

gesundhyte.de

THE MAGAZINE FOR DIGITAL HEALTH IN GERMANY

ISSUE 14 JUNE 2022

focus: research data saves lives

from page 14

personalized treatments vs. transparent patients?

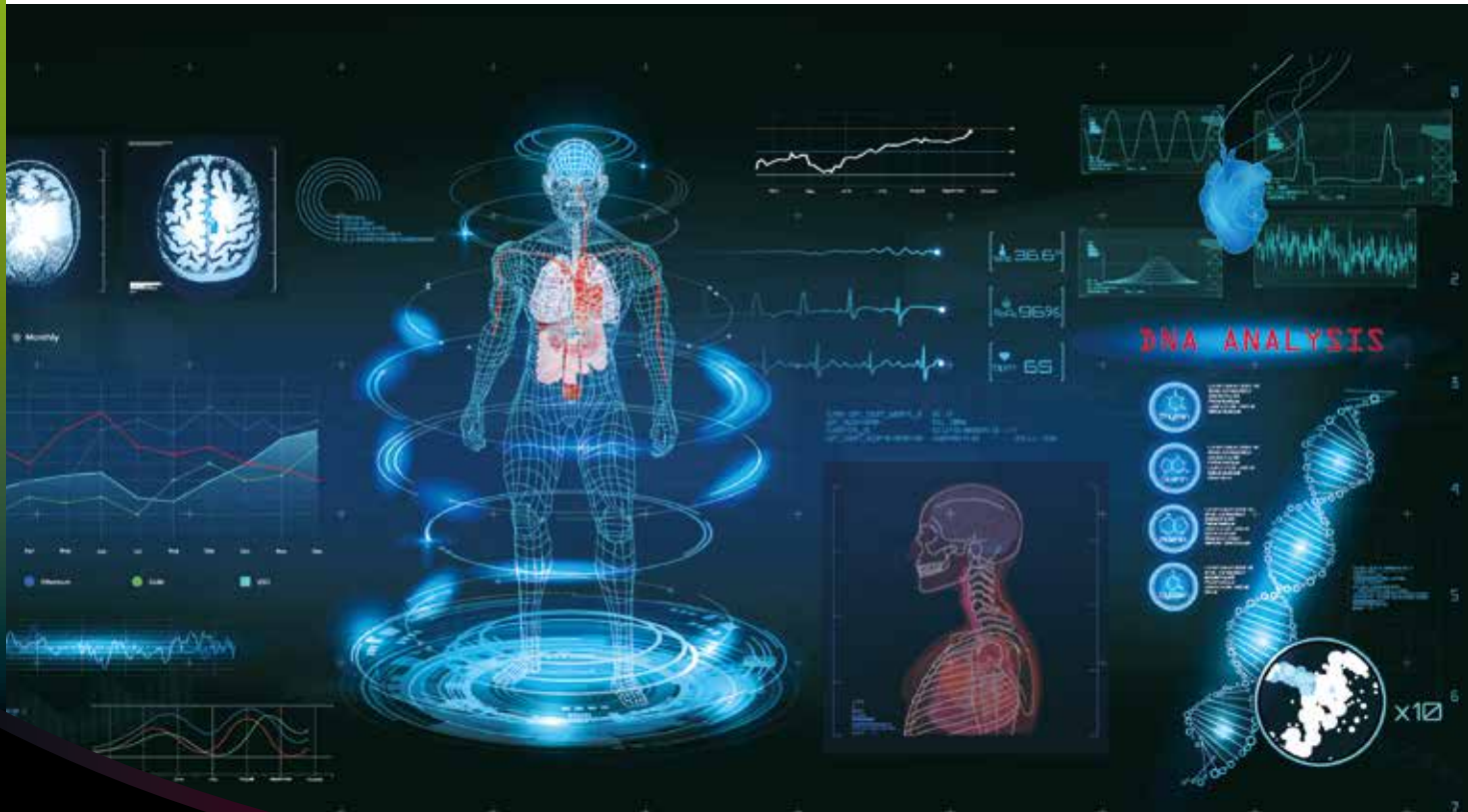
page 32

SmlCS – a smart solution for the automated detection of hospital outbreaks

page 46

interviews with Martin Hager Dana Westphal and Michael Seifert Simon Haas

page 29, 73 and 82



gesundhyte.de

is dedicated to finding solutions for today's most important challenges in health care focusing on Digital Health and Systems Medicine, two young disciplines that are considered to form the medicine of the future. The key is the combination of laboratory research data of different kinds with real-world data – from bench to bed-site. Innovative technologies and methods will pave the way for more precise predictions and personalised therapy. Already today, this approach is successfully used in Oncology and will be extended to other diseases in the future. Here, the establishment of robust and standardized IT infrastructure plays a major role to allow for the secure exchange of patient data with research teams. Read the magazine gesundhyte.de to find out how this innovative branch of science works to provide solutions for our current and future challenges in medicine.



greetings

Dear Reader,



With more than two years of COVID-19 pandemic under our belt, we are looking back on an intensely challenging period in which research has achieved major breakthroughs. The majority of the population has become keenly aware of the need for strong and effective health research and medical care. Rapid access to reliable data and its exchange between research and patient care have been decisive factors for success. Extensive data sets related to the new virus had to be gathered, shared, and analysed in the shortest possible time.

For this task, the scientific community was able to take advantage of research infrastructures that have been actively supported by the Federal Ministry of Education and Research (BMBF) in recent years. For example, the bioinformatics tools and the data cloud of the German Network for Bioinformatics Infrastructure (de.NBI) were used by researchers across Germany to better understand the virus and to study its mechanisms.

BMBF has also established structures for connecting medical data. The core element of the Medical Informatics Initiative, which has been BMBF-funded since 2018, are data integration centres to facilitate analyses across different university hospitals. To support and strengthen the COVID-19 research effort, BMBF further intensified the networking of university hospitals by funding the National Network of University Medicine (NUM). The COVID-19 Data Exchange Platform (CODEX), which makes research data related to COVID-19 available to all university hospitals in Germany, has evolved in a central element of this initiative. The extensive prior developments of the Medical Informatics Initiative, in particular the data integration centres, provided the foundation for this purpose.

This edition of gesundhyte.de provides insight into diverse projects funded under this and other measures. The results highlight the great potential of research data and its analysis in the life sciences and medicine. I wish you all inspiring reading.

A handwritten signature in blue ink, appearing to read 'V. von M.', with a stylized flourish at the end.

Prof. Dr. Veronika von Messling

Director-General for Life Sciences
Federal Ministry of Education and Research



greetings

Dear Reader,

Another year of the pandemic has shown us how vital it is to make healthcare data available to the research community. With the integration of the Berlin Institute of Health (BIH) into Charité, we have taken a major step towards our goal of strengthening translational research. On January 1, 2021, under the motto “We belong together!” the BIH became the translational research unit and has since formed – alongside the university hospital and the medical faculty – Charité’s third pillar.

The collaboration between Charité and BIH has already proven to be extremely valuable during the pandemic. Thanks to the close connection between the BIH and our hospital, we were able to respond quickly and incorporate patient care data directly into research. In recent months, research on scientific and medical solutions to combat the pandemic has been running at full speed, also at Charité, where we contribute to the fight against COVID-19 in many collaborations. The pandemic situation is a prime example of the importance of closely integrating routine clinical and research activities. As a result of the crisis, a strong and collaborative network, the German Network of University Medicine (NUM), has been established under the coordination of Charité. We are confident that this alliance of all German university hospitals will continue to produce success in the future.

In the longer term, the new translational research unit will particularly benefit our patients. For this reason, I look forward to working with the BIH to further accelerate the translation of research findings into clinical applications for our patients. The synergies that are now being formed between Charité and the BIH must be used in a constructive and targeted manner. That is why the integration of the BIH into Charité is so crucially important.

This issue of gesundhyte.de provides numerous examples of promising translational research. I hope you enjoy reading the individual articles.

Prof. Dr. Heyo K. Kroemer

Chief Executive Officer, Charité – Universitätsmedizin Berlin

foreword

Dear Reader,



“The ball is round and a game lasts 90 minutes.” this famous wisdom quote from Sepp Herberger, a German soccer player and coach, describes a fact. Facts are based on data and describe an indisputable truth. However, facts can also describe a generally accepted circumstance. What could be true or false becomes a statement of fact.

The infamous statement by former U. S. president Donald Trump’s chief advisor Kellyanne Conway, who defended Trump’s then spokesperson Sean Spicer against accusations of lying by commenting that he had not lied but offered alternative facts, was the sad birth of an era in which factual assertions are readily put forward without data to back them up. As we all know, such factual claims spread on social media much faster than data-based conclusions, which are often perceived as dry or boring. *“If 50 million people say something foolish, it is still foolish,”* said the famous French writer Anatole France a century ago, and this proclamation has taken on a whole new meaning in the age of social media.

Ascertaining facts by collecting data is the core business of science. Not only there, but also in the field of health management, data provides the basis for prognoses, assessments, and diagnoses. Especially at a time when data-free assertions of fact are readily published and replicated millions of times, it is all the more important to remind people that knowledge and facts must be based on data. However, a disturbing perception among the public is that factual claims can change – even in the field of science. What was considered *“true”* yesterday may appear *“false”* tomorrow. Science has always faced errors and subsequent revisions. There are numerous examples of how a shifting basis of data has generated new findings, and not just in the relatively short history of the current pandemic.

Data reflect a snapshot of the description of reality, and the data landscape is constantly and dynamically evolving. The foundation of data-centric science includes platforms accompanied by frameworks that allow the scientific community broad access to and the best use of data. This issue of gesundhyte.de presents fascinating insights into this. The portrait of Claudia Langenberg shows that scientists sometimes also take an emotional approach to data. *“We love data,”* is an apt characterization of her field of work in just three words – and at the same time, also an excellent summary of this issue.

I hope you enjoy reading the remaining 40,962 words and are able to gain numerous insights.

We look forward to your continued loyalty – both to us and to data!

Yours,

Roland Eils

Editor-in-Chief, gesundhyte.de

index

greetings	3	
Prof. Dr. Veronika von Messling, Director-General for Life Sciences, Federal Ministry of Education and Research		
greetings	4	
Prof. Dr. Heyo K. Kroemer, Chief Executive Officer, Charité – Universitätsmedizin Berlin		
foreword	5	
Prof. Dr. Roland Eils, Editor-in-Chief, Founding Director of the BIH Center for Digital Health at Charité		
“life is unpredictable!”	8	
Portrait: Barbara Di Ventura by Katharina Kalhoff and Gesa Terstiege		
my experience of the pandemic	11	
2 years of COVID-19 from the perspective of a young scientist by Sören Lukassen		
focus: research data saves lives		
improving access to clinically relevant data	14	
BMBF funding measure “i:DSem” summing up positively by Björn Dreesen-Daun, Markus Löffler, Stephan Kiefer, Juliane Fluck, Mischa Ubachs and Maksims Fiosins		
use of routine clinical data for research and care	21	
Prerequisite: digital modular broad consent by Clemens Spitzenfeil, Alexandra Stein, Martin Bialke, Torsten Leddig and Wolfgang Hoffmann		
lost in translation? semantic standardization!	25	
Terminologies like LOINC and SNOMED CT reduce Babylonian Confusion of tongues by Cora Drenkhahn and Josef Ingener		
a learning healthcare system could already be a reality	29	
Interview and Company profile: Martin Hager, Roche Pharma AG		
personalized treatments vs. transparent patients?	32	
How patient advocacy groups assess opportunities and challenges of AI in medicine and care by Klemens Budde, Matthieu-P. Schapranow, Elsa Kirchner and Thomas Zahn – Plattform Lernende Systeme		
innovative software tools for the life sciences	37	
BMBF initiative CompLS provides €34 million in funding for the development of computational methods and analytical tools by Daniel Heinrichs, René Wolf-Eulenfeld, Bernhard Renard, Jakub Bartoszewicz, Melania Nowicka, Birte Kehr and Sebastian Niehus		
future-oriented analysis of research data	42	
Members of the de.NBI network use AI-based technologies by Vera Ortseifen, Peter Belmann and Alfred Pühler		
SmlCS	46	
A smart solution for the automated detection of hospital outbreaks by Antje Wulff, Claas Baier, Michael Marschollek, and Simone Scheithauer on behalf of all members and sites of the HiGHmed project's Infection Control use case and Infection Control Study Group		

national research data infrastructure NFDI Working together to build a FAIR research data management in Germany by Nathalie Hartl	50	
GAIA-X as an enabler for a health data space Launched in 2019, the Franco-German project GAIA-X is entering its third year. Here is a snapshot. by Harald Wagener	54	
FAIR data: experiences from collaborative research projects Requirements and step-by-step implementation for data-driven collaborations by Matthias Ganzinger and Matthieu-P. Schapranow	57	
news from the BMBF	62	
news from the BIH	66	
mTOMADY – a digital health wallet for people without insurance Company profile: mTOMADY by Stefanie Seltmann	70	
“interdisciplinarity is the only way to answer complex questions” Interview: Dana Westphal and Michael Seifert by Marco Leuer	73	 
INCOME – integrative, collaborative modeling in systems medicine Hackathons for a culture of data and model sharing by Nina Fischer, Wolfgang Müller, Dagmar Waltemath, Olaf Wolkenhauer and Jan Hasenauer	77	
e:Med junior research alliance LeukoSyStem Interview: Simon Haas by Silke Argo	82	
“we love data!” Portrait: Claudia Langenberg by Stefanie Seltmann	87	
news	90	
events	94	
imprint	101	
about us	102	
contact	103	

"life is unpredictable!"

A portrait of systems biologist Barbara Di Ventura

by Katharina Kalhoff and Gesa Terstiege

Optogenetics aims to precisely control cell activity using light. The discipline has made considerable progress since it first emerged around 20 years ago. Prof. Barbara Di Ventura is committed to further exploring the potential of optogenetics, particularly for the treatment of disease. Her efforts to advance the promising field are aided by her background in computer engineering and molecular biology and by her passion for her subject. She is convinced that one day optogenetics will considerably improve cancer treatment.

As a child her father thought she should become a lawyer because little Barbara would not stop talking until she had convinced every one of her opinion. "My mother worried that would bring me into contact with mafiosi and other corrupt individuals," explains the scientist Di Ventura from the University of Freiburg and laughs. "But what put me off was a glance at the Italian civil code – it looked so boring." So instead of law, she opted to study computer engineering at La Sapienza University in Rome. Today, she is a professor of Biological Signaling Studies at the University of Freiburg and hopes that her research will lead to improved cancer treatment in the future.

"My supervisor told me few months into the PhD that it would be way better for me to generate my own data rather than having to wait on other people's experiments. I took on the challenge – with no prior knowledge"

Di Ventura began to make the switch to biology when she embarked on a master thesis that focused on modeling gene networks. The young engineer was not originally interested in pursuing a PhD, but an opportunity to go abroad attracted her. The European Molecular Biology Laboratory (EMBL) in Germany was recruiting PhD students from across Europe. "I had never considered Germany as a place to settle down, but I quickly fell in love with Heidelberg," she says. "Life is unpredictable!" Her original plan was to develop computer models based on biological experiments. "My supervisor told me few months into the PhD that it would be way better for me to generate my own data rather than having to wait on other people's experiments," she says. "I took on the challenge – with no prior knowledge. Fortunately, my supervisor at that time, Luis Serrano, was very supportive."

"I always had inspiring mentors and supporters who have pushed me forward and always believed in my scientific vision – science is all about teamwork."

Teamwork is decisive

An important step in the right direction was the establishment of her own research group. This was achieved thanks to the German Federal Ministry of Education and Research, which provided almost €1.5 million in funding for an e:Bio junior research group. "That funding was very important in launching my career. Without the junior research group, I probably would not have gotten the professorship in Freiburg. It enabled me to demonstrate that I could work independently and make progress on relevant topics," says Di Ventura. "I always had inspiring



From Rome to southern Germany: Barbara Di Ventura is Professor for Biological Signalling Studies at the University of Freiburg (Photo: Harald Neumann).

mentors and supporters who have pushed me forward and always believed in my scientific vision – science is all about teamwork.” Her field depends on people who can work together and approach various biological questions from different angles. Some members of her team develop optogenetic tools, others model the processes, and others employ the tools to gain important answers.

