In 2021, HIDSS4Health hosted 61 events (21 networking, 8 courses, 32 lectures).

01.07.2021, "From Science to Start-up to Siemens - The Innovation Journey of AI", Online (35 participants): Event for doctoral researchers interested in business aspects of AI and in questions how to translate ideas into products.

27.07.2021, "Ethics and AI: How to Successfully Integrate Both into Businesses", Online (26 participants): Event for doctoral researchers interested in ethical aspects of AI in business applications.

20.09.-22.09.2021, Retreat, Bad Herrenalb (28 participants): Retreat of the doctoral researchers to network, get to know each other and discuss their projects and possible collaborations

27.09.2021, Doctoral Researcher General Assembly 2021, DKFZ (21 participants): Assembly of the doctoral researchers, where organizational aspects of the school are discussed and representatives for the General Assembly and the Steering Committee are elected.

06.10.2021, General Assembly 2021, Online (41 participants): Assembly of the HIDSS4Health "parliament", consisting of all PIs, two postdocs and one selected doctoral researchers per cohort. Here, the current state of the school is discussed and members of the Steering Committee and new PI candidates are elected.

In addition: 15 regular meetings of HIDSS4Health doctoral researchers online (every 3 weeks).

#### » 8 Courses

28.04.-29.04.2021, "Visualising Science", Online (12 participants): The seminar "Visualising Science" addresses basic design principles and methods to present research results in a simpler, clearer and more lively way.

17.05.-19.05.2021, "Presentation Skills", Online (8 partici- > pants): The seminar "Presentation Skills" addresses core aspects of planning and holding a presentation in various > forms and for different target groups.

18.06.2021, "Managing Up", Online (15 participants): "Managing Upward: Optimizing Student-Supervisor > Communication to Support Wellbeing" is a workshop that introduces to strategies and techniques that will improve the ability to set expectations, establish and renegotiate healthy boundaries, and anticipate and address issues that arise. This course was conducted jointly with HDS-LEE. >

15.07.-16.07.2021 & 19.07.2021, "Scientific Writing Course", Online (12 participants): The seminar "Publishing in scientific journals" covers all aspects of writing, submitting, reviewing and publishing a scientific manuscript.

28.09. + 20.10.2021, "Time & Self-Management", DKFZ/Online (7 participants): This workshop focuses on optimizing the personal time and self-management during the PhD period.

22.10. + 25.10.2021, "Gitlab", Online (18 participants): The "Gitlab" workshop covers the basic tools required for a research software workflow: working with the Shell, Git and project management with GitLab. The course was organized together with HIFIS.

22.-26.11.2021, "Python from 0 to Data Analysis", Online (34 participants): This workshop covers introduction to Python and the two packages pandas and matplotlib. It was conducted jointly with HIFIS, and the two partner graduate schools IHRS Biosoft (FZJ) and BIF-IGS (KIT).

#### » 16 Lectures as Part of the "Data Science & Health" Lecture Series

The lectures were all given online and were open for people affiliated with Helmholtz or universities. The mean number of participants in each lecture was 44, where approx. 29% were from HIDSS4Health, 60% from Helmholtz and 11% from outside Helmholtz. The participants came from 27 different Helmholtz institutions and universities. On average, 37% of the participants were female.

- > 11.01.2021, Michael Beigl (KIT): "Wearable and Mobile Health Data"
- > 11.01.2021, Filip Sadlo (Heidelberg University): "Introduction to Visual Data Science"

- 25.01.2021, Peter Sanders (KIT): "Parallel Algorithms for Dummies"
- 25.01.2021, Achim Streit (KIT): ""ntroduction to Distributed and Parallel Computing"
- 08.02.2021, Lena Maier-Hein (DKFZ): "Surgical Data Science"
- 08.02.2021, Michael Gertz (Heidelberg University):
   "Trends and Topics in Text Analytics"
- > 19.10.2021, Klaus Maier-Hein (DKFZ): "Medical Image Computing"
- > 26.10.2021, Ralf Mikut (KIT): "Time Series Analysis"
- O2.11.2021, Alexander Schug (FZJ): "Data Inference on Sequence Data: How to predict biomolecular structure and function"
- > 09.11.2021, Michael Beigl (KIT): "Wearable Health Data"
- > 16.11.2021, Mark Ladd (DKFZ): "Imaging Physics in Oncology"
- > 23.11.2021, Pascal Friederich (KIT): "Graph Neural Networks for Molecular Design"
- > 30.11.2021, Anne Kaster (KIT): "Single Cell Omics"
- 07.12.2021, Lennart Hilbert (KIT): "Fluorescence Microscopy and Digital Image Processing in Molecular Cell Biology"
- > 14.12.2021, Franziska Mathis-Ullrich (KIT): "Co-Operation with Surgical Robots"
- 21.12.2021, Oliver Jäkel (DKFZ): "Why Data Sciences Will Be Crucial for Modern Image Guided Radiotherapy"

#### » 16 Lectures as Part of the "Advanced Topics in Data Science & Health" Lecture Series

The lectures were given online by the doctoral researchers and Postdocs of HIDSS4Health and were open for all HIDSS4Health doctoral researchers. The average number of participants was 18.

- > 13.04.2021, Jens Petersen (DKFZ): "Transformers for Vision Models"
- 13.04.2021, Ines Reinartz (KIT): "Metaheuristic Optimization Methods"
- > 20.04.2021, David Zimmerer (DKFZ): "Distribution Shift / Out-of-distribution Uncertainty + Why Do I Need It & How Can I Use It?"
- > 20.04.2021, Pia Stammer (KIT): "The Monte Carlo Method and its Applications"
- > 04.05.2021, Silvia Seidlitz (DKFZ): "How to Deal with Confounders in Your Dataset"
- > 04.05.2021, Jan Sellner (DKFZ): "Unsupervised and Supervised Dimensionality Reduction"

- > 11.05.2021, Katharina Löffler (KIT): "Object Tracking"
- > 11.05.2021, Paula Breitling (KIT): "Wearables in the Health Domain"
- > 25.05.2021, Yaroslav Agapov (Heidelberg University): "Parameter Sensitivity Visualization for Ensemble MRI Simulations"
- 25.05.2021, Vojtech Kumpost (KIT): "Theoretical Chronobiology"
- O1.06.2021, Hamideh Hajiabadi (KIT): "Introduction to AutoML"
- > 01.06.2021, Constantin Seibold (KIT): "Weakly Supervised Detection"
- > 15.06.2021, Patrick Godau (DKFZ): "Meta Learning"
- > 15.06.2021, Rene Snajder (DKFZ): "Language Modeling"
- > 22.06.2021, Mikael Beyene (KIT): "Consensus Mechanisms in Distributed Ledger Technology"
- > 22.06.2021, Melanie Schellenberg (DKFZ): "Learning-to-Simulate" Approaches for Medical Imaging"

#### PUBLICATIONS

Doctoral researchers from HIDSS4Health are highlighted.

#### » 9 First-author Publications (doctoral researchers)

Seibold, C., Fink, M.A., Goos, C., Kauczor, H., Schlemmer, H., Stiefelhagen, R., Kleesiek, J. (2021). Prediction of low-keV monochromatic images from polyenergetic CT scans for improved automatic detection of pulmonary embolism. IEEE International Symposium of Biomedical Imaging, https://doi.org/10.1038/s41746-021-00439-y.