“The development of a small, single-component, easy-to-operate optogenetic tool that responds to red light would represent a real breakthrough in optogenetics. Red light can penetrate more deeply into body tissue and with such a tool cancer patients might be treated one day.”

As the name suggests, optogenetics is a combination of optics and genetics. Researchers in the field use light to, for example, “switch on” or “switch off” individual genes. Using light as a trigger for molecular processes has a number of advantages. For a start, it elicits fast responses and is easy to use. In addition, light has no unintended effects on cells – unlike the small molecules that are otherwise frequently used to manipulate cells.

Her enthusiasm for her work is apparent in every sentence she utters: *“The development of a small, single-component, easy-to-operate optogenetic tool that responds to red light would represent a real breakthrough in optogenetics. Red light can penetrate more deeply into body tissue and with such a tool cancer patients might be treated one day.”*

As a computer engineer who got into molecular biology at a relatively late stage in life synthetic biology is the perfect discipline for her. *“It enables me to perfectly combine my know-*



Good science requires good teamwork: Prof. Di Ventura believes it is essential for a lab to be made up of committed researchers from many disciplines working together to achieve the same objectives (Photo: Harald Neumann).

ledge from both areas,” she says with a broad smile. Her lab has now grown to ten members. Eight of them perform wet lab experiments while two work on theoretical models.

“It may sound crazy, but I enjoy every day at work”

The Italian scientist feels very much at home in southern Germany, where she lives with her husband and son. But the international research is still very important to her. She enjoys the freedom and creativity that her research activities offer and loves interacting with her team and the students at the university: *“It may sound crazy, but I enjoy every day at work.”*

There is only one thing that dims the joy of the lively Italian a bit: *“Because I’m such a committed and motivated individual myself, I get frustrated if a student seems unmotivated. I am annoyed by a lack of dedication.”* Outside of lockdown periods, when Di Ventura is not at the university or an international conference, she can be found enjoying her other passion: music. *“I sing in a rock band, which I totally love to do. Of course, it’s not a prerequisite or anything, but any highly motivated scientist who also plays an instrument is really welcome to join our team!”* she jokes.

“Because I’m such a committed and motivated individual myself, I get frustrated if a student seems unmotivated. I am annoyed by a lack of dedication.”

Contact:

Prof. Dr. Barbara Di Ventura

Institute of Biology II

BIOSS and CIBSS Signalling Research Centers

Molecular and Cellular Engineering Group

University of Freiburg

barbara.diventura@biologie.uni-freiburg.de

<http://diventura.bioss.uni-freiburg.de>

my experience of the pandemic

2 years of COVID-19 from the perspective of a young scientist

by Sören Lukassen

When COVID-19 began to spread around the world in early 2020, it was clear that this pandemic was going to have a dramatic impact on our lives – at least in the short term. The situation in Wuhan showed that even a local outbreak in a country with good resources could push the healthcare system to its limits. Plus, very little was known at the time about the transmission, risk factors, and lethality of this novel coronavirus. My concerns back then revolved around existential and logistical questions: How can I prevent infection? What happens if my friends, family, or I get infected? But I didn't realize at this time the biggest upheaval the pandemic would cause in my life: How it would affect my work as a scientist.

As a bioinformatician, I was in the fortunate situation of being able to work from home, a lockdown would not have posed much of an obstacle for my work. Nevertheless, even before the first COVID-19 case was recorded in Berlin, my past caught up with me: Before I began my postdoc research on identifying functional gene sets in single-cell RNA sequencing data, I had worked for a while as a bioinformatician in a core unit where I was responsible for analyzing RNA-seq and single-cell RNA-seq data sets. My research group worked on several projects dealing with lung disease, and now someone was needed who could quickly merge the various control data sets, analyze them and turn them into a reference data set for other researchers. In the meantime, samples from as many COVID-19 patients as possible needed to be taken and sequenced in order to characterize the disease at the single-cell level.

What started as a short-term, impromptu project quickly developed into a Herculean task that involved many different parties. Our work put out several publications and data sets, which in turn were used by dozens of other groups for their own research. Working on this project was certainly exhausting for an early career scientist like me, but it was also a great learning experience. The fast pace of the work magnifies any problems and conflicts, but it also allows one to watch processes play out in fast motion that would usually take years and are difficult for younger scientists to grasp. In this article, I would like to share a few of the things I have learned in the past 24 months.

The urgency of a pandemic demands cooperation

Many research projects have one or a few scientists at the helm who drive the work forward. A major goal of doctoral and post-doctoral students is to be lead author of one's "own" paper. Although there is the possibility to share lead authorship, the fact that a paper's authors are often referred to as "Doe *et al.*" shows that particular importance is placed on being first in the list. Generally, this person is a postdoc or doctoral student who has spent many years gaining the necessary expertise and who has been the main driving force behind the project.

"What started as a short-term, impromptu project quickly developed into a Herculean task that involved many different parties."





Sören Lukassen is a junior group leader in medical omics (Photo: Sören Lukassen).

But this approach doesn't work in a situation like the COVID-19 pandemic. For one thing, it would be too slow, but also the emergence of a new disease means that no single research group has sufficient expertise to adequately analyze and interpret complex data sets. Instead, small-scale consortia need to form that bring together researchers with a wide range of expertise to work in close cooperation. In our case, there was almost daily communication between bioinformaticians, immunologists, virologists, and clinicians, which also provided a much deeper perspective than would have otherwise been possible. However, such intense collaborations also give rise to conflicts, which brings me to my next point.

Research can be a crash course in group dynamics

Management courses like to teach a team-building model, where a new group goes through the phases of tentatively getting to know one another, experiencing conflicts, resolving these conflicts, and eventually functioning as a productive team. This may be nothing new for veteran research group leaders, but early career scientists usually join existing research groups and thus rarely experience the dynamics of a newly formed team – and in collaborative projects, there is usually enough distance between the parties to keep the potential for conflict to a minimum. For a rapid analysis of the data on COVID-19, however, much closer collaboration was required:

About the author

Sören Lukassen is a junior group leader at the BIH Center for Digital Health at Charité. After studying molecular medicine and completing a PhD in human genetics, he worked as a bioinformatician at the Genomics Core Unit of Friedrich-Alexander University Erlangen-Nürnberg. During his postdoc in Heidelberg and Berlin, he focused on machine learning for single-cell RNA sequencing before starting research in the early 2020 on studies of the mechanisms that lead to severe disease in COVID-19.

data generation, analysis and interpretation were conducted simultaneously by groups with a wide range of expertise. In our case, these groups exchanged information on a daily basis. As there were no pre-existing structures, responsibilities had to be defined and clarified as we went. The participants had to agree on a plan for the entire project that represented the interests and capabilities of all groups. This led to heated discussions and conflicts in the early days, but it paid off in the medium term. Instead of avoiding differences of opinion, we had to confront and discuss these issues head-on – which, after an initial period of settling in, created a high level of mutual trust and a very efficient working atmosphere that would probably not have arisen in a looser sort of cooperation.

Short-term projects still require long-term planning

This close collaboration quickly resulted in several joint publications, the volume of data grew, and new ideas for projects kept emerging. In the beginning, our single-cell RNA-seq data sets contained some 20 samples and a few tens of thousands of cells – but this quickly grew to well over a hundred samples and hundreds of thousands of cells. While this provided us with unique insight into the biology of COVID-19, it was also becoming more and more of a problem, as our analysis pipeline couldn't keep up with the growing volume of data. When choosing our tools in early 2020, we had primarily looked for pipelines that were tried and tested and, most importantly, that had backward compatibility with other projects. This allowed for the rapid transfer of healthy control samples from other projects and significantly reduced the amount of work involved in the analysis. We also knew that our analysis scripts would work with up to 150,000 or 200,000 cells, which seemed sufficient given that we envisaged using perhaps 100,000 cells. However, as our initial projects quickly led to others, the pipeline's limitations started to become an issue. At first, we stuck with it anyway because we had already published findings using this approach and wanted to maintain consistency. In the financial world, this would be called a "sunk cost fallacy" – we had invested so much time in the analysis scripts and generated so much data with this pipeline that we couldn't bring ourselves to part ways. But eventually, of course, it got to the point where we had to rebuild the pipeline from scratch. The faster runtime the new pipeline provided far outweighed the time invested in the changeover – and got rid of the horror associated with the request: "Could you run this again with additional samples?"

This experience has made me realize several things:

- 1. Analysis pipelines and infrastructure should be designed to cope with data volumes that may seem unrealistic at first glance.**
- 2. The scalability of an algorithm is as important as its absolute performance. If this information is lacking, you should take the time to determine it yourself.**
- 3. As soon as it becomes apparent that the volume of data might end up too large for the procedures currently in use, you should promptly start planning a changeover; time invested in the old pipeline is not a reason to delay making this move.**

During the pandemic, a number of insights that were not entirely new have gained widespread acceptance. In the world of science, these include the pros and cons of preprints and the associated benefits of post-publication peer review. But in addition to what we have learned as a community, we have also gained smaller, personal insights – be it with regard to time management, working from home, or collaboration and project planning. If we can apply what we've learned from this whole experience, then at least in one respect some good will have come out of the past 2 years.

Contact:

Dr. Sören Lukassen

Group leader Medical Omics
BIH Center for Digital Health
Berlin Institute of Health at Charité
soeren.lukassen@charite.de

www.hidih.org/research/medicalomics



improving access to clinically relevant data

BMBF funding measure "i:DSem" summing up positively

by Björn Dreesen-Daun

Both clinical and experimental primary data, but also secondary data such as specialist publications, are growing explosively. The situation is magnified by the fact that these data are often very heterogeneous, i.e. they are unstructured, in different formats, and of varying quality. The research approach taken by "integrative data semantics" addresses this dilemma. The aim is to develop tools for homogenizing these heterogeneous data and to facilitate predominantly computer-based processing. This is a necessary prerequisite for using all clinically relevant data to develop better therapies in the future.

The funding measure "i:DSem – Integrative Data Semantics in Systems Medicine" is one of the initiatives of the German Federal Ministry of Education and Research (BMBF) aimed at advancing health research through innovations in computer-based data processing. The reason for launching the initiative in 2016 was that while more and more knowledge is being produced, significant hurdles stood in the way to making it available to physicians for problem solving. This challenge requires the development of novel tools, so-called semantic technologies, which are at the heart of the i:DSem measure.

At the end of 2021, the funding came to an end after just over five years and quite a number of successes. During the funding period, the BMBF spent more than €20 million to support eight interdisciplinary research projects.

The projects in the funding measure were divided into two phases. There was initially a three-year development phase, which was followed in 2019 by an interim evaluation. All projects then successfully entered a two-year translation phase, in which the developments made were validated for specific applications. With the completion of this phase, the projects have come to an end in the course of 2021.

Overall, the BMBF can look back on a successful measure, whose achievements were highlighted at the closing event in summer 2021. Since the launch of i:DSem, the BMBF has initiated further funding schemes to strengthen computer-based health research. For example, the measure "CompLS – Computational Life Sciences" aims in particular to explore the considerable potential of artificial intelligence for life sciences research. In addition, the BMBF-funded "Medical Informatics Initiative" is establishing important IT infrastructures at university hospitals. The goal is, among other things, to prepare and process routine clinical data in such a way that they can be used for medical research across institutions.

Using three projects as examples, we will illustrate the successes of the i:DSem scheme on the following pages.

Contact:



Dr. Björn Dreesen-Daun
Project Management Juelich
Forschungszentrum Juelich GmbH
b.dreesen@fz-juelich.de

www.ptj.de/idsem



Structured research (meta)data, models and applications:
the Health Atlas provides resources and follows the FAIR principles

As an interoperable health research platform based on the FAIR data principles, the Health Atlas (www.health-atlas.de) gives clinical and epidemiological researchers access to publications, structured data, models and applications.

Through the work of the Health Atlas, research resources become accessible on a platform that makes quality-assured data and metadata available for interoperable use along with publications. After registering on the platform, users can up-

What's unique about the Health Atlas?

In addition to publications and data, the Health Atlas provides access to numerous models, including disease prediction and risk assessment models as well as novel models for 3D-laser scan based body typing and COVID-19 pandemic models. All models have been created with modern web applications that allow for experimentation and use in research projects. This enables researchers to choose, for example, between various published genetic prediction models for familial breast

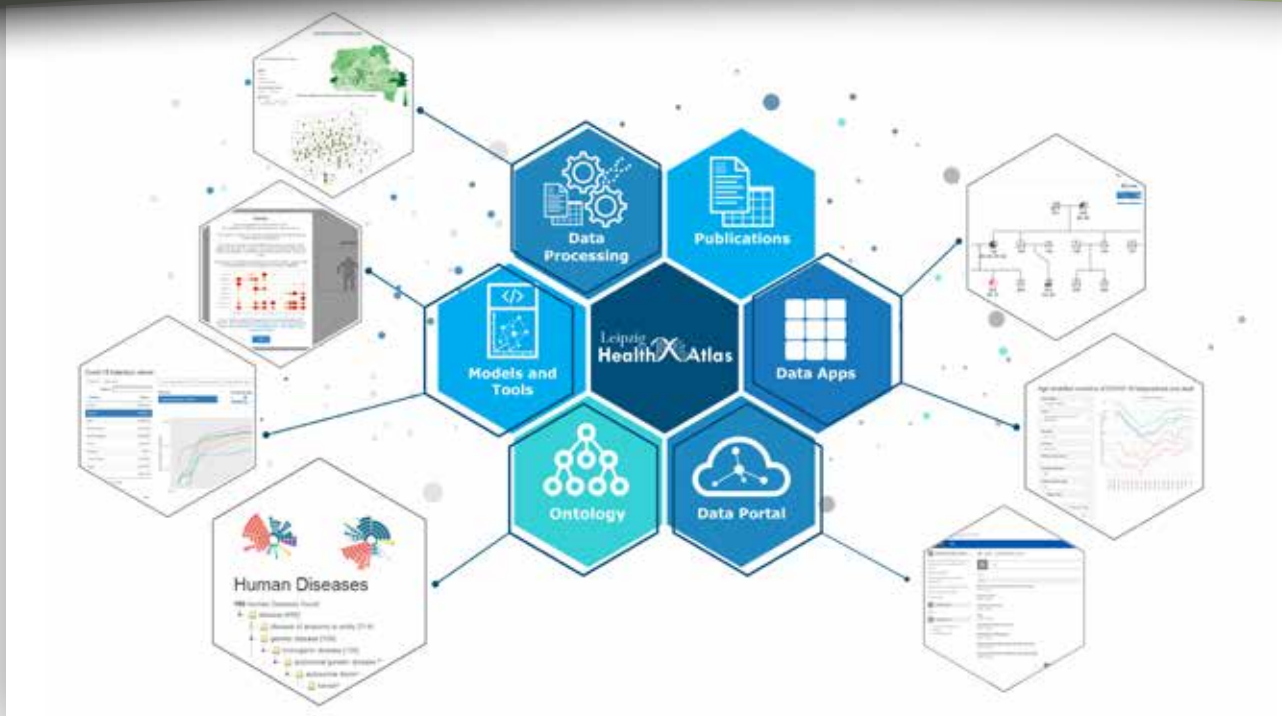


Figure 2: As an interoperable platform for health research, the Health Atlas provides a diverse range of resources and information. Key components include the data portal, models and applications to enable online data processing, and ontologies to improve searching (Source: René Hänsel, IMISE).

and ovarian cancer and compare the obtained results without having to implement the models themselves. Other models focus on normative data – available as tables in many publications – that are used to classify measured values from specific measurements. Researchers may use and include these data in more advanced analyses and visualizations. The platform also includes models that visualize the COVID-19 case statistics or provide classifications of the current pandemic situation.

What makes us FAIR?

The objective of the Health Atlas is to enable the shared use of research data. Consequently, we are committed to fulfilling the FAIR criteria. Unique identifiers (the “Health Atlas ID”) are used to reference content. The SEEK platform facilitates access via current interfaces in addition to user interfaces. Data access can be tiered.

At the online workshop “FAIR Data Infrastructures for Biomedical Communities” on October 15, 2020, Prof. Carole Goble of the University of Manchester, who among other roles is the Co-Director of the FAIRDOM Initiative, recommended the Health Atlas as an example of successful data sharing.