**Löffler, K.**, Scherr, T., Mikut, R. (2021). A graph-based cell tracking algorithm with few manually tunable parameters and automated segmentation error correction. Plos One, https://doi.org/10.1371/journal.pone.0249257.

**Stammer, P.**, Burigo, L., Jäkel, O., Frank, M. and Wahl, N. (2021). Efficient uncertainty quantification for Monte Carlo dose calculations. Physics in Medicine and Biology, https://doi.org/10.1088/1361-6560/ac287f\_

Petersen, J., Köhler G., Zimmerer, D., Isensee, F., Jäger,<br/>P.F., Maier-Hein, K. (2021). Better generalization for condi-<br/>tional convolutional Neural Processes on time series data.Photonics West 2021: Photons Plus Ultrasound: Imaging and<br/>Sensing 2021.Proceedings of Machine Learning Research (Conference<br/>Proceeding).Godau, P., Maier-Hein, L. (2021). Task Fingerprinting for Meta<br/>Learning in Biomedical Image Analysis, MICCAI 2021.

**Petersen, J.**, Isensee, F., Köhler G., Jäger. P.F., Zimmerer, D., Neuberger, U., Wick, W., Debus, J., Heiland, S., Bendszus, M., Vollmuth, P., Maier-Hein, K. (2021). Continuous-Time Deep Glioma Growth Models. Medical Image Computing and Computer Assisted Intervention - MICCAI 2021.

**Warsinsky, S.**, Schmidt-Kraepelin, M., Rank, S., Thiebes, S., Sunyaev, A. (2021). Conceptual Ambiguity Surrounding Gamification and Serious Games in Health Care: Literature Review and Development of Game-Based Intervention Reporting Guidelines (GAMING). Journal of Medical Internet Research.

**Warsinsky, S.**, Schmidt-Kraepelin, M., Rank, S., Thiebes, S., Sunyaev, A. (2021). Are Gamification Projects Different? An Exploratory Study on Software Project Risks for Gamified Health Behavior Change Support Systems. Proceedings of the 54th Hawaii International Conference on System Sciences.

Scheikl, P.M., Gyenes, B., Davitashvili, T., Younis, R., Schulze, A., Müller-Stich, B.P., Neumann, G., Wagner, M., Mathis-Ullrich, F. (2021). Cooperative Assistance in Robotic Surgery through Multi-Agent Reinforcement Learning. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2021).

## » 3 First-author Publications (associated doctoral researchers)

**Kumpošt, V.**, Vallone, D., Babu Gondi, S., Foulkes, N.S., Mikut, R., Hilbert, L. (2021). A stochastic oscillator model simulates the entrainment of vertebrate cellular clocks by light. Scientific Reports, https://doi.org/10.1038/s41598-021-93913-2.

**Schellenberg, M.**, Gröhl, J., Dreher, K., Holzwarth, N., Tizabi, M.D., Seitel, A. and Maier-Hein, L. (2021). Generation of training data for quantitative photoacoustic imaging. SPIE Photonics West 2021: Photons Plus Ultrasound: Imaging and Sensing 2021.

#### » 12 Co-author Publications (doctoral researchers)

Salg, G.A., Ganten, M., Bucher, A.M., Kenngott, H.G., Fink, M.A., **Seibold, C.**, Fischbach,R., Schlamp, K., Velandia, C.A., Fervers, P., Doellinger, F., Luger, A., Afat, S., Merle, U., Diener, M.K., Pereira, P.L., Penzkofer, T., Persigehl, T., Othman, A., Heußel, C.P., Baumhauer, M., Widmann, G., Stathopoulos, K., Hamm, B., Vogl, T.J., Nikolaou, K., Kauczor, H., Kleesiek, J. (2021). A reporting and analysis framework for structured evaluation of COVID-19 clinical and imaging data. npj Digital Medicine.

Marinov, Z., Vasileva, S., Wang, Q., **Seibold, C.**, Zhang, J., Stiefelhagen, R. (2021). Pose2Drone: A Skeleton-Pose-based Framework for Human-Drone Interaction. Proceedings of European Signal Processing Conference.

Tan, H., Chen, C., Luo, X., Zhang, J., **Seibold, C.,** Yang, K., Stiefelhagen, R. (2021). Flying Guide Dog: Walkable Path Discovery for the Visually Impaired Utilizing Drones and Transformer-based Semantic Segmentation. Proceedings of IEEE International Conference on Robotics and Biomimetics.

Roitberg, A., Schneider, D., Djamal, A., **Seibold, C.**, Reiß, S., Stiefelhagen, R. (2021). Let's Play for Action: Recognizing Activities of Daily Living by Learning from Life Simulation Video Games. Proceedings of IEEE International Conference on Intelligent Robots and Systems, https://doi.org/10.1148/ radiol.2021211013.

Fink, M.A., Mayer, V.L., Schneider, T., **Seibold, C.**, Stiefelhagen, R., Kleesiek, J., Weber, T.F., Kauczor, H. (2021). CT Angiography Clot Burden Score from Data Mining of Structured Reports for Pulmonary Embolism. Radiology.

Reiß, S., **Seibold, C.**, Freytag, A., Rodner, E., Stiefelhagen, R. (2021). Every annotation counts: Multi-label deep supervision for medical image segmentation. Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition.

Kades, K., **Sellner, J.**, Koehler, G., Full, P.M., Lai, T.Y.E., Kleesiek, J., Maier-Hein, K.H. (2021). Adapting Bidirectional Encoder Representations from Transformers (BERT) to Assess Clinical Semantic Textual Similarity: Algorithm Development and Validation Study. JMIR Med Inform.

Reinke, A., Eisenmann, M., Tizabi, M.D., Sudre, C.H., Rädsch, T., Antonelli, M., Arbel, T., Bakas, S., Cardoso M.J., Cheplygina, V., Farahani, K., Glocker, B., Heckmann-Nötzel, D., Isensee, F., Jannin, P., Kahn, C., Kleesiek, J., Kurc, T., Kozubek, M.,

Landman, B.A., Litjens, G., Maier-Hein, K., Martel, A.L., Müller, H., **Petersen, J.**, Reyes, M., Rieke, N., Stieltjes, B., Summers, R.M., Tsaftaris, S.A., van Ginneken, B., Kopp-Schneider, A. Jäger, P., Maier-Hein, L. (2021). Common limitations of performance metrics in biomedical image analysis. Medical Imaging with Deep Learning 2021 (Conference Proceeding).

Isensee, F., Jaeger, P.F., Kohl, S., **Petersen, J.**, Maier-Hein, K.H., (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, https://doi.org/10.1038/s41592-020-01008-z.

Preetha, C.J., Meredig, H., Brugnara, G., Mahmutoglu, M.A., Foltyn, M., Isensee, F., Kessler, T., Pflüger, I., Schell, M., Neuberger, U., **Petersen, J.**, Wick, A., Heiland, S., Debus, J., Platten, M., Idbaih, A., Brandes, A.A., Winkler, F., van den Bent, M.J., Nabors, B., Stupp, R., Maier-Hein, K.H., Gorlia, T., Tonn, J., Weller, M., Wick, W., Bendszus, M., Vollmuth, P. (2021). Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. Lancet Digital Health, https://doi.org/10.1016/S2589-7500(21)00205-3\_

Hermann, S.; **Breitling, P.**; Röddiger, T.; Beigl, M. (2021). Cardiopulmonary Resuscitation Support: Comparison of Wrist-, Chest-, and Ear-Worn Devices and Estimation Algorithms. ISWC '21: 2021 International Symposium on Wearable Computers

Besso, M.J., **Bitto, V.**, Koi, L., Hadiwikarta, W., Euler-Lange, R., Bonrouhi, M., Linge, A., Krause, M., Kurth, I., Baumann, M. (2021). OC-0062 Potential predictive biomarkers for Nimorazole-modified radiochemotherapy in head and neck cancer. Radiotherapy and Oncology, http://dx.doi. org/10.1016/S0167-8140(21)06756-6.

## » 9 Co-author Publications (associated doctoral researchers)

Bracher, J., **Wolffram, D.**, Deuschel, J. et al. (2021). A pre-registered short-term forecasting study of COVID-19 in Germany and Poland during the second wave. Nature communications, https://doi.org/10.1038/s41467-021-25207-0.

Bracher, J., **Wolffram, D.**, Gneiting, T. and Schienle, M. (2021). Vorhersagen sind schwer, vor allem die Zukunft betreffend: Kurzzeitprognosen in der Pandemie. Mitteilungen der Deutschen Mathematiker-Vereinigung, https://doi. org/10.1515/dmvm-2021-0073\_ Maier-Hein, L., Wagner, M., Ross, T., Reinke, A., Bodenstedt, S., Full, P.M., Hempe, H., Mindroc-Filimon, D., **Scholz, P.**, Tran, T.N. and others (2021). Heidelberg colorectal data set for surgical data science in the sensor operating room. Scientific data.

Gröhl, J., **Schellenberg, M.**, Dreher, K. and Maier-Hein, L. (2021). Deep learning for biomedical photoacoustic imaging: A review. Photoacoustics.

Holzwarth, N., **Schellenberg, M.**, Gröhl, J., Dreher, K., Nölke, J., Seitel, A., Tizabi, M.D., Müller-Stich, B.P. and Maier-Hein, L. (2021). Tattoo tomography: Freehand 3D photoacoustic image reconstruction with an optical pattern. International Journal of Computer Assisted Radiology and Surgery.

Holzwarth, N., **Schellenberg, M.**, Gröhl, J., Dreher, K.K., Nölke, J., Biegger, P., Tizabi, M.D., Seitel, A. and Maier-Hein, L. (2021). Tattoo tomography: an optical pattern approach for context-aware photoacoustics. SPIE Photonics West 2021: Photons Plus Ultrasound: Imaging and Sensing 2021.

Rachel, Z., Holzwarth, N., Held, T., **Schellenberg, M.**, Regnery, S., Lang, K., Weusthof, K., Maier-Hein, L., Debus, J., Adeberg, S. (2021). Multispectral Optoacoustic Imaging for the Diagnosis and Determination of the Therapy Response in Patients with malignant Head and Neck Tumors. SPIE Photonics West 2021: Photons Plus Ultrasound: Imaging and Sensing 2021.

Gröhl, J., Dreher, K.K., **Schellenberg, M.**, Seitel, A. and Maier-Hein, L. (2021). SIMPA: an open source toolkit for simulation and processing of photoacoustic images. SPIE Photonics West 2021: Photons Plus Ultrasound: Imaging and Sensing 2021.

Gröhl, J., **Schellenberg, M.**, Dreher, K.K., Holzwarth, N., Tizabi, M.D. and Maier-Hein, L. (2021). Semantic segmentation of multispectral photoacoustic images using deep learning. SPIE Photonics West 2021: Photons Plus Ultrasound: Imaging and Sensing 2021.

#### » 20 Preprints/submitted

**Hajiabadi. H.**, Mamontova, I., Prizak, R., Pancholi, A., Koziolek, A., Hilbert, L. (2021). Deep-learning microscopy image reconstruction with quality control reveals second-scale rearrangements in RNA polymerase II clusters. bioRxiv (preprint, submitted)

Seidlitz, S., Sellner, J., Odenthal, J., Özdemir, B., Studier-Fischer, A., Knödler, S., Ayala, L., Adler, T., Kenngott, H.G., Tizabi, M., Wagner, M., Nickel, F., Müller-Stich, B.P., Maier-Hein, L. (2021). Robust deep learning-based semantic organ segmentation in hyperspectral images. arXiv (preprint, submitted)

**Seibold, C.**, Reiss, S., Sarfraz, S., **Sellner, J.**, Mayer, V., Murray, J.M., Moon-Sung, K., Fink, M., Maier-Hein, K., Kleesiek, J., Stiefelhagen, R. (2021). A Case for Anatomy-guided Visual Understanding of Chest-Radiograph Medical Reports. (submitted)

**Schellenberg, M.**, Gröhl, J., Dreher, K., Holzwarth, N., Tizabi, M.D., Seitel, A. Maier-Hein, L. (2021). Data-driven generation of plausible tissue geometries for realistic photoacoustic image synthesis. arXiv (preprint)

**Schellenberg, M.**, Dreher, K.K., Holzwarth, N., Isensee, F., Reinke, A., Seitel, A., Tizabi, M.D., Maier-Hein, L. and Gröhl, J. (2021). Semantic segmentation of multispectral photoacoustic images using deep learning. Arxiv (preprint, submitted)

**Müller, L., Petersen, J.**, Yamlahi, A., Wise, P., Adler, T.J., Seitel, A., Kowalewski, K., Müller, B., Kenngott, H., Nickel, F., Maier-Hein, L. (2021). Robust hand tracking for surgical telestration. IPCAI 2022 (conditionally accepted)

Hutschenreiter, L., **Haller, S.**, Feineis, L., Rother, C., Kainmüller, D., Savchynskyy, B. (2021). Fusion Moves for Graph Matching. arXiv

Feineis, L., **Haller, S.**, Hutschenreiter, L., Bernard, F., Swoboda, P., Rother, C., Kainmüller, D., Savchynskyy, B. (2021). A Comparative Study of Graph Matching Algorithms in Computer Vision (submitted)

Bracher, J., **Wolffram, D.**, Deuschel, J., Görgen, K., Ketterer, J.L., Ullrich, A., Abbott, S., Barbarossa, M.V., Bertsimas, D., Bhatia, S., Bodych, M., Bosse, N.I., Burgard, J.P., Fiedler, J., Fuhrmann, J., Funk, S., Gambin, A., Gogolewski, K., Heyder, S., Hotz, T., Kheifetz, Y., Kirsten, H., Krueger, T., Krymova, E., Leithäuser, N., Li, M.L., Meinke, J.H., Miasojedow, B., Mohring, J., Nouvellet, P., Nowosielski, J.M., Ozanski, T., Radwan, M., Rakowski, F., Scholz, M., Soni, S., Srivastava, A., Gneiting, T., Schienle M. (2021). National and subnational short-term forecasting of COVID-19 in Germany and Poland, early 2021. medRxiv, https://doi.org/10.1101/2021.11.05.21265810 (preprint, submitted)

Dietrich, M., **Seidlitz, S.**, Schreck, N., Wiesenfarth, M., **Godau, P.**, Tizabi, M., **Sellner, J.**, Marx, S., Knödler, S., Allers, M.M., Ayala, L., Schmidt, K., Brenner, T., Studier-Fischer, A., Nickel, F., Müller-Stich, B.P., Kopp-Schneider, A., Weigand, M.A., Maier-Hein, L. (2021). Machine learning-based analysis of hyperspectral images for automated sepsis diagnosis. arXiv (preprint)