The Health Atlas, formerly known as Leipzig Health Atlas (LHA), was developed by the Institute for Medical Informatics,

Statistics and Epidemiology (IMISE), the Interdisciplinary Centre for Bioinformatics (IZBI) and the Leipzig Research Centre for Civilization Diseases (LIFE), under the auspices of the BMBF scheme “i:DSem – Integrative Data Semantics in Systems Medicine.” The Health Atlas is an integral part of other consortia of the Medical Informatics Initiative (SMITH) and the National Research Data Infrastructure (NFDI4Health).

Interested in becoming a part of the Health Atlas? Sign up now for free at:

<https://www.health-atlas.de/signup>

Contact:



Prof. Dr. Markus Löffler
Director of the institute for
Medical Informatics, Statistics and
Epidemiology (IMISE)
University Leipzig
info@health-atlas.de

www.imise.uni-leipzig.de/en

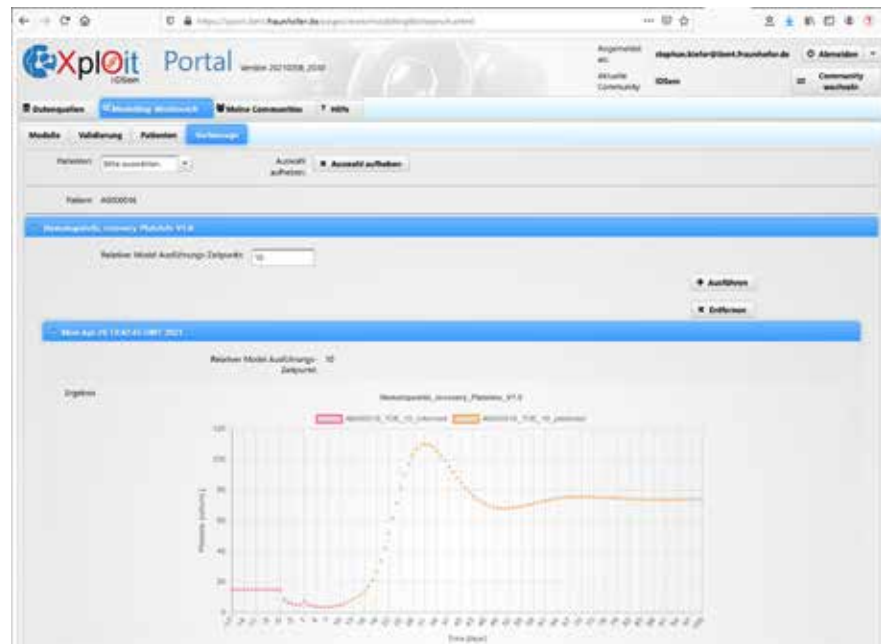


Figure 3: Results from a sample execution of the hematopoiesis model on the XploIt platform. The model was initialized using the red data points up to day 10 after stem cell transplantation. The yellow data points represent platelets predicted by the model up to day 100 after stem cell transplantation (Graph: Fraunhofer IBMT and Saarland University).

XploIt: health prediction models, developed efficiently

Computer-aided development and validation of predictive models in stem cell transplantation

by Stephan Kiefer

By assessing individual risks of complications and response to specific therapies, predictive computer models can help physicians provide personalized treatments. Although such models are urgently needed in stem cell transplantation, the wide range of data required, combined with the small number of cases, makes their design costly and clinical implementation difficult.

The recently completed XploIt project aims to tackle this challenge, focusing on the development of a new dedicated IT platform. This XploIt platform will make validated computer models available for stem cell transplantation medicine.

The XploIt platform

Developed by IT experts at the Fraunhofer Institute for Bio-medical Engineering (IBMT) and their colleagues at Saarland University, the web-based XploIt platform (Weiler *et al.*, 2018) allows clinicians to merge diverse clinical data for predictive models, share them with model developers, and test the resulting models – all in compliance with privacy regulations. Clinicians can use local tools to pseudonymize the required data, relativizing dates in relation to the transplant date, and make the data available in a data warehouse via upload through dedicated data pipelines. Information from semi-structured written reports can be extracted through tools provided by the industrial partner Averbis. This allowed the participating

transplant centers and virology institutes at Essen University Hospital and Saarland University Hospital to capture data from over 2,500 patients after allogeneic stem cell transplantation, with more than 15 million data points for model development.

In order to harmonize different data formats and terminologies, the project followed a data integration approach that employs ontology as a knowledge base for identifying the technical terms in transplantation medicine and their respective relationships. Using a designated tool, concepts that are required for this project-specific ontology can be extracted from established ontologies. This allows developers to annotate and merge data relevant to their planned model in the provided data sets, and then use analysis and visualization tools to find correlations and generate hypotheses. The data needed to train models can be transferred to the relevant environments. Newly developed models are then uploaded to a model repository and, after annotation of the required input parameters, fed with data from the platform's data warehouse and executed. To check a model's usefulness, it can be run retrospectively for different points in time in a treatment course and compared with the patient's actual health development or with assessments made by the clinicians treating the patient at the respective dates.

Predictive models for post-hematopoietic stem cell transplantation

Project partners at the University of Tübingen and Saarland University developed the first predictive models for patients after cell transplantation that individually predict possible life-threatening complications, such as the dreaded graft-versus-host disease and CMV reactivation on the one hand and new blood formation and survival on the other hand. This

helps clinicians to initiate life-saving measures earlier. Both machine learning models for predicting events and systems biology models for time course predictions were developed. Figure 3 shows an example of the prediction of platelet regeneration (highlighted as yellow data points) in a hypothetical patient after stem cell transplantation.

The XplOit platform supports evaluations that show how the models developed in XplOit stack up against clinician assessments and actual outcomes. The first such evaluation is currently nearing completion.

References:

Weiler, G., Schwarz, U., Rauch, J., Rohm, K., Lehr, T., Theobald, S., Kiefer, S., Götz, K., Och, K., Pfeifer, N., Handl, L., Smola, S., Ihle, M., Turki, AT., Beelen, DW., Rissland, J., Bittenbring, J. and Graf, N. (2018). XplOit: An Ontology-Based Data Integration Platform Supporting the Development of Predictive Models for Personalized Medicine. *Stud Health Technol Inform.*247:21-25.

Contact:



Stephan Kiefer

Deputy Head of the WG
Biomedical data and bioethics
Fraunhofer Institute for
Biomedical Engineering
stephan.kiefer@ibmt.fraunhofer.de

www.ibmt.fraunhofer.de/en.html



Figure 4: Overview of the IDSN architecture (A), examples of results from interactive analysis of data from SCA (B) and gene expression data (C). (A) provides an overview of the architecture described in the section “Architecture” in Madan *et al.*, 2018. (B) shows visual analyses of the comparison of SCA disease onset as a function of SCA gene variants 1, 3, and 6 and of corresponding allele length (left). (C) visualizes the expression of the MAPT gene as a function of different diseases as a table (left) and as a graph (right) (Source: IDSN).

the relevance of FAIR data for neurodegenerative research

BMBF project “Integrative Data Semantics for Neurodegeneration Research” paves way for complex data analysis

by Juliane Fluck, Mischa Uebachs and Maksims Fiosins

The vast amount of personal health data now available on Alzheimer’s, Parkinson’s and other neurodegenerative diseases makes new preventive and therapeutic interventions possible. By using semantic technologies and FAIR principles, the IDSN project breaks new ground for the development of therapies and the application of artificial intelligence (AI) in this field.

To reduce the suffering of patients and their families as well as the economic burden posed by an aging demographic, discoveries in the field of neurodegenerative diseases and dementia need to be translated rapidly into new therapies. The key aim of the BMBF-funded project “Integrative Data Semantics for Neurodegeneration Research” (IDSN, www.idsn.info) is to integrate data from different areas of research and combine these with existing disease information. Researchers at the German Center

for Neurodegenerative Diseases (DZNE), the Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), the University Hospital of Bonn, the University Medical Center Hamburg-Eppendorf and the ZB MED – Information Centre for Life Sciences have joined forces in IDSN to (1) develop standards, methods, and software for integrating data from high-throughput screening, clinical cohorts, and/or routine clinical data; and (2) use these data in new analyses. After a brief outline of the IDSN architecture, two application examples are illustrated.

IDSN architecture

The primary data are analyzed and stored locally. To make data interoperable, the results of the analyses are annotated semantically using standard ontologies and terminologies. These standards are stored in the separate semantic lookup platform SemLookP (<https://semanticlookup.zbmed.de>). This allows researchers to retrieve relevant concepts for annotating

data, to make information about the concepts available, and to provide mappings between different standards (for an overview, see Figure 4A). The semantically enriched data, including provenance of the original data, are transferred to a central data management platform (Madan *et al.*, 2018). Semantic indexing facilitates fast data queries and retrievals, while interactive graphical user interfaces allow for dedicated data visualization and fast and flexible data analysis (see Figures 4B and 4C).

Mechanisms of frontotemporal dementia

Frontotemporal dementias (FTD) are a heterogeneous group of dementia diseases associated with loss of function in the frontal and temporal lobes, often with early onset (before age 65) and frequently genetically determined or caused by metabolic abnormalities. One focus of IDSN is the combined AI analysis of clinical data and genetic, epigenetic, gene and protein expression data to gain new insights into FTD. The semantic annotation in IDSN allows these different data types to be related, enabling integral analysis of expression data and clinical patient data. Another benefit of IDSN semantics is the ease with which secondary data (e.g., associations between genes and diseases) can be linked and used. The combination of semantically annotated metadata, large data sets and novel AI models allowed us to discover new potential disease mechanisms of FTD, such as greatly increased vascularization of the diseased brain (Menden *et al.*: 2020 [1] and [2]).

Analysis of cohort data on spinocerebellar ataxias

Spinocerebellar ataxias (SCA) are rare inherited disorders of the cerebellum that result in a progressive impairment of balance and coordination along with physical disability. Collecting sufficient amounts of clinical data is extremely time- and resource-intensive. Within the IDSN project, four globally available clinical trials were merged by means of a specially developed semantic standard and import software to allow integrated analyses of the data (Uebachs *et al.*, 2020). Developing and mapping to this standard is far from trivial – it requires both clinical expertise and the capturing of similar information collected by different methods and scales. The result is a combined analysis in an intuitive interactive user interface developed in close cooperation with clinicians in IDSN (see Figure 4).

References:

Madan, S. *et al.* (2018). A Semantic Data Integration Methodology for Translational Neurodegenerative Disease Research. 11th Semantic Web Applications and Tools for Healthcare and Life Sciences (SWAT4(HC)LS). Antwerp, Belgium, December 3–6, 2018. CEUR Workshop Proceedings, Vol-2275, C.J.O. Baker, A. Waagmeester, A. Splendiani, O. Beyan, M Scott-Marshall, eds. Paper 9.

Menden, K. *et al.* [1] (2020). Deep learning-based cell composition analysis from tissue expression profiles. *Science advances* 6(30).

Menden, K. *et al.* [2] (2020). Integrated multi-omics analysis reveals common and distinct dysregulated pathways for genetic subtypes of Frontotemporal Dementia. <https://www.biorxiv.org/content/10.1101/2020.12.01.405894v3>

Uebachs, M. *et al.* (2020). A new tool for prompt and comprehensive visualization and comparison of cohort characteristics in spinocerebellar ataxias. SCA &ARCA Global Conference 19.-21.10.2020.

Contact:



Prof. Dr. Juliane Fluck

Head of Knowledge Management
Information Centre for Life Sciences
(ZB MED)

and Fraunhofer Institute for Algorithms and
Scientific Computing (SCAI)
and University of Bonn
fluck@zbmed.de

www.scai.fraunhofer.de/en/business-research-areas/bioinformatics.html



Dr. Mischa Uebachs

Specialist neurology
Clinic and Polyclinic for Neurology
University Hospital Bonn (UKB)
m.uebachs@gmail.com

<https://neurologie.uni-bonn.de>



Dr. Maksims Fiosins

Scientific associate
Medical Center Hamburg-Eppendorf (UKE)
Maksims.fiosins@gmail.com

www.uke.de/english/departments-institutes/institutes/medical-systems-biology/index.html

use of routine clinical data for research and care

Prerequisite: digital modular broad consent

by Clemens Spitzenfeil, Alexandra Stein, Martin Bialke, Torsten Leddig and Wolfgang Hoffmann

Information on disease progression and on the effect of therapies is of great importance for physicians and researchers alike. Until now, clinical trials and epidemiological studies have provided the basis for such information, whereas routine treatment data have typically not played a role. The advantage of these data is that their many different forms allow realistic insights into daily clinical practice. That is precisely the focus of the MII's efforts in the area of broad consent: it seeks to make such data available for research purposes. To ensure the success of this project, both an understanding of consent procedures by patients and an efficient implementation of the consent process are indispensable.

What is the potential of healthcare data?

Understanding the impact of risk and influencing factors on disease development and progression as well as the interactions between various pathologies is essential for improving diagnostics and therapy. Different clinical trials are currently compared to ensure a sufficient amount of data as reference material. However, as knowledge advances, the complexity of disease patterns is also increasing, and with it the demands on clinical trials. Data collection is further complicated by comorbidities or delayed effects, which are not always known in advance. The requirement for studies to specify a precise purpose cannot be fulfilled when such disease connections are not yet known. For many patients, this means participating in numerous studies and repeatedly undergoing extensive interviews and examinations.

Healthcare data, as it represents the entire spectrum of diseases encountered in clinical settings, offers the potential of significantly reducing effort on the part of both researchers and patients. Thus, they enable the investigation of disease interactions and the potential discovery of previously unknown illnesses. The data complement clinical trials, allowing researchers to discover new connections while also generating targeted hypotheses.

A crucial prerequisite for exploiting this potential is the adequate mapping of disease progression and treatment pathways through long-term, cross-sector collection of healthcare data. This data collection only works on the basis of pseudonymized data sets. The secondary use of such personal data is permitted under certain circumstances in accordance with Article 6 of the EU's General Data Protection Regulation (GDPR), with the informed consent of patients usually being the method of choice. The challenge in broad consent lies in obtaining consent for data use for as-yet undefined research projects over a long time period compared to consent for specific clinical trials that are limited in time and content. In addition, patient data can be stored and used for research purposes for up to 30 years after consent has been given, unless this consent is withdrawn. An independent ethics committee and a panel of experts at each participating hospital decide whether the data may be used for a research project.

The patient's perspective

As Richter *et al.* (2017) were able to show, patients generally have a positive attitude toward broad consent without any restrictions on the research area. The reasons for giving consent

are primarily prosocial, i.e., resulting from altruism, solidarity, or gratitude toward the treating physician. About 75% of patients also have a positive attitude toward allowing research with healthcare data without patient consent, pursuant to Art. 9 (2) lit j of the GDPR (Richter *et al.*, 2019). So theoretically, from the patient's perspective, it would be conceivable to abolish informed consent. However, from a legal and ethical perspective, and in order to achieve the most representative coverage of the patient population possible, it is advisable to offer patients an opportunity to make informed decisions. This can be realized through a broad consent policy – as already established in the field of biobanks, and currently being introduced as part of the Medical Informatics Initiative (MII) with a stronger focus on healthcare data. This approach seems particularly sensible, as the MII also provides for recontacting patients to report incidental findings, obtain additional information or even to recruit patients for topic-specific clinical and epidemiological studies. To ensure that patient goodwill is not lost, such recontacting should only be carried out with prior explicit permission (opt-in).

Digitalization of the consent process

A consistent and informed patient consent is essential for the use of healthcare data in a research context. It must be ensured that this process is as simple as possible. The same applies, of course, to the procedure of withdrawal of a consent.

Electronic collection and processing of consent information not only allows the patient's wishes to be correctly mapped, but it also enables the withdrawal of consent to be rapidly implemented. The GDPR – and some statutory provisions in Mecklenburg-Western Pomerania state law and other German states that make reference to it – as well as its implementation in German law currently do not provide for an explicit written form requirement in the meaning of Section 126 of the German Civil Code (BGB). This means that electronic consent is possible (Dierks *et al.*, 2021), thus allowing consent to be obtained through fully digital means (e.g., using a tablet, see Figure 1).