Studier-Fischer, A., **Seidlitz, S., Sellner, J.**, Wiesenfarth, M., Ayala, L., Özdemir, B., Odenthal, J., Knödler, S., Kowalewski, K.F., Haney, C.M., Camplisson, I., Dietrich, M., Schmidt, K., Salg, G.A., Kenngott, H.G., Adler, T.J., Schreck, N., Kopp-Schneider, A., Maier-Hein, K., Maier-Hein, L., Müller-Stich, B.P., Nickel, F. (2021). Spectral organ fingerprints for intraoperative tissue classification with hyperspectral imaging. bioRxiv (preprint, submitted)

Rix, T., Hübner, M., Dreher, K.K., Nölke, J., Ayala, L., Schellenberg, M., Sellner, J., Seidlitz, S., Studier-Fischer, A., Müller-Stich, B., Nickel, F., Seitel, A., Maier-Hein, L. (2021). Deep learning for spectral image synthesis. SPIE Photonics West 2022: Multimodal Biomedical Imaging XVII (submitted)

Ayala, L., Wirkert, S., Vemuri, A., Adler, T., **Seidlitz, S.**, Pirmann, S., Engels, C., Teber, D., Maier-Hein, L. (2021). Video-rate multispectral imaging in laparoscopic surgery: First-in-human application. arXiv (preprint)

Gröhl, J., Dreher, K., **Schellenberg, M.**, Rix, T., Holzwarth, N., Vieten, P., Menjivar, L.A., Bohndiek, S., Seitel, A. and Maier-Hein, L. (2021). SIMPA: An Open-source Toolkit for Simulation and Processing for Photonics and Acoustics. (submitted)

Nickel, F., Studier-Fischer, A., Oezdemir, B., Odenthal, J., Mueller, L., Knoedler, S., Kowalewski, K., Camplisson, I., Allers, M.M., Dietrich, M., Schmidt, K., Salg, G.A., Kenngott, H.G., Billeter, A., Gockel, I., Sagiv, C., Etz-Hadar, O., Gildenblat, J., Ayala, L., **Seidlitz, S.**, Maier-Hein, L., Mueller-Stich, B.P. (2021). Optimization of anastomotic technique and gastric conduit perfusion with hyperspectral imaging in an experimental model for minimally invasive esophagectomy. bioRxiv (preprint, submitted)

Vieten, P., Dreher, K., Holzwarth, N., **Schellenberg, M.**, Nölke, J., Seitel, A., Gröhl, J., Rachel, Z.E., Siea, A.C., Held, T., Adeberg, S., Debus, J., Maier-Hein, L. (2021). Deep learning-based semantic segmentation of clinically relevant tissue structures leveraging multispectral photoacoustic images. SPIE Photonics West 2022: Photons Plus Ultrasound: Imaging and Sensing 2022 (accepted)

Holzwarth, N., **Schellenberg, M.**, Gnirs, R., Biegger, P., Korzowski, A., Franke, V.L., Dreher, K.K., Nölke, J., Seitel, A., Bachert, P., Schlemmer, H. and Maier-Hein, L. (2021). Tattoo image fusion: an optical pattern approach to image registration of freehand photoacoustics with other medical imaging modalities. SPIE Photonics West 2022: Photons Plus Ultrasound: Imaging and Sensing 2022 (accepted)

Nickel, F., Studier-Fischer, A., Özdemir, B., Odenthal, J., **Müller, L.R.,** Knödler, S., Kowalweski, K.F., Camplisson, I., Allers, M.M., Dietrich, M., Schmidt, K., Salg, G.A., Kenngott, H,G., Billeter, A.T., Gockel, I., Sagiv, C., Hadar, O.E., Gildenblat, J., Ayala, L., **Seidlitz, S.** Maier-Hein, L., Müller-Stich, B.P. (2021). Optimization of anastomotic technique and gastric conduit perfusion with hyperspectral imaging in an experimental model for minimally invasive esophagectomy. (submitted)

Roß, T., Bruno, P., Reinke, A., Wiesenfarth, M., Koeppel, L., Full, P.M., Pekdemir, B., **Godau, P.**, Trofimova, D., Isensee, F., Moccia, S., Calimeri, F., Müller-Stich, B.P., Kopp-Schneider, A., Maier-Hein, L. (2021). How can we learn (more) from challenges? A statistical approach to driving future algorithm development. arXiv (preprint)

Fallon, M., Schmidt-Kraepelin, M., Thiebes, S., **Warsinsky, S.**, Sunyaev, A. (2021). Social Comparison in mHealth for Physical Activity: The Role of Similar Others and Feelings of Envy. (submitted)

#### FURTHER NEWS FROM HIDSS4HEALTH

## » The School was Present at the Following Conferences or Events

- > 2021, Asian Conference on Computer Vision, poster
- 2021, International Symposium of Biomedical Imaging, poster
- > 2021, GIANA 2021 Challenge
- > Jan. 2021, 54th Hawaii International Conference on

System Sciences, oral presentation

- O6-11.03.2021, SPIE Photonics West, Melanie Schellenberg, oral presentation and poster
   Presentation at "Junge Talente Wissenschaft und Musik", DR: Paul Maria Scheikl
- O7.03.2021, BVM Bildverarbeitung in der Medizin, oral presentation
   20.05.2021, Oxford Nanopore Technologies London
   Workshop on Girl's day "Daten von heute-Medizin von morgen", DRs: Alexandra Walter, Pia Stammer, Katharina Löffler
- 20.05.2021, Oxford Nanopore Technologies Londor Calling 2021, oral presentation
- 05.06.2021, PTCOG 2021, oral presentation
- > 09.06.2021, Zeiss Symposium 2021, oral presentation
- 27.09.-01.10.2021, IROS 2021, oral presentation
- 27.09.-01.10.2021, MICCAI Medical Image Computing and Computer Assisted Intervention - MICCAI 2021, oral presentations & poster
- > 14.-19.11.2021, CompSysBio: Advanced Lecture Course on Computational Systems Biology (2021), Aussois, Frankreich, oral presentation

## » The Following Prizes were Awarded to the School's Members

- > 1st place in ISBI Cell Tracking Challenge, PI: Ralf Mikut, DR: Katharina Löffler
- KHYS Research Travel Grant in cooperation with Prof. Paolo Fiorini (Verona, Italy), PI: Franziska Mathis-Ullrich, DR: Paul Maria Scheikl

#### » Further Cooperations

- Cooperation with companies solidfluid & Tripleye, PI: Franziska Mathis-Ullrich, DR: Balazs Gyenes
- Start of an academic partnership between i4health (University College London) and HIDSS4Health, PI: Martin Frank, DR: Alexandra Walter
- > Involvement in the Covid-19-forecast-hub, PI: Melanie Schienle, DR: Daniel Wolffram
- > Press release: "More than the Sum of its Parts: Combining Models to Improve COVID-19 Forecasts", https://www.h-its.org/2021/09/29/covid-19-forecast-hub/
- CODAG Bericht Nr. 24: "Eine kollaborative Plattform zur korrigierten Schätzung der Sieben-Tage-Hospitalisierungsrate", https://www.covid19.statistik. uni-muenchen.de/pdfs/codag-bericht-24.pdf
- > RKI-Wochenbericht (Seite 12): "Bewertung des Trends der 7-Tage-Hospitalisierungsinzidenz unter Berücksichtigung verzögert berichteter Hospitalisierungen",https://www. rki.de/DE/Content/InfAZ/N/Neuartiges\_Coronavirus/ Situationsberichte/Wochenbericht/Wochen bericht\_2021-11-25.pdf?\_\_blob=publicationFile

#### » Science Communication Activities

- > Workshop on Girl's day "Wie kann man ohne schädliche Strahlung einen Blick unter die Körperoberfläche werfen?", DRs: Silvia Seidlitz, Melanie Schellenberg
- Contribution to the KIT Science week with an active museum, DRs: Tim Ortkamp, Pia Stammer, Alexandra Walter

Doctoral researchers from HIDSS4Health in total supervised 28 **Bachelor's and Master's theses** in 2021.

HIDSS4Health has a website (www.hidss4health.de), a Twitter account with 366 followers (jan 2022), and a mailing list.

#### 0 1 1 0 HEIBRIDS 0 0 0 0 HELMHOLTZ EINSTEIN INTERNATIONAL 0 1 1 0 HELMHOLTZ EINSTEIN INTERNATIONAL berlin RESearch school in data science

#### 5. HEIBRIDS - HELMHOLTZ EINSTEIN INTERNATIONAL RESEARCH SCHOOL ON DATA SCIENCE

HEIBRIDS brings together six Helmholtz centers and four university partners from the Einstein Center Digital Future (ECDF), which focuses on core digitization technologies, from digital health to digital industry and the digital humanities. The participating Helmholtz centers have first-class expertise in the fields of molecular medicine, astrophysics, polar and marine research, aerospace, materials science and geosciences. The goal of HEIBRIDS is to train a new generation of researchers who are both qualified Data Scientists and understand the requirements and challenges of those disciplines in which Information & Data Science has become indispensable.

#### **APPLICANT SITUATION AND RECRUITMENT**

There has been no recruitment in 2021, as all 25 HEIBRIDS PhD positions have been filled:

13 PhD researchers started in 2018 (one followed his supervisor to TU Dortmund in 2020 and his status has, since, changed to associated), two positions were awarded in 2019 (one additional doctoral researcher joined with supervisor's own funding) and 10 PhD researchers started in 2020.

Among the HEIBRiDS PhD researchers, there are 15 international and eight female researchers. In addition, there are one female and five male associated HEIBRiDS researchers.

#### **EVENTS AND NETWORKING**

#### » 6 Training and Networking Events

**HEIBRIDS Spring School,** Online, March 22-25 (26-30 participants):

The HEIBRIDS Spring School included lectures and interactive workshops on Scientific Writing, Research Data Management, Using Containers in Science, High-Performance Computing, and Data Cleaning. https:// www.heibrids.berlin/events-training/heibrids-workshops/ heibrids-spring-school-2021/

## **Workshop on Scientific Integrity,** Online, June 23 (8 participants):

A workshop on Scientific Integrity was held for the third cohort of HEIBRIDS PhD researchers. Through discussions and group exercises, we dealt with what distinguishes good behavior in science, how moral conflicts can be resolved, and what are the underlying values of good scientific behavior and misconduct. https://www. heibrids.berlin/events-training/heibrids-workshops/ workshop-on-scientific-integrity/

HIDA/HEIBRiDS Carpentries Workshop: **Programming with Python,** Online, August 31 – September 3 (25 participants): This workshop was open to participants from all Helmholtz Centers and partners. It covered data analysis and visualization in Python, focusing on working with core data structures, using conditionals and loops, writing custom functions, and running Python programs from the command line. https://www.heibrids.berlin/events-training/ heibrids-workshops/programming-with-python/

## Carpentries Workshop: **Programming with R,** Online, October 26-29 (15 participants):

This workshop was open to participants from all Helmholtz Centers and partners. It covered data analysis and visualisation in R focusing on working with core data structures, using conditionals and loops, writing custom functions, and running R programs from the command line\_https:// www.heibrids.berlin/events-training/heibrids-workshops/ programming-with-r/

## **HEIBRIDS Retreat,** Online/hybrid & BIMSB, November 3-4 (53 participants):

The first day of this two-day event included PICO presentations from all participating PhD researchers, followed by in-depth discussions with PIs in breakout rooms, and a keynote lecture. On the second day, a workshop on storytelling took place (see below), as well as a short presentation from HIDA regarding events and training offers.

#### Workshop on **Storytelling for Data Scientists,** BIMSB/MDC, November 4 (21 participants):

An interactive workshop that teaches scientists how to use storytelling to communicate to their peers and the public at large. https://www.lafabbricadellarealta.com/ case-studies/storytelling-data-scientists/

#### » 13 HEIBRIDS Lectures

- > 06.01.2021, Ingmar Nitze (AWI): "Machine-learning for mapping permafrost landscape dynamics"
- > 20.01.2021, Annika Bande (HZB): "Data Science and Robotics in Chemistry"
- > 03.02.2021, Dagmar Kainmüller (MDC): "Learning Cell Segmentation and Matching"
- > 17.02.2021, Begüm Demir (TU Berlin): "Deep Earth Query: Information Discovery from Big Earth Observation Data Archives"
- > 21.04.2021, Guillermo Gallego (TU Berlin / ECDF):
   "Space-Time Measurement of Ocean Waves Using Stereo Vision Systems"
- > 05.05.2021, Kashif Rasul & Julia Lasserre (Zalando): "Neural Time Series Forecasting & Algorithmic Size Advice at Zalando"
- > 19.05.2021, Wojciech Samek (Fraunhofer Institut): "Recent Advances in Explainable AI"
- O2.06.2021, Andrea Volkamer (Charité Universitätsmedizin): "Data-Driven Methods for Compound Design and Risk Assessment"
- > 30.06.2021, Stephan Rasp (Climate AI & TU Munich):
   "What Makes a Successful AI Project? 5 Questions to Ask Before Training Your Neural Network"
- > 14.07.2021, Heiko Mueller (NYU): "openclean An Open-Source Data Cleaning Library for Python"
- O3.11.2021, Frank Leymann (Uni Stuttgart): "Basics of Quantum Computing and Potentials in Machine Learning"
- 17.11.2021, Florian Tschorsch (TU Berlin): "Next Generation Data Networks: BitTorrent Meets Content Addressing"
- > 15.12.2021, Daniela Nicklas (Uni Bamberg): "IoT and the City: Smart Data Processing for Smart Applications"

#### » 2 HIDA Lectures @ HEIBRiDS

- > 16.06.2021, Alan Akbik (HU Berlin): "TARS: Few-Shot Learning for Natural Language Processing"
- > 20.10.2021, Dennis Shasha (NYU): "SafePredict and Friends"

#### » 15 PhD Researcher Seminars

### PUBLICATIONS

Doctoral researchers from HEIBRiDS are highlighted.

#### » 14 First-author Publications

**Weber, L.,** Sänger, M., Münchmeyer, J., Habibi, M., Leser, U., and Akbik, A. (2021). HunFlair: An easy-to-use tool for state-of-the-art biomedical named entity recognition. Bioinformatics, btab042. https://doi.org/10.1093/bioinformatics/btab042.

**Redyuk, S.,** Kaoudi, Z., Markl, V., and Schelter, S. (2021). Automating data quality validation for dynamic data ingestion. Proceedings of the International Conference on Extending Database Technology, March 23-26. ISBN 978-3-89318-084-4 on OpenProceedings.org.

**Agarwal, S.,** Tosi, N., Kessel, P., Padovan, S., Breuer, D., and Montavon, G. (2021). Towards constraining Mars' thermal evolution using Machine Learning. Earth and Space Science. https://doi.org/10.1029/2020EA001484.