For the technical implementation of the consent process, a variety of technology standards and a small selection of free, established tools, such as the open-source software “generic Informed Consent Service” (gICS; Rau *et al.*, 2020), are available. The standardized mapping of consent data is essential for multisite projects such as the MII, which is why a generic mapping of modular consent was proposed (Bild *et al.*, 2020). This proposal has been expanded in the current work of the MII's Task Force for consent implementation¹ and is already being incorporated in current implementations.

¹ <https://www.ths-greifswald.de/en/update-gics-in-version-2-13-4-available>



Figure 1: Thanks to the open-source software gICS®, the process of obtaining broad consent is fully digital at the University Medicine Greifswald (Photo: A. Stein).

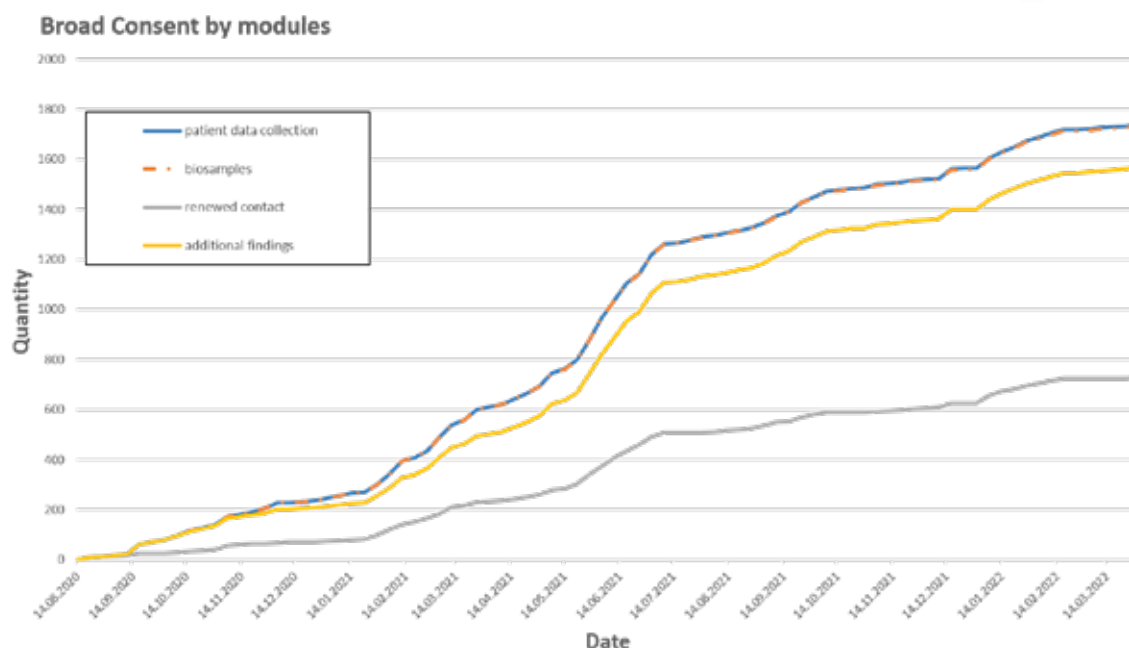


Figure 2: Number of consents obtained at the UMG in the period from August 14, 2020, to April 15, 2022, broken down by consent module (Source: C. Spitzenpfeil).

Initial experiences with e-broad consent

As part of the MII, several German university medical centers are already starting to ask patients for electronic consent (e-broad consent) when they are admitted to the hospital. We launched this process at the University Medicine Greifswald (UMG) in August 2020. This important milestone was preceded by an investigation into the setting in which a free and uninfluenced decision on broad consent is possible for patients. Specifically, the study examined the extent to which a spatial and personnel separation from the hospital admission process would be necessary so that patients can make a free decision. The study showed that, from the patient's point of view, the obtaining of broad consent directly following admission to the hospital would allow unrestricted and free decision-making. Nevertheless, it is important to make a clear distinction between admission, as a mandatory prerequisite for treatment, and information about broad consent, as a voluntary addition. Conducting a suitable discussion with patients is sufficient for this purpose.

To reduce the duration of the information process and also the effort for patient admission staff and patients, the MII's educational film on broad consent is shown in the waiting

area of the central admissions point. Experience has shown that this enables a good level of understanding of the patient regarding consent to be achieved.

Using this approach, the UMG has already been able to collect more than 1,700 broad consents with a manageable amount of effort on all sides. As a basis for future research projects, it is interesting to note that there is a high level of consent for the use of healthcare data and residual biomaterials for research, whereas recontacting of patients for the collection of additional data or inclusion in further studies is currently desired by only about a third of the consenting patients (see Figure 2).

Experience at the UMG has shown that the broad consent model can be well integrated into the inpatient admission process. For the success of broad consent, however, it will be essential not only to include inpatients, but also to cover the entire patient population. Corresponding field trials in selected outpatient clinics at the UMG are currently in preparation.



The authors (left to right): Clemens Spitzenpfeil, Prof. Dr. Wolfgang Hoffmann, Dr. Martin Bialke, Dr. Torsten Leddig, Dipl.-Jur. Alexandra Stein (Photo: Martin Bialke).

References:

Richter G. *et al.* (2017): Broad consent for health care-embedded biobanking: understanding and reasons to donate in a large patient sample. *Genetics in Medicine*, Volume 20, Number 1

Richter G. *et al.* (2019): Patient views on research use of clinical data without consent: Legal, but also acceptable?. *European Journal of Human Genetics* (2019) 27:841-847

Rau H. *et al.* (2020): The generic Informed Consent Service gICS®: implementation and benefits of a modular consent software tool to master the challenge of electronic consent management in research. *Journal of Translational Medicine* (2020) 18(1):287

Dierks C. *et al.* (2021): Data Privacy in European Medical Research: A Contemporary Legal Opinion. (Berlin: Schriftenreihe der TMF, Band 18)

Bild R. *et al.* (2020): Towards a comprehensive and interoperable representation of consent-based data usage permissions in the German medical informatics initiative. *BMC Medical Informatics and Decision Making* (2020) 20:103

Contact:



Prof. Dr. Wolfgang Hoffmann, MPH
Institute for Community Medicine
Dept. Epidemiology of Health Care and
Community Health
Greifswald University Hospital
wolfgang.hoffmann@uni-greifswald.de

www2.medizin.uni-greifswald.de/icm/index.php?id=18&L=1

About the Department of Epidemiology of Health Care and Community Health

The Department of Epidemiology of Health Care and Community Health (ICMVC) of the Institute for Community Medicine at the University Medicine Greifswald (UMG) deals with topics such as data protection and its implementation in medical research. For this purpose, the ICMVC set up the Independent Trusted Third Party of the University Medicine Greifswald in 2014, and has developed open source software tools (under AGPLv3 license) to support the protection of identifying data, to implement pseudonymization processes and to manage patient consent. In addition to facilitate the technical aspect of the consent process, the ICMVC worked early on to establish the practice of broad consent at the UMG. The ICMVC is a member of the MIRACUM consortium, funded by the BMBF, grant number: 01ZZ1801M, and also serves as its Consent Competence Center.

lost in translation? semantic standardization!

Terminologies like LOINC and SNOMED CT reduce Babylonian Confusion of tongues

by Cora Drenkhahn and Josef Ingenerf

Healthcare data are increasingly being stored digitally and thus contain a treasure trove of hidden insights for research and quality assurance. Yet a lack of semantic standardization and shortcomings in the currently used coding systems prevent a targeted analysis of these data. In order to exploit this wealth of knowledge, it is essential that we record the data in a meaning-preserving way using terminologies like LOINC or SNOMED CT. In combination with the internationally emerging standard HL7 FHIR, a number of promising interoperable software solutions are currently arising. In this article we will illustrate why it is worth to invest in the use of codes from more sophisticated terminologies.

Routine healthcare continuously generates substantial amounts of medical data on hundreds of thousands of individual patient cases. With the advance of digitalization, these data are being documented and communicated electronically at an increasing pace. This includes essential core information such as diagnoses, procedures, laboratory tests and medications that are documented and billed by the healthcare provider. Ideally, this and other information would be reused for further purposes. Firstly, an improvement of the quality of patient care itself could be achieved by avoiding unnecessary redundant examinations and medical errors due to a lack of documented diseases and medications – as long as the relevant information is available in the right form, at the right time and place. Secondly, in the long term an improvement in treatment

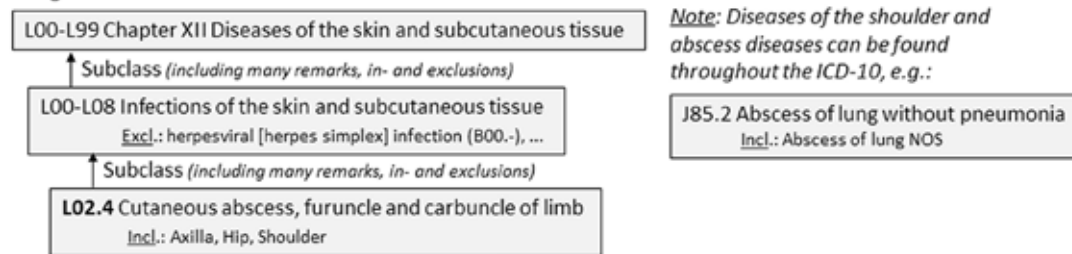
options results from the comprehensive use of patient data for clinical research – provided the patient's consent is obtained. By pooling large amounts of data and analyzing them jointly, researchers can identify overarching developments and deduce disease-relevant correlations, which inevitably lead to new scientific discoveries that will ultimately benefit individual patients (Safran *et al.*, 2007).

Natural language variability hinders the reusability of collected data

Yet the vast potential of such healthcare data, which are available in principle, has so far remained largely untapped. One major reason is the vagueness of human communication in general, and medical language in particular. Abbreviations such as “PIN” contribute further to this **Babylonian confusion of tongues**. The ambiguous abbreviation (homonymy) can be interpreted as meaning “Prostatic intraepithelial neoplasia” or “Pressure-induced nystagmus”, depending on the context. Conversely, a condition such as a skin abscess can be paraphrased in different ways (synonymy), for example, with “abscess of the skin”, “cutaneous abscess”, or “Hautabszess” in German. Multilingual synonyms are particularly relevant for international research. Many other phenomena such as quasi-synonymy (e.g., “abscess” versus “furuncle”: are the two the same thing?) and negation (e.g., “not an extensive abscess”: is the abscess not extensive or is there no abscess at all?) make reliable interpretation of linguistic data difficult. This is true for humans, but it is especially true for computers, which is the subject of this article.

Coding of „Abscess of left shoulder“:

using ICD-10:



using SNOMED CT:

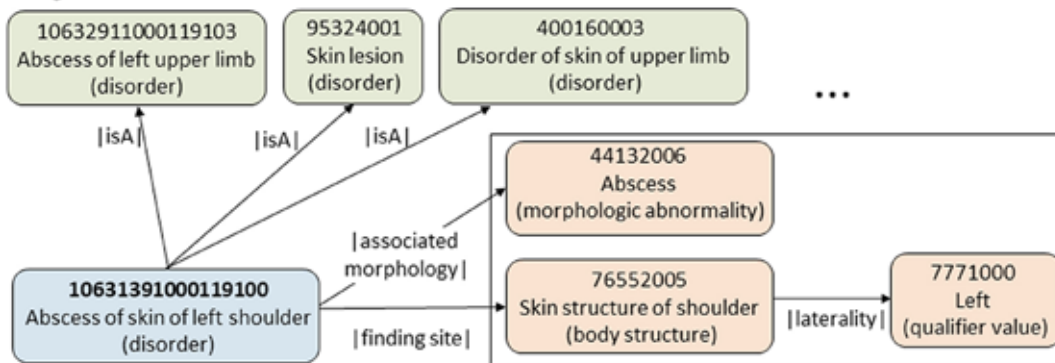


Figure 1: A comparison of ICD-10 classification versus the more expressive SNOMED CT terminology (Source: Created by Josef Ingenerf).

In this respect, the widespread application of free text in medical documentation complicates the desired reuse of healthcare data for purposes such as research. A central aim of the Medical Informatics Initiative (MII) is thus to facilitate the secondary use of routine healthcare data for research (Semler *et al.*, 2018). Although advanced techniques in natural language processing (NLP), text mining, and machine learning are quite successful by now, essential information can currently not be extracted reliably from free text documentation under the premise of ensuring their validity and completeness (Névéol *et al.*, 2018). However, in the view of the aforementioned peculiarities of natural language – synonymy and homonymy, implicit context, abbreviations, and negations – this finding is not surprising. The physician's favored flexibility of free-text descriptions permitting individual expressiveness becomes informatics' downfall.

Why coding with ICD-10 is not a solution

Currently, the usage of classifications like the International Classification of Diseases, 10th Revision, German Modification (ICD-10-GM) is legally binding in Germany. Diagnoses coded in this way are primarily utilized for billing purposes or statistical analyses. However, this practice is insufficient to actually facilitate diverse possibilities of reuse. The approximately 13,000 predefined codes represent disjoint and exhaustive classes organized in strict hierarchies, so that a clear assign-

ment of each diagnosis to exactly one ICD-10 class is achieved. Although useful for statistics – each diagnosis is only counted once – the use of classes as usually rather coarse categories means a loss of information that hinders or prevents the data's dynamic application for further analysis. Furthermore, the semantic meaning of the classes can only be evaluated by computers to a very limited extent, especially since their inclusion/exclusion criteria can only be interpreted by humans. In the upper half of Figure 1, one can see ICD code L02.4, which constitutes a rather fuzzy category in which related conditions (e.g., “furuncles”) are also included, but a causation by herpes viruses is excluded. The exact semantics of the code is “invisible” to computers, since it is noted partly in textual comments. It is neither possible to access all abscess diseases as relevant classes are scattered throughout the classification, nor to specify the anatomical location (including laterality) more precisely.

Only semantic standardization using terminologies ensures interoperability and reusability of patient data

Hence, semantic interoperability cannot be achieved by using simple classification systems. Rather, it is necessary to represent relevant medical facts in a standardized, purpose-agnostic format that can be interpreted and further processed by computers, if possible without loss of information. Various **terminologies**, of which LOINC and SNOMED CT are two of the best known and most widely used systems, meet this requirement. Due to the worldwide acceptance of these terminology standards, the

LOINC	Long Common Name	Component	Property	Timing	System	Scale	Method	Class
595-9	Bacteria identified in Abscess by Aerobe culture	Bacteria identified	Prid	Pt	Abscess	Nom	Aerobic culture	MICRO

Figure 2: LOINC code and description, six axes and values, and clinical term class (here “Microbiology”) (Source: <https://loinc.org/search/>).

data can even be communicated and used for computer-based analyses on an international level. Nevertheless, medical documentalists are not confronted with long, obscure codes as the corresponding medical concepts in the standards are linked to natural language terms that can be read and understood by human users. By ensuring a good integration of these standards into (existing) software systems, the burden of documentation can be relieved while preventing incorrect entries.

Logical Observation Identifiers Names and Codes (**LOINC**) is a freely available international standard created by the U.S.-based Regenstrief Institute. It is primarily used for the standardized representation of laboratory tests and other clinical observations. A LOINC code represents the unique combination of at least five axes (highlighted in Figure 2) that explicitly state the test’s characteristics.

SNOMED Clinical Terms (SNOMED CT) is of particular importance and relevance (Drenkhahn and Ingenerf, 2020). In early 2021, Germany became a member of the governing body SNOMED International, so the terminology standard can now be used nationwide by licensed organizations and individuals. Unlike other semantic standards limited to specific disciplines, SNOMED CT is a reference terminology that aims to represent information and knowledge in the medical field as comprehensively as possible. Moreover, SNOMED CT has a complex structure as well as special features that, when combined, enable advanced functionalities.

Essentially, SNOMED CT provides so-called concepts that represent clinical ideas to which unique identifiers are assigned. Each of these approximately 350,000 concepts has several natural language designations (descriptions) linked to it, which are very helpful when used in search queries as synonyms, for example, during data entry. However, it is mainly the third component of SNOMED CT that contributes to its superiority among terminologies: Through the help of relationships, a machine-processable definition of the concept’s meaning is achieved. This logical formalization results in complex links between the concepts (see lower part of Figure 1), which can then be used for inferring computer-based conclusions (with the aid of a description logic reasoner). SNOMED CT therefore provides the best means for both performing complex data analytics and the multi-purpose reuse of coded data.