**Pfalz, G.,** Diekmann, B., Freytag, J.-Ch., and Biskaborn, B.K. (2021). Harmonizing heterogeneous multi-proxy data from lake systems. Computers & Geosciences. https://doi.org/10.1016/j.cageo.2021.104791.

**Münchmeyer, J.,** Bindi, D., Leser, U., and Tilmann, F. (2021). Earthquake magnitude and location estimation from real time seismic waveforms with a Transformer Network. Geophysical Journal International, 226(2), 1086-1104. https://doi.org/10.1093/gji/ggab139.

**Ghosh, B.,** Haghshenas Haghighi, M., Motagh, M., and Maghsudi, S. (2021). Using Generative Adversarial Networks for extraction of InSAR signals from large-scale Sentinel-1 Interferograms by improving tropospheric noise correction. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-3-2021, 57-64.

https://doi.org/10.5194/isprs-annals-V-3-2021-57-2021.

**Rettelbach, T.,** Langer, M., Nitze, I., Jones. B., Helm, V., Freytag, J-Ch., and Grosse, G. (2021). A Quantitative Graph-Based Approach to Monitoring Ice-Wedge Trough Dynamics in Polygonal Permafrost Landscapes. Remote Sens. 2021, 13, 3098. https://doi.org/10.3390/rs13163098. Ghosh, B., Motagh, M., Haghshenas Haghighi, M., Stefanova Biskaborn, B.K., Nazarova, L., Kröger, T., Pestryakova, L.A., Vassileva, M., Walter T., and Maghsudi, S. (2021). Automatic detection of volcanic unrest using blind source separation with a minimum spanning tree based stability analysis. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. doi.org/10.1109/JSTARS.2021.3097895.

Tillmann, P., Bläsi, B., Burger, S., Hammerschmidt, M., Höhn, 0., Becker, C. and Jäger, K. (2021). Optimizing metal grating back reflectors for III-V-on-silicon multijunction solar cells. Opt. Express, Vol.29, p. 22517.\_doi: 10.1364/0E.426761.

Weber, L., Sänger, M., Garda, S., Barth, F., Alt, Ch., and Leser, U. (2021). Humboldt @ DrugProt: Chemical-Protein Relation Extraction with Pretrained Transformers and Entity Descriptions. Proceedings of the 7th BioCreative Challenge Evaluation Workshop.

Agarwal, S., Tosi, N., Kessel, P., Breuer, D., and Montavon, G. (2021). Deep learning for surrogate modeling of two-dimensional mantle convection. Physical Review Fluids, 6, 113801. https://doi.org/10.1103/PhysRevFluids.6.113801.

Weber, L., Garda, S., Münchmeyer, J., and Leser, U. (2021). Extend, don't rebuild: Phrasing conditional graph modification as autoregressive sequence labelling. Proceedings of the 2021 Conference on Empirical Methods in Natural Vyse S.A., U. Herzschuh, G. Pfalz, L.A. Pestryakova, B. Language Processing, 1213-1224.

Styp-Rekowski, K., Stolle, C., Michaelis, I., and Kao, O. Calibration of the GRACE-FO Satellite Platform Magnetometers and Co-Estimation of Intrinsic Time Shift in Data. To appear in proceedings of the IEEE International Conference on Big Data 2021.

Münchmeyer, J.\*, Woollam, J.\*, Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J., and Soto, H. (2021). Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. Journal of Geophysical Research: Solid Earth. (accepted) \*Equal contribution

#### » 5 Co-author Publications

Rumberger, J.L., Yu, X., Hirsch, P., Dohmen, M., Guarino, V.E., Mokarian, A., Mais, L., Funke, J., and Kainmueller, D. (2021). How Shift Equivariance Impacts Metric Learning for Instance Segmentation. IEEE/CVF International Conference on Computer Vision (ICCV). (accepted)

Syrykh, L., Pfalz, G., Herzschuh, U., and Diekmann, B. (2021). Late Quaternary Climate Reconstruction and Lead-Lag Relationships of Biotic and Sediment-Geochemical Indicators at Lake Bolshoe Toko, Siberia. Front. Earth Sci., 9, 703. https://doi.org/10.3389/feart.2021.737353.

Rautenstrauch, P., Vlot, A.H.C., Saran, S. and Ohler, U. (2021). Intricacies of single-cell multi-omics data integration. Trends in Genetics. https://doi.org/10.1016/j.tig.2021.08.012.

Foster, W. J., Ayzel, G., Münchmeyer, J., Rettelbach, T., Kitzmann, N., Isson, T. T., Mutti, M., and Aberhan, M. Machine learning identifies ecological selectivity patterns across the end-Permian mass extinction. Paleobiology. (accepted)

Baunsgaard, S., Boehm, M., Chaudhary, A., Derakhshan, B., Geißelsöder, S., Grulich, P. M., Hildebrand, M., Innerebner, K., Markl, V., Neubauer, C., Osterburg, S., Ovcharenko, O., Redyuk, S., Rieger, T., Mahdiraji, A. R., Wrede, S. B., Zeuch, S. (2021). ExDRa: Exploratory Data Science on Federated Raw Data. Proceedings of the 2021 International Conference on Management of Data, 2450-2463.

#### » 6 Pre-prints / Publications under Review

Diekmann, N. Nowaczyk, and B.K. Biskaborn (2021). Sediment and carbon accumulation in a glacial lake in Chukotka (Arctic Siberia) during the late Pleistocene and Holocene: Combining hydroacoustic profiling and down-core analyses. Biogeosciences Discuss. https://doi.org/10.5194/bg-2021-39.

Jaurigue, L., Robertson, E., Wolters, J., and Lüdge, K. (2021). Reservoir Computing with Delayed Input for Fast and Easy Optimisation. Entropy, 23, 1560. https://doi.org/10.3390/e23121560.

Pfalz, G., Diekmann, B., Freytag, J.-C., Sryrkh, L., Subetto, D.A., and Biskaborn, B.K. (2021). Improving age-depth correlations by using the LANDO model ensemble. Geochronology Discuss. https://doi.org/10.5194/gchron-2021-40.

Abolt, C. J., Atchley, A. L., Harp, D. R., Liljedahl, A. K., Jorgenson, T. M., Witharana, C., Bolton, W. R., Schwenk, J., Rettelbach, T., Grosse, G., Boike, J., Nitze, I., Rumpca, C. T., Wilson, C. J., Bennet, K. E. Topography drives variability in circumpolar permafrost thaw pond expansion. Nature. (under review)

Woollam, J.\*, **Münchmeyer, J.\*,** Tilmann, F., Rietbrock, A., > Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J. & Soto, H. (2021). SeisBench - A Toolbox for Machine Learning in Seismology. Arxiv preprint. https://arxiv.org/abs/2111.00786 \*Equal contribution

Kirschbaum, T., Petit, T., Dzubiella, J., and Bande, A. Effects of Oxidative Adsorbates and Cluster Formation on the Electronic Structure of Nanodiamonds. The Journal of Computational Chemistry. (under review)

#### FURTHER NEWS FROM HEIBRIDS

- > HEIBRIDS doctoral researchers have participated in 29 national and international conferences, with poster or oral presentations.
- > AWI press release about the first prize winners, Tabea Rettelbach and Alexandra Zuhr, at the prestigious European "Copernicus Masters" competition.
- > HEIBRIDS doctoral researcher Leon Weber and associated doctoral researcher Mario Sänger participated in the MEDIQA Shared Task (Task 1) and their contribution achieved a superb second position among 23 participating teams.
- > Binayak Ghosh and his colleague Shagun Garg from the GFZ (team shagun1511), ranked in the third place at the NASA ETCI 2021 Competition on Flood Detection from Sentinel-1 data. The event, organized by NASA IMPACT in collaboration with the IEEE GRSS, seeks to develop approaches to delineate open water flood areas as an effort to identify flood extent, an impactful disaster that occurs frequently throughout the world.
- > HEIBRIDS PI Guillermo Gallego is among the five awardees in the category "Outside Associate Editors" that are selected annually by the Institute of Electrical and Electronics Engineers (IEEE). The work of Binayak Ghosh on automatic detection of volcanic unrest has been covered by various news outlets: https://scienmag.com/volcano-analysis-in-real-time/ https://www.maz-online.de/Brandenburg/Wie-Potsdamer-Geoforscher-die-Beobachtung-von-Vulkanenverbessern (restricted access) https://www.innovations-report.de/fachgebiete/

geowissenschaften/vulkananalyse-in-echtzeit/

- HEIBRIDS doctoral researcher Tabea Rettelbach, Christian Utama, and Brian Groenke attended the first H^3, Helmholtz Herbst Hackathon, which brought together for four days MSc students, doctoral researchers, and postdocs from Helmholtz and elsewhere to apply their skills in solving real-world scientific challenges with AI/ML. The event featured several keynote talks by prominent researchers in the field of scientific machine learning. Team HEIBRiDS participated in a challenge involving image-based analysis of simulated plant root systems, and in cooperation with another team, was able to achieve excellent results on the given task.
- > Under the lead of the Alfred Wegener Institute (AWI), and in scientific collaboration with the German Aerospace Center (DLR) and the University of Fairbanks (UAF), a crew of five researchers and logisticians (including HEIBRIDS PhD researcher Tabea Rettelbach and her supervisor Prof. Dr. Guido Grosse) from AWI headed to Kotzebue, AK, USA in June and July 2021 to conduct an aerial imaging campaign with AWI's Polar-6 aircraft. The goal of the expedition was to acquire very high-resolution imagery with the Modular Aerial Camera System (MACS; designed by DLR) and high-resolution digital terrain models with laser scanners, to map and monitor (changing) permafrost landscapes in West Alaska. With such high-resolution data, it is possible to monitor permafrost related activities such as coastal erosion, changes in lakes, thaw slumps, or ice-wedge thaw, all of which are especially vulnerable in the face of a changing climate. Overall, more than half a million images were captured and ca. 3.600 km<sup>2</sup> mapped.
- > HEIBRIDS doctoral researcher Peter Tilmann was funded by the HIDA Trainee Network to carry out a three-month stay at the DLR Institute of Solar Research in Almeria, Spain. During his stay, he conducted research in the group of Stefan Wilbert on the classification of optical cloud transparency from sky images.



HEIBRiDS doctoral researche Tabea Rettelbach (in front) with her colleagues in Alaska (Photo: T. Rettelbach).

# MARDATA HELMHOLTZ SCHOOL FOR MARINE DATA SCIENCE

#### **6. HELMHOLTZ SCHOOL FOR MARINE DATA SCIENCE (MARDATA)**

The Helmholtz School for Marine Data Science (MarDATA) pools scientific marine expertise and offers young scientists > 54 supervisors and PIs the unique opportunity to refine Information & Data Science methods specifically for the marine research. MarDATA has the goal of defining a new type of "Marine Data Scientists" and training them in a structured doctoral program. Scientists **EVENTS AND NETWORKING** from computer science, information technology and mathematics work together on marine topics. This includes modelling on supercomputers, (bio-) computer science and robotics  $\rightarrow$  An Introduction to GIS or statistics and big data methods.

Helmholtz Centres: GEOMAR + AWI

**Partner Universities:** Kiel University, University of Bremen, > An Introduction to Marine Geophysics Jacobs University Bremen

#### APPLICANT SITUATION AND RECRUITMENT

In 2021, MarDATA recruited the second cohort of doctoral researchers. Recruitment was done in two separate rounds > at GEOMAR and AWI. At both cases, recruitment started with an internal call for marine Information & Data Science project ideas among scientists at the Helmholtz Centers and partner universities. To increase the diversity of PIs/supervisors at GEOMAR, MarDATA issued a "Young Female Investigator Call" (YFIC) that primarily targeted young female PostDocs as PIs. Here, MarDATA was able to present an adjusted co-financing scheme that minimizes the necessary co-funding by PIs to significantly enlarge the group of potential MarDATA PIs. In 2021, MarDATA advertised eight positions at AWI and 11 at GEOMAR. Thereby, all 20 positions available at GEOMAR in the initial funding period have been advertised. At AWI, five additional positions remain open and will be advertised in 2022. The internal project call for these five positions at has already been completed. Applications (summarized across all 2021 projects) have multiplied compared to the first round of recruitment in 2019, with a total of 450 applications for 19 projects in 2021 (only 80 in 2019, 20 projects). Projects at  $\rightarrow$ AWI started between June and October 2021, while projects at GEOMAR started between October 21 and January 2022.

#### **Current Status:**

- > 33 Doctoral researchers (13 female/20 male) from 12 countries (German 17/international 16)
- > 1st cohort: 15, 2nd cohort: 18 (plus1 project still in recruitment process)
- GEOMAR: 19, AWI: 14 >

#### » 9 MarDATA Online Lecture Series:

- > Ice Sheet Dynamics
- > My Data, My Publication, My Rights
- > An Introduction to Reinforcements Learning
- > Marine Data Visualization: Ocean Data View
- An Introduction to Marine Geology
- > An Introduction to Marine Chemistry
- > An Introduction to Physical Oceanography

#### » 3 MarDATA Workshops:

- Practical Innovation Creative Thinking
- Reproducible Computational Science
- Machine Learning for Marine Sciences (cooperation with Helmholtz AI at DKRZ)

Additional training was provided by the cooperating graduate programs ISOS (Kiel) & POLMAR (AWI).

#### » 3 Special Events/ Networking Events:

- Women in Data Science | Seminar Series: 1. "AI-enabled Medical Imaging" (invited speaker: Prof. Dr. Julia Schnabel, Helmholtz Munich/TUM)
- MarDATA Weekly (weekly video meeting with the MarDATA doctoral researcher community)
- PI/Supervisor onboarding (two meetings: Dec 2021 and Jan 2022)
- Individual onboarding meetings (remote) with all newly recruited MarDATA doctoral researchers

#### » 4 Events with Participation of MarDATA:

- Digital Science Monday @ GEOMAR (re-occurring lecture/ seminar series)
- HIDA Data Science Career Day
- HIDA Annual Conference
- > H3 Helmholtz Herbst Hackathon

#### PUBLICATIONS

MarDATA doctoral researchers are **highlighted**. MarDATA currently does not have any associated members

#### » 4 Peer-reviewed Publications

Verwega, M.T., Trahms, C., Antia, A.N., Dickhaus, T., Prigge, E., Prinzler, M.H.U., Renz, M., Schartau, M., Slawig, T., Somes, C.J., and Biastoch, A. (2021). Perspectives on Marine Data Science as a Blueprint for Emerging Data Science Disciplines. Front Mar Sci, 8:678404.

Verwega, M.T., Somes, C.J., Schartau, M., Tuerena, R. E., Lorrain, A., Oschlies, A., and Slawig, T. (2021). Description of a global marine particulate organic carbon-13 isotope data set. Earth Syst. Sci. Data, 13, pp. 4861-4880.