Recent developments: FHIR advances terminology standards

As is already becoming apparent, the lacking adoption of semantic standards within clinical IT-systems is not the only reason for their slow uptake. In particular, the complex and heterogeneous nature of the various terminologies has so far hampered a widespread implementation (Rector, 1999). But recently, there are promising developments in this area: At the level of necessary information models, the structural standard Health Level 7 – Fast Healthcare Interoperability Resources (**HL7 FHIR**) has been providing new momentum for several years. In addition to standardized data structures – called resources – for describing medical facts (e.g., for patients, allergies or medications), FHIR also offers a specialized module for terminology services. With the data structures and functionalities defined here, terminology content can be uniformly represented as well as queried. Based on these premises, semantic components can be outsourced to a dedicated terminology server with which relevant applications can easily communicate via FHIR’s standardized operations, as shown in Figure 3.

Several current projects and initiatives are already demonstrating the synergistic benefits achievable by combining semantic and structural standards with terminology servers. Firstly, the MII’s HiGHmed consortium is using the FHIR-based Ontoserver from the Australian company CSIRO as a central hub for providing semantic content relevant to its three use cases: Cardiology, Oncology, and Infection Control (Haarbrandt *et al.*, 2018). Secondly, a similar approach applies to the CODEX project of the German Network of University Medicine (NUM) which aims to establish a national research data platform for COVID-19. And thirdly, the German National Association of Statutory Health Insurance Physicians (NASHIP, German: Kassenärztliche Bundesvereinigung (KBV)) also sets an increasing focus on standardization using FHIR on the one hand and terminologies – primarily SNOMED CT – on the other hand for the specification of so-called medical information objects (MIOS).

In summary, much needed progress in semantically interoperable health information systems is becoming apparent in recent times. At the beginning of last year, Germany became the 40th country to join the non-profit organization SNOMED

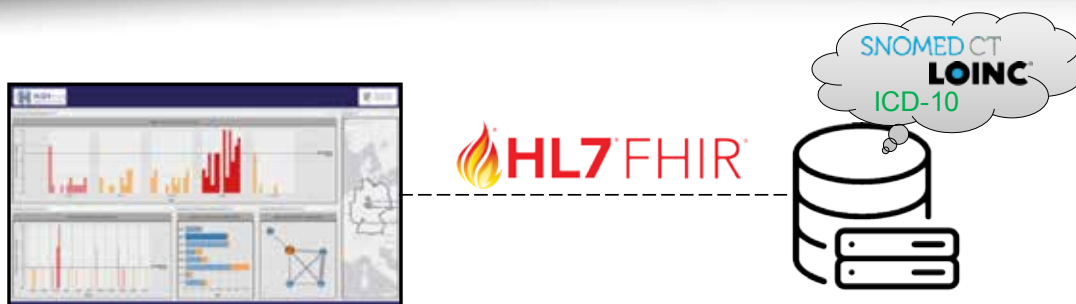


Figure 3: Applications use semantic standards via FHIR-compliant terminology servers

(Created by Cora Drenkhahn; sources: https://www.informatik.tu-darmstadt.de/gris/forschung_1/visual_search_and_analysis/_visual_search_and_analysis_projects_1/project_highmed/_project_highmed___interactive_visualization_for_infection_control_1.en.jsp, <https://www.hl7.org/fhir/overview.html>, <https://de.cleanpng.com/png-j1o44m/>, <https://loinc.org/>, <https://commons.wikimedia.org/wiki/File:Snomed-logo.png>).

International. Using structural standards such as HL7 FHIR, data structures (including codes from semantic standards like LOINC and SNOMED CT) are defined for the exchange and sharing of electronic health information; both for legally required purposes (e.g., the implementation of a vaccination passport scheme) and for research initiatives such as the aforementioned projects MII and NUM. Like this, data from both “worlds” should become increasingly interoperable and reusable so as to benefit both patient care and research efforts.

Safran, C., Bloomrosen, M., Hammond, W.E., Labkoff, S., Markel-Fox, S., Tang, P.C., and Detmer, D.E. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association* 14, 1–9.

Semler, S.C., Wissing, F., and Heyder, R. (2018). German Medical Informatics Initiative: A national approach to integrating health data from patient care and medical research. *Methods of Information in Medicine* 57, e50.

References:

- Drenkhahn, C., and Ingenerf, J. (2020). Referenzterminologie SNOMED CT. *Forum der Medizin_Dokumentation und Medizin_Informatik* 22, 81–84.
- Haarbrandt, B., Schreiweis, B., Rey, S., Sax, U., Scheithauer, S., Rienhoff, O., Knaup-Gregori, P., Bavendiek, U., Dieterich, C., Brors, B., *et al.* (2018). HiGHmed – An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. *Methods Inf Med* 57, e66–e81.
- Névél, A., Dalianis, H., Velupillai, S., Savova, G., and Zweigenbaum, P. (2018). Clinical Natural Language Processing in languages other than English: opportunities and challenges. *J Biomed Semantics* 9, 12.
- Rector, A.L. (1999). Clinical terminology: why is it so hard? *Methods Inf Med* 38, 239–252.

Contact:



Cora Drenkhahn
Research Associate
Institute of Medical Informatics
IT Center for Clinical Research
University of Lübeck
c.drenkhahn@uni-luebeck.de

www.imi.uni-luebeck.de

About the research project

The campus Lübeck of the University Hospital of Schleswig-Holstein (UKSH) joined the HiGHmed consortium of the Medical Informatics Initiative (MII) in 2019. In the use case “Infection Control,” seven partner sites are working together to establish an early detection system for nosocomial infection outbreaks. The IT Center for Clinical Research in Lübeck is providing support to this effort, primarily in the form of terminology management and terminology server setup. With a similar focus, the Lübeck site has been involved in the COVID-19 Data Exchange Platform (CODEX) subproject of the German Network of University Medicine (NUM) since 2020.

"a learning healthcare system could already be a reality"

Interview with Dr. Martin Hager

Medical Therapeutic Area Lead in the Personalized Healthcare (PHC) division
at Roche Pharma

Dr. Martin Hager is convinced that personalized healthcare, where every instance of medical treatment or care generates new knowledge for the future, has already been possible for a long time. In this interview, he talks about the current opportunities and challenges and emphasizes the central role of networking and exchange.

gesundhyte.de: Dr. Hager, Swiss multinational Roche is committed to personalized medicine. What role does this already play in medical treatment today?

Dr. Martin Hager: First of all, I have to say that enormous progress has unquestionably been made in personalized medicine in recent years. Look at oncology, for instance, which has undoubtedly played a pioneering role. Nowadays, we have a wide range of biomarker-based therapies available. And every year, new therapies are added that are specifically tailored to even small patient populations. We have witnessed remarkable progress in our understanding of diseases ushering in an era of new treatment approaches that would have been unimaginable just a few years ago. But that's not the whole story. In reality, these advances are not benefiting all patients, and the concept of personalized medicine has not been widely incorporated in all aspects of our healthcare system – at least not to the extent it could be.

"We need to work together right now to develop solutions targeted purely at achieving better outcomes for patients."

gesundhyte.de: How do you explain that?

There are many reasons, but a lot of them have to do with the fact that healthcare structures are simply not keeping up with the pace of innovation and technological progress. Take molecular diagnostics, for instance. Comprehensive tumor profiling using next generation sequencing is still the exception in routine clinical care. And so we have to ask ourselves: What is the benefit of having more and more personalized therapies targeting rare genetic alterations if the broader healthcare system does not identify the patients who could benefit from those therapies? What is needed is the courage to initiate



Photo: Roche

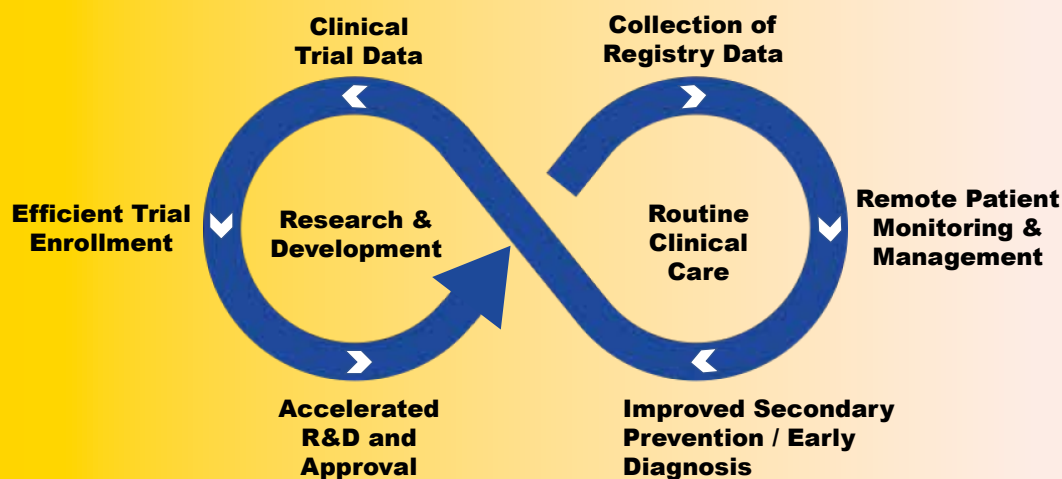


Figure 1: A knowledge-generating healthcare system (Source: Roche).

change. We must work together right now to develop solutions that have the potential to achieve better outcomes for patients.

gesundhyte.de: How is that possible?

There already exist some excellent initiatives to improve transfer of top-level medicine and clinical care at university hospitals to the broader healthcare system. I'm thinking, for instance, about the Hauptstadt Urologie network, which pursues precisely that goal – and also shows how important it is for different actors in the healthcare system to collaborate. We are currently working with Hauptstadt Urologie to develop precision oncology solutions that can help prostate cancer patients regardless of where they live. The project is addressing a wide range of relevant questions from digital support for quality-assured healthcare processes in the context of precision diagnostics to issues relating to data privacy and security.

gesundhyte.de: You mentioned digital solutions. What role do these play in personalized medicine?

"The chances for truly achieving and implementing data-based personalized healthcare have never been better."

Digital transformation of our healthcare system together with the systematic use of healthcare data are two components that play a pivotal role in the process as they represent the very foundation of personalized healthcare i.e. healthcare needs of individual patients.

gesundhyte.de: That sounds rather utopian!

Not at all! The technologies to achieve that have already been around for a few years. Think about modern IT solutions that structure and interconnect millions of data sets. Or digital health solutions that collect individual health data in real-time via wearables and smartphones. Additionally, artificial intelligence helps us analyze all these data. There has never been a better time to implement data-based personalized healthcare and make it a reality.

gesundhyte.de: A central aspect is the exchange of medical data with the healthcare system.

Absolutely. Data generated by the healthcare system play a central role. For instance, we assume that around 95 percent of potentially available oncological data is generated during routine examinations. If we are able to collect these data in a structured way in accordance with jointly developed quality standards, to interconnect and share them, we will be able to establish a healthcare system in which every medical treatment or care generates new knowledge for the future. In the end we have created a learning healthcare system.



Roche is creating personalized medicine (Photo: Roche).

gesundhyte.de: What do we need in order to do that?

Of course, we need the right infrastructure to replace isolated, stand-alone solutions. We also need rules to ensure the secure exchange of anonymized and pseudonymized health data. We believe it is crucial that researchers in the private sector are recognized as equal partners in the innovation process and have equal access to the data for their research. More than anything else, what is needed is the willingness of all participants in the healthcare system to communicate and cooperate beyond their own immediate interests.

Collaboration and knowledge transfer across different sectors is indeed possible – and worthwhile. We can clearly see that with a recent project with the University of Munich. Together, we showed how it is possible to apply knowledge gained from anonymized clinical genomic data to the treatment of future patients. Specifically, we developed a prediction algorithm that used the relevant data to determine which patients would benefit from molecular testing and when. I am convinced that these types of collaborative projects will

“Not only do the digital transformation of our healthcare system and the systematic use of healthcare data play a decisive role, but they are also a prerequisite if we want the idea of *personalized medicine* to come to life.”

become standard, as this is the only way we can shape the future of the healthcare system in a way that benefits the patients.

Franziska Müller conducted the interview.

Contact:

Dr. Martin H. Hager, MBA

Medical Therapeutic Area Lead Personalized Healthcare (PHC)
Roche Pharma AG

www.roche.de/ueber-roche/was-uns-antreibt/personalisierte-medizin

“We believe it is crucial that researchers in the private sector are recognized as equal partners in the innovation process and have equal access to the data for their research.”

personalized treatments vs. transparent patients?

How patient advocacy groups assess opportunities and challenges of AI in medicine and care

by Klemens Budde, Matthieu-P. Schapranow, Elsa Kirchner and Thomas Zahn – Plattform Lernende Systeme

Learning systems built upon latest artificial intelligence (AI) algorithms can be effective software tools for patients, physicians and medical professionals. Amongst others, they can support early prevention, improved decision making or the better selection of individual treatments. A qualitative survey shows that patient advocates believe AI in medicine and care holds potential for more personalized treatments as well as for faster, more comprehensive diagnoses. However, Deployment challenges include data misuse, cybercrime, erroneous and discriminatory decisions made by AI systems, poor data sets and poorly optimized algorithms. To overcome these challenges, experts call for greater cooperation among stakeholders involved in the research, development and application of AI technologies.

The debate over digital vaccination certificates has shown that Germany still faces major challenges in the digital transformation of healthcare. Pushing the digital transformation forward is important as it not only helps to control the spread of infections, but it is also a key prerequisite for adoption of learning systems in healthcare. Patient-specific medical data and AI-based systems have the potential to improve healthcare quality. They can support medical experts to optimize treatment outcomes through early prevention, improved decision making or better selection of individual treatments. Furthermore, they can accelerate the discovery of previously unknown medical correlations and innovative preventive approaches. Nowadays, learning systems make use of machine learning (ML) algorithms – a special subset of AI algorithms – to identify correlations within high-dimensional data, which otherwise requires manual and time-consuming involvement of medical experts. Thus, the adoption of learning systems can release medical and nursing staff from manual repetitive work, help them to spend more time in direct patient care, and provide



Publication “AI in medicine and care from the perspective of patients”

The paper is based on a report by the Working Group on Health Care, Medical Technology, Care of German Plattform Lernende Systeme. This report on a round-table discussion with representatives of various patient advocacy groups summarizes the participants' assessment of the opportunities and challenges for using AI systems in healthcare.

Klemens Budde *et al.* (ed.): “AI in medicine and care from the perspective of patients – A report on the round-table discussion with patient advocates,” Munich, 2020. The full German version is available online at: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG6_Whitepaper_Medizin_Pflege_Tagungsbericht.pdf

An executive summary in English is available at: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen_EN/AG6_WP_Patienten_Round_Table_Executive_Summary.pdf

The study's key findings were also presented in a paper published in *Deutsches Ärzteblatt*, which is available online: <https://www.aerzteblatt.de/archiv/216998/Kuenstliche-Intelligenz-Patienten-im-Fokus>



Platform members Elsa Kirchner (DFKI; in the background) and Klemens Budde (Charité – Universitätsmedizin Berlin; right) discuss AI and healthcare issues with patient advocates (Source: Plattform Lernende Systeme).

patients with long-term support in their daily lives. Building trust in these new technologies among patients and medical professionals requires high security standards for the relevant IT systems. Concerns regarding the use of AI-based medical decision support systems need to be identified, and the needs of patients and users must be considered during development of such solutions. Analyses of potentials and challenges of digital transformation and adoption of AI in medicine and care have largely ignored the perspectives of patients and medical professionals so far. Despite technological progress, concerns of patients and those in need of care have to be in focus also with regards to service quality as well as satisfaction with medical care. Patients must be involved in the various processes so that treatments can be optimally aligned with their needs.