Verwega, M.T., Somes, C. J., Tuerena, R. E., Lorrain, A. (2021). A global marine particulate organic carbon-13 isotope data product. PANGAEA: https://doi.org/10.1594/ PANGAEA.929931.

von See, T.B., Meurer, T., and Greinert, J. (2021). Marine boundary layer tracking using an AUV with UKF based extremum seeking. IFAC-PapersOnLine, 54 (16), pp. 320-326.

#### » 2 Conference Proceedings

Trahms, C., Handmann, P., Rath, W., Visbeck, M., and Renz, M. (2021). Where have all the larvae gone? Towards Fast Main Pathway Identification from Geospatial Trajectories. In: 17th International Symposium on Spatial and Temporal Databases (SSTD '21) 23.-25.08.2021, pp. 126-129.

Hiremath, D., Martin, C., Hasselbring, W. and Rath, W. (2021). Towards Automated Metamorphic Test Identification for Ocean System Models. In: 2021 IEEE/ACM 6th International Workshop on Metamorphic Testing (MET). pp. 42-46.

#### » 8 Conference Contributions (talks, posters and session chairs)

Trahms, C., Renz, M., and Visbeck, M. 5th Data Science Symposium, Geomar (22. Jan 2021). A Whole New World: Leveraging the Power of Data Links with Heterogeneous Information Networks (talk).

Trahms, C., Handmann, P., Rath, W., Visbeck, M. and Renz, M. EGU General Assembly 2021 (19. - 30. April 2021). Efficient Pathway Identification from Geospatial Trajectories (talk).

Kinzel, L., Schlindwein, V., Fromm, T. and Maaß, P. 37th General Assembly of the European Seismological Commission (19. - 24. Sept 2021). P-wave arrival picking with deep learning and uncertainty quantification (talk).

Kinzel, L., Schlindwein, V., Fromm, T. and Maaß, P. 37th General Assembly of the European Seismological Commission (19. - 24. Sept 2021). Unsupervised Deep Representation Learning for Icequake Detection at Neumayer Station, Antarctica (talk).

Trahms, C. & Verwega, M.T. ICYMARE2021 (21. - 24. Sept 2021) Session: Marine Data Science (session chairs).

von See, T.B., Meurer, T. and Greinert, J. 13th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles (CAMS) (22. - 24. Sept 2021). Marine boundary layer tracking using an AUV with UKF based extremum seeking (talk).

Khachaturyan, M. ISE Symposium 2021 (29. Sept - 01. Oct 2021). Conserved mitochondrial chromosome stoichiometry in a circumglobally distributed marine plant (talk).

Mohn, H., Kreyling, D., Wohltmann, I., Maass, P., Rex, M., Quadrennial Ozone Symposium 2021 (03. - 09. Oct 2021). AI For Fast Stratospheric Ozone Predictions (poster).

#### FURTHER NEWS FROM MARDATA

- Part of the MarDATA mission is to define and establish the scientific profile of marine Information & Data Science. Accordingly, members of MarDATA regularly discuss tasks, challenges and, opportunities for marine data scientists. In a workshop in 2020, members discussed the potential of marine Information & Data Science to become a blueprint for emerging Information & Data Science disciplines. In 2021, the results of this workshop were published in Frontiers of Marine Sciences (Verwega & Trahms et al, 2021).
- Yvonne Jenniges received a Scientific Telegram Award for her video presentation at the HIDA Annual Conference (30.11.2021).
- In 2021, MarDATA started a new seminar series: Women in Data Science. The seminar series introduces outstanding female (marine) data scientists and their research.

In addition to a scientific talk, seminars also feature roundtable discussions that address challenges in a still often male-dominated Information & Data Science community. First speaker of the series was Prof. Julia Schnabel (Institute of Machine Learning in Biomedical Imaging, Helmholtz Center Munich). Since December 2021 on, the seminar series has been organized in cooperation with the Women Executive Board at GEOMAR. Upcoming speakers include MarDATA Advisory Board member Prof. Robin Teigland (Professor of Strategy and Management of Digitalization in the Entrepreneurship and Strategy Division at Chalmers).

- A MarDATA Block Course originally planned to happen in person in Nov 2021 had to be cancelled (again). The block course was supposed to feature "Intro to Marine Science" lectures for the newly starting 2nd cohort of doctoral researchers, and a "MarDATA Data Science Sprint" format for the 1st cohort doctoral researchers.
- Communication/Marketing: MarDATA has a website http://www.mardata.de, a Twitter account with 310 followers (Jan 2022), and a news mailing list.

# V. OUTLOOK

n 2022, HIDA will focus its activities in view of the upcoming evaluations. HIDA will prepare and support the first schools (HEIBRIDS and HIDDS4Health) in this process commencing in autumn 2022. The schools anticipate their first doctoral researchers to complete their PhDs over the course of the year.

In 2022, HIDA will expand its activities and successes from 2021, as well as launch further new initiatives to provide additional training and networking opportunities in the field of Information & Data Science for all Helmholtz Centers and research fields. For 2022, we expect to hold more face-toface events so that doctoral researchers can broaden their personal networks. In this context, we also plan to further expand the HIDA Trainee Network. Another goal will be to intensify and increase cooperation between researchers and international partners, in order to create spaces for knowledge exchange and training, and to actively support the Helmholtz Centers in recruiting new Information & Data Science talent. HIDA aims to maximize the perception of the Helmholtz Association as a key player in the field of Information & Data Science in cooperation with the HIDSS and the Incubator Platforms and in the light of the upcoming evaluations in 2022/23.

HIDA plans to continue its networking events (e.g. Virtual Career Day Vol. III; Virtual Datathon Challenge Vol. IV, HIDA Annual Conference Vol. II) and to take active part in and design recruiting events for the entire Helmholtz Association. Substantial marketing and communication initiatives complement these activities.

# VI. RESPONSIBILITY FOR REPORT / HIDA MAIN CONTACT

The members of the Steering Committee of the Helmholtz Information & Data Science Academy (HIDA-Steer) checked and approved this report in February 2022.

#### **HIDA Main Contacts:**

Susan Trinitz Adviser, Strategic Initiatives/HIDA Constanze Fröhlich Press Officer, HIDA

Helmholtz Information & Data Science Academy Friedrichstraße 171 10117 Berlin



#### Imprint

Published by:

Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren e.V. Head Office and Bonn Office Ahrstrasse 45, 53175 Bonn Phone: +49 (0) 228 30818-0 Email: info(at)helmholtz.de

Strategic Initiatives Helmholtz Association Helmholtz Information & Data Science Academy (HIDA) Friedrichstraße 171 D-10117 Berlin Phone: +49 30 5498226-0 Email: hida(at)helmholtz.de

V.i.S.d.P. (Person responsible according to the German Press Law): Franziska Broer Project Managers: Constanze Fröhlich, Susan Trinitz Editors: Constanze Fröhlich, Susan Trinitz

Design: Kai Kullen

Printing: Druck- und Verlagshaus Zarbock GmbH & Co. KG Frankfurt am Main

Correct as of March 2022

The illustration is based on data calculating ozone rates of change in the atmosphere.

HELGE MOHN from MarDATA determines changes in ozone concentration with low computational effort and high accuracy by means of neural networks. His goal is to enable climate models to make better predictions on this basis.

This image is an example of an artistic implementation of research data used by the doctoral researchers of the six Helmholtz Data Science Schools in the fields of Energy, Earth and Environment, Health, Key Technologies, Structure of Matter, and Aeronautics, Space and Transport.