The Working Group on Health Care, Medical Technology, Care of Germany's Platform for Artificial Intelligence (Plattform Lernende Systeme) in cooperation with the German Research Center for Artificial Intelligence (DFKI) and Charité – Universitätsmedizin Berlin, conducted a qualitative expert survey amongst patient representatives in Germany. Their objective was to identify the respondents' perspective on opportunities and challenges in the use of AI systems in medicine and care. Survey participants were asked to provide their perspective in a multi-stage, qualitative process following the Delphi survey method. The first step consisted of individual questionnaire-based interviews with twelve patient representatives from different advocacy groups. Respondents were interviewed on selected topics of digitalization and AI-based assistance systems

in healthcare. In the fall of 2019, Plattform Lernende Systeme and DFKI organized a round-table discussion as part of a second study based on these interviews. At this round table, ten of the twelve initial survey respondents and six other patient advocates discussed their various points of view based on the results of the qualitative preliminary survey. The participating patient advocates were selected to cover as wide a range of diseases as possible.

Using AI to improve treatment quality and strengthen patient autonomy

Overall, the results of the survey indicate that patient advocates have a rather positive, and in some cases ambivalent, view on AI technologies in healthcare. Respondents mentioned that AI technologies have great potential, but also pose some risks. At the same time, the level of knowledge about AI technologies varies among patient representatives as their use is still in its infancy. The majority of interviewed patient advocates were unable to answer the questions on the safety of AI-based medical decision support systems, which indicates a need for further education and awareness-raising. In addition, the majority of respondents called for more methodological transparency in AI-based technologies and applications in medicine and care. Therefore, additional efforts should be made to build up competence for AI technologies in medical professionals as well as patients equally.

Table 1: Potentials of AI in medicine from the perspective of patient representatives

IMPROVING THE QUALITY OF TREATMENT

- Personalized medicine
- Cross-sector processes and networked treatment
- Diagnostics
- Process and workflow optimization
- Knowledge diffusion and knowledge gain
- Counteracting underuse
- Reduction of side effects of treatment

STRENGTHENING PATIENT SOVEREIGNTY

- Strengthening patient rights
- Social participation and inclusion
- Support in everyday life

In the early stage of the discussion among all 16 participants of the second study, topics relating to **opportunities and patient advocates' hopes** regarding AI in medicine and care were rated more important than possible **risks and privacy concerns**. This shows that the potential opportunities of AI in healthcare are seen as outweighing the risks. Patient advocates are open-minded and have a generally positive attitude towards new technologies and their potential.

Patient advocates saw two key areas in which AI-based learning systems in medicine and care offer opportunities: AI-based medical decision support systems may improve the quality of treatment and strengthen patient autonomy. According to patient advocates, the most promising benefits of such systems would be improved diagnoses and a replacement of one-size-fits-all treatments towards more personalized treatments resulting in reduced side effects and treatment failures. Factors

rated as important by the respondents include personalized treatment goals, optimized therapy selection and timing, early detection of disease exacerbations, and new options for more holistic treatments. Frequently mentioned topics also included new diagnostic possibilities to detect disease earlier and more comprehensively. Just as frequently, respondents expressed a desire for improved cross-sectional treatment processes. They also argued for using learning systems to increase the transparency and efficiency of healthcare processes. For example, healthcare staff could make use of automated care documentation based on voice recognition to reduce documentation overhead. In addition, learning systems could promote better faster adoption of clinical guidelines in standard care, e.g. through medical decision support systems that are automatically kept up to date. Equally important for patient advocates is optimized access to knowledge (knowledge diffusion) and the acquisition of new scientific knowledge, such as through AI-supported analysis of study data and health-related data, including genetic data. Somewhat less importance was assigned to the possibility of offsetting shortages in medical care with new ways of accessing treatment, such as telemedicine and AI-based applications.

Furthermore, patients' rights could be strengthened by using AI-based medical devices. Particularly important in this context, according to the patient advocates surveyed, is providing patients with easy access to their electronic health records. New possibilities for inclusion and social participation as well as the resulting increased autonomy through AI technologies are also rated as important in this context – exoskeletons that

Table 2: Prioritizations of the patient representative groups

RISKS AND CHALLENGES OF KI IN MEDICINE

- ❗ Erroneous and discriminatory decisions of AI systems
- ❗ Risks to quality of care
- ❗ Dependencies and economization of the health care system
- ❗ Ethical issues

support the recovery of motoric skills were mentioned as an example. Furthermore, daily living aids received great attention, such as new ways of maintaining health with AI-based applications.

Concerns about inadequate privacy protection and misuse of data

A main concern reported by patient advocates was that medical or nursing staff might make erroneous or biased treatment decisions based on recommendations of learning systems (see Table 2). Patient advocates fear that incorrect interpretations of AI results or inferior data quality sets and algorithms may lead to wrong decisions or decisions that are detrimental to the patient. Thus, algorithms might use insufficiently traceable decision-making processes or be trained on irrelevant data points. While a critical analysis of the purpose, quality and size of data sets is required, there is a lack of universally accepted metrics and standards. Experts from the Plattform Lernende Systeme investigated specific attacks against AI systems and protective measures against such attacks in a recent publication.

Another aspect that patient advocacy groups are particularly critical of is the risk to the quality of care that AI technologies might pose. The greatest concern is that human interaction may be restricted by standardized treatment. Study participants also feared the dangers resulting from the fact that patients' concerns are not addressed sufficiently during the development of and the use of AI-based learning systems in medicine and care. Ethical questions about the social consequences of unhealthy self-optimization were also rated as relevant, and patient advocates declared it necessary that an ethical framework be developed to prevent adverse effects.

Patient advocates also expressed concerns that the deployment of AI systems may lead to new dependencies, such as regarding the access to data and software. In this context, study participants warned that overly strict regulations could hinder innovation and prevent the emergence of new start-ups.

The greatest danger the representatives of the various patient advocacy groups see in AI in healthcare is the possible lack of data security, data misuse, and cybercrime (see Table 3). To prevent this, they called for solid IT security as well as penalties for data misuse. Health data is essential for the development and deployment of AI-based medical decision support systems. Consequently, strengthening AI-based care while protecting data subjects requires judicious privacy protection. Key considerations in this context are the data subjects' informational self-determination as well as clear rules on who may use the sensitive data. Study participants pressed for clear data storage requirements on issues, such as confidentiality, availability, legally binding nature and data access rights. They fear that without such measures, data may be collected without subjects' consent or that the right to be forgotten may no longer be guaranteed. In addition, a contact point for legal assistance should be created, so that those affected can assert their rights in a self-determined manner.

The results of the survey of various stakeholder group representatives suggest options for action and the perspective of stakeholders on political regulatory requirements for the deployment of AI in medicine and care. The exchange with the patient advocates also revealed an ongoing need for awareness and education on how AI-based decision support systems work and what they can be used for already today. Thus, the poten-

Table 3: Prioritizations of patient advocacy – Key privacy concerns for AI in medicine

DATA MISUSE AND CYBERCRIME

- ! Sanctioning of data misuse
- ! Secure and stable IT systems

INFORMATIONAL SELF-DETERMINATION

- ! Ownership and use of data
- ! Lack of data storage requirements
- ! Legal assistance for data subjects
- ! Data collection without consent
- ! Right to be forgotten

tial and limitations of AI systems in medicine and care must be communicated to the public and discussed by policymakers. This will help to strengthen the perception of AI's social relevance and further increase acceptance. To this end, we need a broad public dialog that includes AI developers, future users as well as additional stakeholders.

Contact:



Prof. Dr. med. Klemens Budde is Co-Head of the Working Group on Health Care, Medical Technology, Care of Plattform Lernende Systeme and Deputy Medical Director of the Medical Department, Division of Nephrology and Internal Intensive Care Medicine at Charité – Universitätsmedizin Berlin.
klemens.budde@charite.de

https://nephrologie-intensivmedizin.charite.de/metast/person/person/address_detail/budde-1/



Dr.-Ing. Matthieu-P. Schapranow is member of the Working Group on Health Care, Medical Technology, Care of Plattform Lernende Systeme, Working Group Leader “In-Memory Computing for Digital Health”, and Scientific Manager Digital Health Innovations at HPI Digital Health Center, Hasso Plattner Institute, University of Potsdam.
schapranow@hpi.de

<https://hpi.de/digital-health-center/members/working-group-in-memory-computing-for-digital-health/dr-ing-matthieu-p-schapranow.html>



Prof. Dr. Thomas Zahn is member of the Working Group on Health Care, Medical Technology, Care of Plattform Lernende Systeme, and Vice Rector Research @ bbw as well as Professor of Business Information Systems, Data Science and Health Information Systems at the bbw University of Applied Sciences Berlin.
thomas.zahn@bbw-hochschule.de

www.bbw-hochschule.de/Home.html



Prof. Dr. Elsa Andrea Kirchner is member of the Working Group on Health Care, Medical Technology, Care of Plattform Lernende Systeme and Professor of Medical Technology Systems at University Duisburg-Essen as well as team lead of the team “Intelligent Healthcare Systems” at DFKI Robotics Innovation Center, Bremen.
elsa.kirchner@dfki.de

<https://www.dfki.de/en/web/research/research-departments/robotics-innovation-center>

ABOUT PLATTFORM LERNENDE SYSTEME

Initiated in 2017 by the Federal Ministry of Education and Research (BMBF), the Plattform Lernende Systeme brings together expertise from science, industry, government and society and supports Germany's further path to becoming an internationally leading technology provider. Approx. 200 members of the platform are organized in working groups coordinated by a steering committee. They demonstrate the personal, social and economic benefits of self-learning systems and identify challenges and design options.

innovative software tools for the life sciences

BMBF initiative CompLS provides €34 million in funding for the development of computational methods and analytical tools

by Daniel Heinrichs and René Wolf-Eulenfeld

To foster the development of innovative computational methods and analytical tools in the life sciences, the German Federal Ministry of Education and Research (BMBF) launched its Computational Life Sciences (CompLS) funding scheme in 2018. To date, 53 projects have been granted a total of €34 million in four funding rounds.

In the life sciences, advances in experimental methods and modern (high-throughput) technologies play a significant role in the development of innovative new tools and ideas. Today, more and more quantitative, time-resolved measurements of cellular systems are being carried out. At the same time, the volume of digitized data in patient care and clinical research is growing at a rapid pace. This increases the need for new bioinformatics tools and intelligent algorithms that enable the efficient processing and analysis of the data obtained, as well as new methods for the mathematical modelling and simulation of complex biological systems.

With its Computational Life Sciences (CompLS) funding scheme, the BMBF wants to further advance the development of innovative methods and software tools for bioinformatics processing, modelling and simulation in the life sciences. The aim is to provide life science research in Germany with efficient and reliable tools – particularly ones that incorporate artificial intelligence (AI) – that can adequately model and analyze data sets from various sources and data obtained using the latest experimental methods.

While the first call for proposals was open to all research topics, each subsequent call has focused on a different key area: deep learning in biomedicine, machine learning for cancer research, and AI methods for infection research.

The following two examples serve to illustrate the successes of the initiative so far.

Contact:



Dr. René Wolf-Eulenfeld
Project Management Juelich
Forschungszentrum Juelich GmbH
r.wolf-eulenfeld@fz-juelich.de



Dr. Daniel Heinrichs
Project Management Juelich
Forschungszentrum Juelich GmbH
d.heinrichs@fz-juelich.de

<https://www.ptj.de/computational-life-sciences>

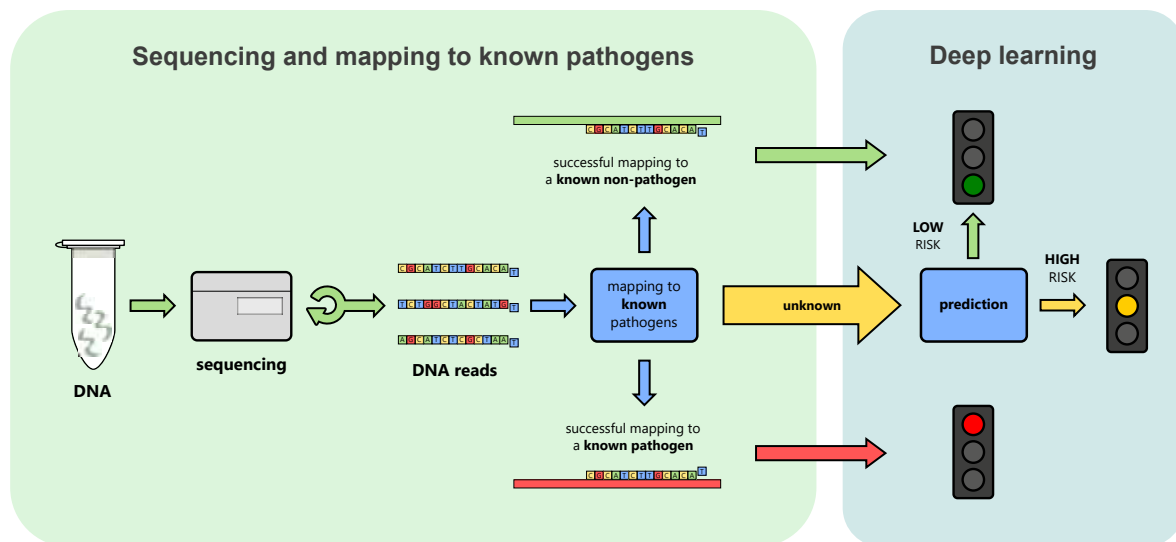


Figure 1: Workflow for predicting infectivity.

A biological sample is sequenced and the resulting reads are matched to genomes of known pathogens and non-pathogens. Reads derived from unknown viruses are then passed to a neural network to predict whether the new virus is capable of infecting humans (Source: Jakub Bartoszewicz and Melania Nowicka).

predicting the infection potential of novel viruses

Interpretable neural networks can quickly identify new viral pathogens from their genomes

by Jakub Bartoszewicz, Melania Nowicka and Bernhard Renard

New diseases are constantly emerging. However, standard molecular diagnostic methods rely on finding very specific matches to previously studied organisms and viruses, and can therefore only detect pathogens that are already known. Any new, significantly different viruses are easily missed. Researchers at the Hasso Plattner Institute are now developing new deep learning methods that can predict whether a novel virus has the potential to infect humans. The DeepPath project should allow scientists to identify pathogens very quickly from just a small amount of data, and to analyze entire genomes.

Over the past year, we have seen rapid progress in the development of tests to detect the SARS-CoV-2 virus: PCR tests, antibody tests, and eventually rapid or even self-administered

antigen tests quickly became part of everyday life in many countries. However, at the very beginning of the pandemic, no targeted tests had been developed. The best method for identifying and analyzing a completely novel threat is genome sequencing. This technology is particularly important in a time when more new pathogens – many of which are still unidentified and which evolve extremely quickly – are expected to emerge.

Genome sequencing and neural networks

Instead of looking for very specific pathogen signatures, genome sequencing aims to decipher the pathogen's entire genomic information. The sequencers read billions of nucleotides per sample, but the resulting reads only show fragments of the original genome – orders of magnitude shorter and pulled out of context. Putting these reads back in their original order is still a complicated task. Standard bioinformatics

methods for identifying pathogens therefore rely on matching these reads to databases of known genomes – and, as a result, can only identify pathogens that have already been studied. The DeepPath project, which is being funded by the BMBF and conducted by the Data Analytics and Computational Statistics research group at the Hasso Plattner Institute (HPI) in Potsdam, aims to change this.

The deep neural networks developed as part of the project can predict whether reads come from viruses capable of infecting humans or not. These machine learning processes recognize patterns that they associate with pathogens. They can therefore predict the “infectious potential” of any sequence – not just of those that resemble known genomes. This results in far more accurate identification. It even enables partial results to be analyzed while the sequencing device is still running, providing near real-time results. However, while the direct matching of similar sequences is relatively easy for an expert to interpret, neural networks only predict whether a sequence could be indicative of a new pathogen, without explaining why.

Interpretability for networks and new genomes

There are two different ways to address this challenge. First, the standard methods can be used in combination with the neural networks. This approach can both accurately predict the pathogenicity of all reads in a sample and match the reads to the genomes of closely related viruses. The second approach uses computational methods that visualize the most important patterns learned by the network. Regions with particularly high infectious potential can even be detected in previously unseen genomes. This allows scientists to both understand how the neural network arrives at its predictions, and to see which of the analyzed pathogens’ genes are markers for their ability to infect humans.

In the future, similar technology could be used to simultaneously analyze clinical samples that contain a mix of several different classes of pathogens (e.g., viruses, bacteria and fungi), or even to address safety issues in synthetic biology – as synthetic DNA sequences could be screened for potential hazards before they are actually synthesized. While the detection and study of novel pathogens will always require the participation of human experts, the interpretable deep learning approaches

developed as part of the DeepPath project offer new possibilities for extremely fast predictions from extremely little data. This could save precious time and allow us to thwart future biological threats before they even start to spread.

Project profile:

The **DeepPath project** is funded by the BMBF as part of its Computational Life Science funding scheme and is being carried out at the Hasso Plattner Institute with the support of the Robert Koch Institute. It is coordinated by Prof. Dr. rer. nat. Bernhard Renard, head of the Data Analytics and Computational Statistics research group at the HPI.

Contact:



Prof. Dr. rer. nat. Bernhard Renard

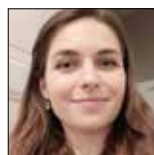
Hasso Plattner Institute,
Digital Engineering Faculty,
University of Potsdam
Bernhard.Renard@hpi.de

<https://hpi.de/forschung/fachgebiete/data-analytics-and-computational-statistics.html>



Jakub Bartoszewicz

Hasso Plattner Institute,
Digital Engineering Faculty,
University of Potsdam
Jakub.Bartoszewicz@hpi.de



Melania Nowicka

Hasso Plattner Institute,
Digital Engineering Faculty,
University of Potsdam
Melania.Nowicka@hpi.de

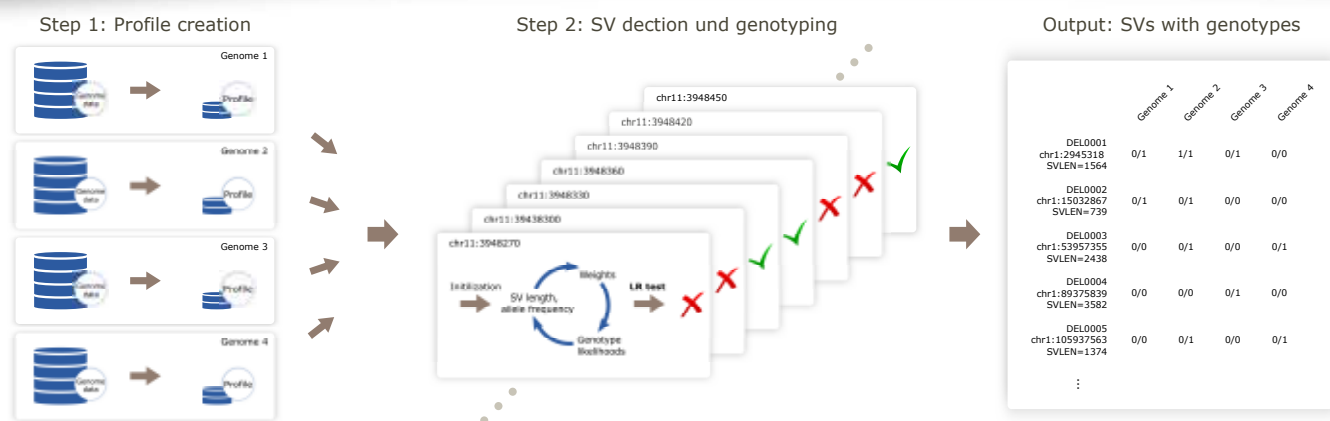


Figure 2: Overview of the calculation steps in PopDel.

First, PopDel reduces the input data of each genome into a small profile. Next, it uses likelihood ratio (LR) tests to determine for all genomes at once whether structural variants are present in each possible genomic region. The output of PopDel contains information on the position and size of the variants as well as the corresponding genotypes of all genomes (Source: Birte Kehr).

searching for variation in the sequence structure of our genomes

How the PopDel software can harness the potential of large volumes of data

by Sebastian Niehus and Birte Kehr

The DNA sequence of every human being contains a multitude of genetic variants that significantly influence our appearance and health. To uncover these differences and better understand their effects, the genomes of tens of thousands of individuals need to be searched and analyzed – a difficult and time-consuming task that has pushed traditional computational methods to their limits. However, new methods for analyzing huge amounts of data can overcome this challenge – providing valuable research insights and assisting in the development of new therapies for serious diseases.

In addition to the millions of variations at single DNA positions, thousands of variants that alter the structure of our genome sequences exist. These so-called structural variants (SVs) can take a wide variety of forms: DNA segments may be reversed (inversions), doubled (duplications), newly introduced (insertions) or lost (deletions). Certain SVs are associated with genetic diseases like Alzheimer's, muscular dystrophies and autism. Therefore, these variants are of great medical interest. To better understand SVs and their effects, thousands of human genomes are sequenced and analyzed.

Special computer programs analyze digitized DNA sequences and provide information about the presence of SVs. The rapidly growing amount of available sequencing data represents a great opportunity – but also poses an enormous technical challenge. As more and more genomes are sequenced, special approaches are required to cope with the vast quantities of data produced.

Information from tens of thousands of genomes

Our project Cohort-SV addresses this challenge. We developed a program called PopDel (Niehus *et al.*, 2021), which can jointly analyze the genomes of tens of thousands of people thanks to efficient algorithms, clever data management, and a tailored statistical model. This allows us to find out what SVs are present in each individual genome while also gaining information about the inheritance and distribution of SVs within a particular population.

To handle such huge quantities of data, PopDel works in two steps. In the first step, it extracts the relevant information from each sequenced genome and creates small individual “profiles” in which to store this information. This reduces the amount of data that needs to be analyzed to one to two percent of the original size, allowing PopDel to subsequently analyze the profiles of a

great many genomes simultaneously. In the second step, it reads the genetic information of all available profiles in small genomic sections and compares them against each other using a statistical model to provide reliable information about the presence of SVs. A particular challenge here is that, when analyzing very large numbers of genomes, different SVs may occur at the same site that need to be identified as different variants. In addition, it must be ensured that SVs that only occur in a single genome are not obscured by the thousands of other genomes. Our aim is to ensure that the quality of the individual results obtained when jointly analyzing tens of thousands of genomes is at least as good as those obtained by analyzing a single genome.

Discovery of a new genetic variant

To prove the quality of the results and their suitability for practical use, we put PopDel to the test in a variety of scenarios. Using simulated as well as real data from a single genome and from almost 50,000 genomes, PopDel succeeded in providing high-quality results. The most significant PopDel finding to date was a previously unknown deletion in the so-called LDL receptor gene by our partner deCODE genetics (Björnsson *et al.*, 2020). Since then, deCODE genetics has conducted further experiments to confirm that the deletion has a positive influence on the function of the LDL receptor. Given that the LDL receptor plays a central role in the cholesterol metabolism of our body, the discovery of this deletion can serve as a starting point for future research into the treatment of cardiovascular diseases.

Looking ahead

PopDel is presently able to reliably identify genomic deletions. In the current project phase, we are expanding the program to identify other types of SVs. Thanks to its fast running time and the fact, it is explicitly designed for use with large amounts of data, we will be able to apply PopDel to the ever-larger DNA sequencing studies of the future, helping us better understand SVs and the diseases associated with them.

Project profile:

Name of the project:

SV Cohort – Identifying and genotyping structural variation in whole genome sequencing data in large cohorts

Supporting institutions:

Leibniz Institute for Immunotherapy (formerly RCI); Berlin
Institute of Health at Charité – Universitätsmedizin Berlin

Participating partner:

Prof. Dr. Birte Kehr

References:

Niehus, S., Jónsson, H., Schönberger, J., Björnsson, E., Beyter, D., Eggertsson, H. P., Sulem, P., Stefánsson, K., Halldórsson, B. V., and Kehr, B. (2021). PopDel identifies medium-size deletions simultaneously in tens of thousands of genomes. *Nature communications*, 12(1), 1-10.

Björnsson, E., Gunnarsdóttir, K., Halldórsson, G. H., Sigurðsson, Á., Árnadóttir, G. A., Jónsson, H., Olafsdóttir, E. F., Niehus, S., Kehr, B., Sveinbjörnsson, G., Gudmundsdóttir, S., Helgadóttir, A., Andersen, K., Thorleifsson, G., Eyjolfsson, G. I., Olafsson, I., Sigurdardóttir, O., Saemundsdóttir, J., Magnusson, O. Th., Masson, G., Stefánsson, H., Gudbjartsson, D. F., Thorgeirsson, G., Holm, H., Halldórsson, B. V., Melsted, P., Norddahl, G. L., Sulem, P., Thorsteinsdóttir, U., and Stefánsson, K. (2020). Lifelong Reduction in LDL Cholesterol Due to a Gain-of-Function Mutation in LDLR. *Circulation: Genomic and Precision Medicine*.

Contact:



Prof. Dr. Birte Kehr

Head of Algorithmic Bioinformatics
Leibniz Institute for Immunotherapy
(formerly RCI)
Birte.Kehr@ukr.de



Sebastian Niehus

PhD Student Algorithmic Bioinformatics
Leibniz Institute for Immunotherapy
(formerly RCI)
Sebastian.Niehus@ukr.de

<https://www.rcii.de>

future-oriented analysis of research data

Members of the de.NBI network use AI-based technologies

by Vera Ortseifen, Peter Belmann and Alfred Pühler

In the digital age, research data are the basic currency of data-driven innovation. This new gold is playing a significant role in the rapid advancement of artificial intelligence (AI). But this also applies vice versa – the analysis of data can benefit immensely from AI-based methods. The potential is evident in the context of the German Network for Bioinformatics Infrastructure (de.NBI) with regard to AI-based projects. These projects also benefit from the excellent resources of the de.NBI Cloud, which is available free of charge to all researchers in the life sciences.

The importance of research data for AI-based developments

Making research data usable in a sustainable and long-term manner is the order of the day. The so-called FAIR data principles (Findable, Accessible, Interoperable, and Reusable) define the fundamental criteria for achieving this objective, and these are also lived out in the German Network for Bioinformatics Infrastructure (de.NBI) and the European Life Science Infrastructure for Biological Information (ELIXIR). In this context, the storage and management of data according to the FAIR data principles play an essential role in the current development of artificial intelligence (AI). It can be expected that the further development of AI will influence the progress of many research fields worldwide. Machine learning (ML) and deep learning (DL), for instance, thrive on the availability of data. The larger the amount of data, or the larger the training and testing data sets, the better the AI programs can be trained. For years this simple relationship has been leading to the increasing importance of AI approaches, driven by the constant growth in the size of data sets. The use of AI in the life sciences is therefore not only indis-

pensable, but also holds enormous potential for advancing areas such as personalized medicine, drug development, and basic biological research.

In this context, de.NBI (see “About the German Network for Bioinformatics Infrastructure” below) is growing in response to current developments and has recognized early on the opportunities that AI-based methods present for the analysis of research data. For this reason, the network’s members are also increasingly using AI-based technologies to analyze research data. The scope of AI projects results from a query within the network, according to which 25 projects from the areas of modelling, phenotyping, omics as well as meta-analysis, databases and support were reported back (Figure 1). Some of these AI-based projects of the de.NBI members will be explained in more detail below.

Identification of new antimicrobial resistance targets by high-throughput deep learning

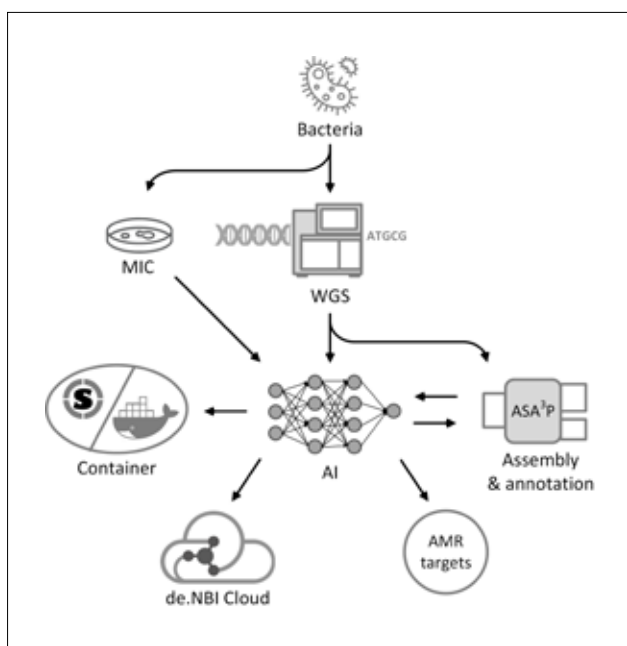
Scientists in the research groups of Alexander Goesmann, Dominik Heider, and Trinad Chakraborty at the Universities of Giessen and Marburg are working to identify new antimicrobial resistance targets. Antibiotic-resistant bacteria have become a serious threat to public health worldwide. Their surveillance and containment, as well as the timely identification of new genetic determinants of antibiotic resistances, are important tools in combating both emergent and established pathogens. To address these challenges, researchers from the Universities of Giessen and Marburg have launched a BMBF-funded project called “Deep-iAMR – Identification of new antimicrobial resistance targets by high-throughput deep learning.” This project combines expertise in medical microbiology, machine learning, and cloud computing techniques (Figure 2) to develop



Figure 1: The members of the de.NBI network are conducting 25 AI-based projects in different life science areas (Source: de.NBI office; background image: lpopba – stock.adobe.com).

automated, scalable bioinformatics workflows. Novel deep learning approaches are being used to rapidly and reliably predict antimicrobial resistances and uncover previously hidden signaling pathways and novel targets for the development of new antibiotics. This involves examining data for complex patterns between genomic and phenotypic data to find the genetic determinants of hard-to-detect or previously unknown anti-

microbial resistances. Yet the application of modern artificial intelligence methods to find these patterns requires massive computational resources as well as specialized hardware, such as graphics processing units (GPUs) and tensors. Researchers within the Deep-iAMR project are taking advantage of the de.NBI network's cloud computing infrastructure, which provides both the computational resources needed and the specialized hardware on demand.



Further information can be found on the project web page:

<https://www.gesundheitsforschung-bmbf.de/de/deep-iamr-identifizierung-von-neuen-antimikrobiellen-resistenz-targets-durch-deep-learning-10900.php>

Figure 2: The data flow within the Deep-iAMR project.

The project aims to combine various omics data sets with clinical and phenotypic information for a large, well-characterized set of multidrug-resistant *E. coli* isolates. These data will then be used to train deep neural networks. Our software is deployed via user-friendly containerization techniques and scalable cloud solutions. WGS = whole-genome sequencing, MIC = minimal inhibitory concentration, AMR = antimicrobial resistance (Source: O. Schwengers, K. Brinkrolf, S. Doijad, T. Chakraborty, D. Heider, A. Goesmann).

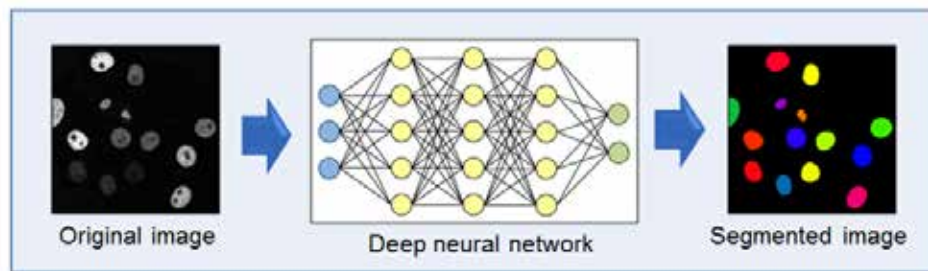


Figure 3: Deep neural network for segmentation of cell microscopy images (Source: C. Krug, K. Rohr).

Deep learning for microscopy image analysis and cellular phenotyping

The automated analysis of high-throughput and high-content microscopy image data is important to elucidate cellular processes. Yet analyzing such data poses a number of challenges. Deep learning methods within the field of artificial intelligence have recently emerged that yield superior results compared to classical image analysis methods.

The Biomedical Computer Vision Group at Heidelberg University, led by Karl Rohr, develops DL methods for the computer-based analysis of cell microscopy images. The aim is to improve the automated quantification of cellular phenotypes at the single-cell level and process large-scale microscopy data efficiently. This has mainly involved the development of deep learning methods for cell segmentation. The methods combine different types of neural network architectures, such as convolutional neural networks and recurrent neural networks.

The developed DL methods for cell segmentation have been used to analyze high-throughput and high-content microscopy image data. An example of a segmentation result for a cell microscopy image is shown in Figure 3. It can be seen that cells with high and low image contrast can be segmented well. The developed methods have been employed, for example, to segment cells in tissue images to subsequently quantify the length of the telomeres (chromosome ends), which can be used in medical diagnosis. This work was carried out as part of the BMBF project “CancerTelSys – Identifying cancer telomere maintenance networks for diagnosis, prognosis, patient stratification and therapy response prediction.” The development and application of the DL methods benefited from the de.NBI computing infrastructure and the de.NBI Cloud.

Further information can be found on the Biomedical Computer Vision Group’s web pages:

http://www.bioquant.uni-heidelberg.de/research/groups/biomedical_computer_vision.html

de.NBI Cloud: Computational resources for the application of artificial intelligence in the life sciences

Even though many important software programs in the life sciences employ machine and deep learning, there are requirements for their successful, high-performance application. In addition, artificial intelligence methods can be accelerated by the use of GPUs and benefit from access to enormous amounts of data, especially in DL. As already indicated in the two projects described above, the de.NBI Cloud (<https://cloud.denbi.de>) offers optimal computational resources for this purpose. GPUs are available in different versions and types and can be used flexibly depending on the requirements of the ML application. The de.NBI Cloud stores great volumes of scientific data, such as the Sequence Read Archive (SRA) metagenome data sets, allowing fast access to the information and eliminating the need to download data from different resources. GPUs, virtual processors (vCPUs), random access memory (RAM), and storage solutions are available free of charge to researchers in Germany and help foster the development and application of artificial intelligence methods in the life sciences (Figure 4).

The de.NBI Cloud already provides resources for a wide range of ML applications. Bioinformatics topics range from image analysis to antimicrobial resistance prediction to the modeling of 3D genome analyses, fluorescence microscopy, cancer diagnosis, and antiviral peptide prediction. Of the more than 600 projects computed in the de.NBI Cloud, at least 55 projects are employing machine learning methods within their research activities (status 02/2022). Numerous projects using DL or ML have already been published. Due to its federated structure, the fully academic de.NBI Cloud operates across six sites (Figure 4). Substantial staff support enables even

inexperienced researchers to access the resources. A total of nearly € 50 million has been invested in setting up the de.NBI Cloud.

Acknowledgments

The work on this article was supported by O. Schwengers, K. Brinkrolf, A. Goesmann (Deep-iAMR project), and C. Krug and K. Rohr (CancerTelSys subproject).

About the German Network for Bioinformatics Infrastructure

The German Network for Bioinformatics Infrastructure (de.NBI) was established in 2015 by the German Federal Ministry of Education and Research (BMBF) to provide researchers in the life sciences with an appropriate infrastructure for analyzing large amounts of data. In doing so, de.NBI offers comprehensive state-of-the-art bioinformatics services to users in basic and applied life science research from academia, industry and medicine. In addition, de.NBI promotes the cooperation of the German bioinformatics community with international network structures. Currently, the network consists of over 300 member scientists and 40 projects located across eight thematically focused service centers. The de.NBI infrastructure encompasses the areas of services, training, and cloud computing. The de.NBI Cloud is available free of charge to all life science researchers (see <https://www.denbi.de/cloud>). This unique computational resource is supplemented by analysis tools, workflows, and databases offered in the services area, whose use is supported by training courses.

Further information:

www.denbi.de

Contact:



Dr. Vera Ortseifen

Node Coordinator of the German Node of ELIXIR (European Life Sciences Infrastructure for Biological Information)
de.NBI Administration Office
c/o Center for Biotechnology
Bielefeld University
vera@cebitec.uni-bielefeld.de



Peter Belmann

de.NBI Cloud Governance
de.NBI Administration Office
c/o Center for Biotechnology
Bielefeld University
pbelmann@cebitec.uni-bielefeld.de



Prof. Dr. Alfred Pühler

Coordinator of the German Network for Bioinformatics Infrastructure
de.NBI Administration Office
c/o Center for Biotechnology
Bielefeld University
puehler@cebitec.uni-bielefeld.de

www.denbi.de

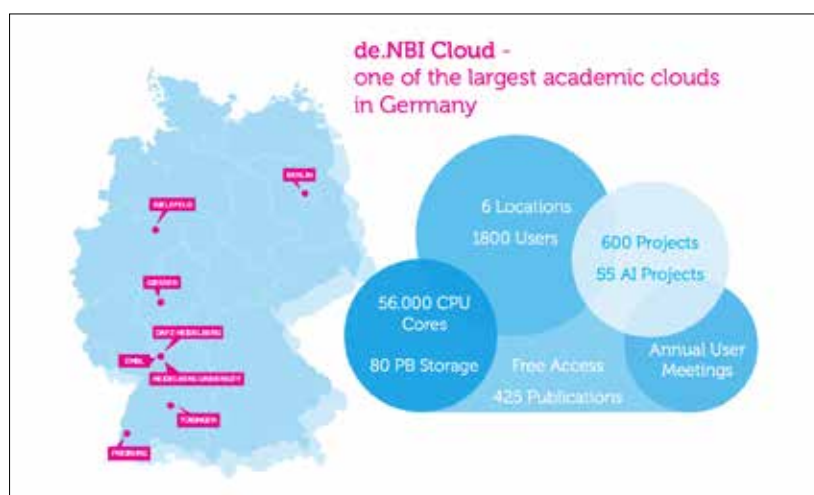


Figure 4: Information on the de.NBI Cloud

Sites across Germany, technical data, and computing facility usage (Source: de.NBI Administration Office, status 02/2022).

SmlCS

A smart solution for the automated detection of hospital outbreaks

by Antje Wulff, Claas Baier, Michael Marschollek, and Simone Scheithauer on behalf of all members and sites of the HiGHmed project's Infection Control use case and Infection Control Study Group

Background

It has been a well-known phenomenon for centuries that wherever people live together in groups, clusters of infectious diseases occur. To limit the spread of such diseases, it is important to trace the transmission routes and understand how clusters emerge. Depending on the epidemiological circumstances, we refer to such occurrences as endemic, epidemic or – as in the case of the current COVID-19 situation – pandemic. The term “outbreak” is often to describe clusters of pathogens and infections. Detecting such outbreaks at an early stage is a central challenge.

The challenge

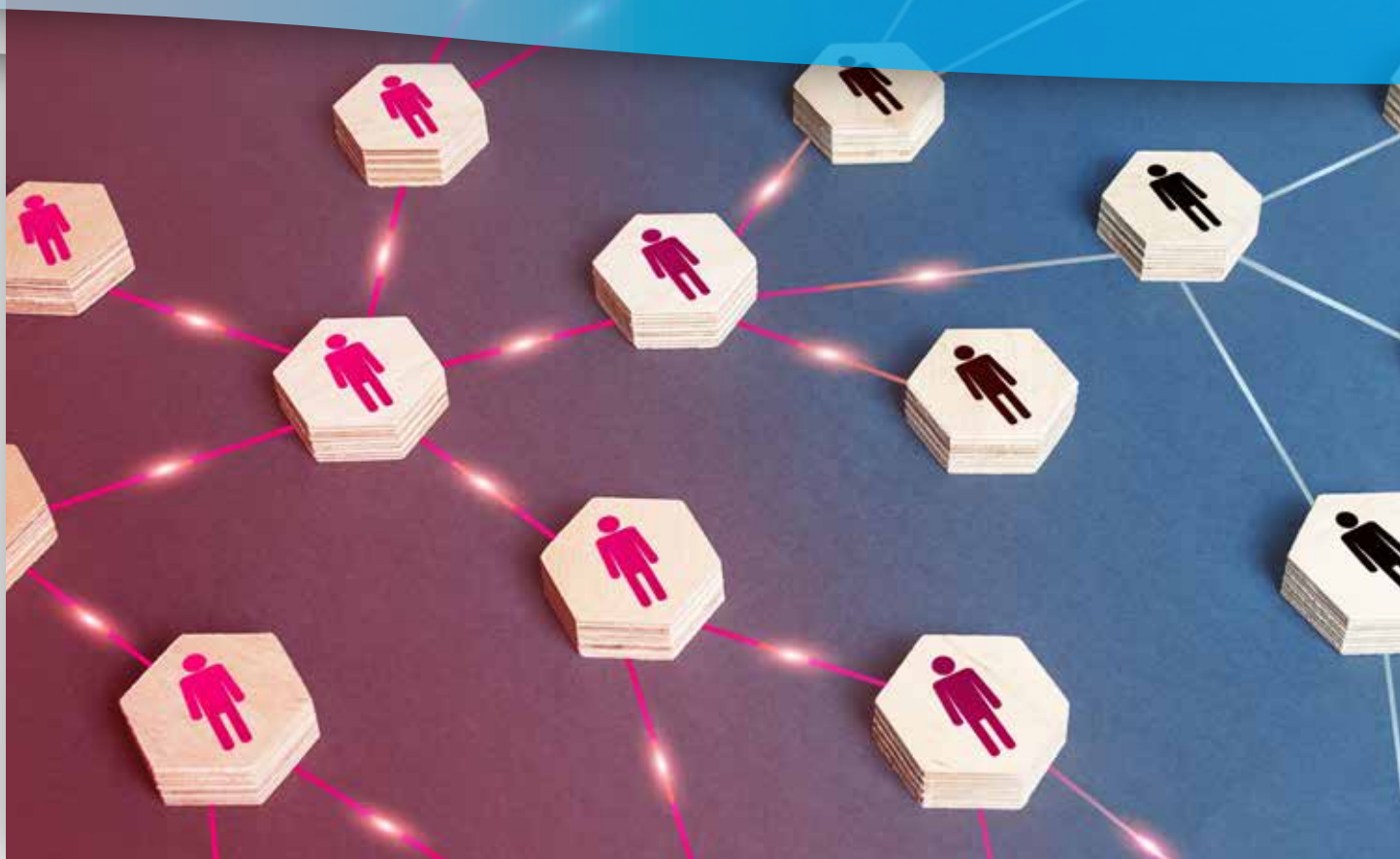
For inpatient healthcare facilities like hospitals and rehabilitation clinics, outbreaks – particularly those caused by the spread of multidrug-resistant bacteria (MDRB) – have been a major topic in recent years (David *et al.*, 2019). This has continued to be the case throughout the COVID-19 pandemic. An MDRB outbreak poses enormous challenges for all health care workers. In addition, such outbreaks must be reported to the relevant public health authority, regardless of where they occur. Judging by the overview of all reported outbreaks presented by the Robert Koch Institute in its annual report, it seems that the number of outbreaks in hospitals is significantly underestimated. We are therefore developing an IT solution – a kind of smart contact tracing system – that can provide early warning of potential outbreaks and perhaps prevent them from occurring in the first place.

As stated above Outbreaks can be caused by MDRB – pathogens like methicillin-resistant *Staphylococcus aureus* (MRSA) or, increasingly in recent years, the group of multidrug-resistant Gram-negative bacteria or vancomycin-resistant enterococci

(VRE). These MDRB represent a problem in patient care for many reasons – on a global level, but also locally in Germany and Europe. From an infection medicine standpoint, the main clinical feature of bacterial multidrug resistance is the loss of efficacy of established, highly effective, and well-tolerated antibiotics in treating an infection. As a result, MDRB infections are often associated with poorer clinical outcomes and increased costs for the health system – as, for example, less effective and more expensive “reserve” antibiotics have to be used, or adequate therapy is delayed. If an infection occurs in a hospital setting (e.g., as a result of medically necessary procedures like surgeries or the insertion of central vascular catheters), it is referred to as a “nosocomial infection”. In Germany, the majority of nosocomial infections are not caused by MDRB. Often, MRB detections are merely so-called colonizations with these pathogens on the mucous membranes, e.g. in the intestine (especially with MRGN and VRE) and in the nasopharynx (MRSA).. Due to a lack of symptoms, these colonizations are not always promptly detected, which can lead to the spread of MDRB pathogens. However, colonization with multidrug-resistant pathogens can also develop into an infection that requires treatment.

The Central aim of the Infection Control use case

In the Infection Control use case (<https://www.highmed.org/highmed-use-case-infection-control>) of the HiGHmed consortium (<https://www.highmed.org/>), we have researched and developed a system for the early detection of nosocomial transmission chains – an automated outbreak detection system. A digital solution is being developed that can automatically detect and visualize the clustered spread of multidrug-resistant bacterial pathogens and, perhaps in the future, susceptible bacterial pathogens (Baumgartl *et al.*, 2021). The idea is that



Digitalization helps detect pathogen spread at an early stage and break transmission chains.

(Photo: shutterstock © Andrii Yalanskyi)

by recognizing an infection chain at a very early stage, an outbreak can be minimized or even – in some cases completely avoided – by the appropriate measures being implemented to break this chain early on, thus protecting patients, preventing infections, reducing costs, and allowing resources to be used more wisely.

A technical solution and platform

The algorithm-based Smart Infection Control System, or SmICS (see Figure 1), enables the automated detection of outbreaks. Our system utilizes data that is already collected as part of routine care. The HiGHmed consortium is working in close cooperation with university and non-university institutions to set up so-called medical data integration centers (MeDICs; *gesundhyte.de* featured an article on this topic in Issue 13, July 2021) – initially at German university hospitals. These centers fulfill a variety of purposes in the project. On the one hand, they serve as communication nodes, allowing the participating locations to network with one another. On the other, they function as local repositories and integration sites for the administrative and care-relevant data that is generated at each respective site during treatment (Haarbrandt *et al.*, 2018). Particularly important for the detection of outbreaks are movement data (i.e., the whereabouts of patients in the hospital) and microbiological

findings. Linking these two data sources is a core element of SmICS, as it allows to determine the specific temporal and local context in which pathogens are detected. Using algorithms and standardized – e.g., threshold-based – queries, the system can then automatically detect potential clusters or outbreaks.

At the MeDICs, all relevant data (e.g., microbiological findings) (Wulff *et al.*, 2021) are modeled in full granularity using an open standard for the representation and storage of medical data (openEHR) (Beale, 2022), and are thus available via the MeDICs data platform for complex queries from the field of healthcare and research. We are also currently implementing a system that takes into account molecular typing results from microbiological findings – i.e., the exact breakdown of whether it is the identical pathogen. Robust and consistent modeling with the openEHR standard (semantic interoperability) allows for the integration of source data from different systems – including commercial laboratory information systems – to create a harmonized, standardized data platform that is accessible via open source interfaces. From the very beginning, we have placed great importance on uniform data management and the consistent use of standardized query options and interfaces, so the SmICS application can be used at all participating sites without the need for significant programming modifications.

Improving infection prevention

The early detection of potential outbreaks is of central importance, as it functions as an “alert system” that can limit the number of patients affected. In addition, an automated system has the advantage of being able to detect transmission chains that are difficult to identify using manual tracking methods, and thus often remain undetected. Manual outbreak detection – which is not automated but may, in a best-case scenario, be supported by digital tools – is also very time-consuming and ties up many of the hygienists’ resources. Time, for example, can be spent much more effectively verifying or falsifying the alerts generated by SmICS. SmICS can therefore help significantly reduce the number of previously undetected and/or unreported outbreaks, protect patients from infection, and ensure that personnel resources in infection prevention are put to more effective use. It must be taken into account that there will be an increase in the number of suspected cases, which must then be analyzed by qualified experts. Overall, this results in a higher quality infection prevention system with no extra strain on resources.

Looking ahead

Established as part of the Medical Informatics Initiative, the HiGHmed project has used concrete use cases to create an open IT infrastructure (platform) that is rigorously built according to international standards and that has many possible applications in networked medical care. In our initial use case Infection Control, we focused on creating a system for smart, automated outbreak detection (SmICS), which we developed from 2018 to 2021. Other use cases from the field of infection prevention, such as the automated surveillance of nosocomial infections or syndromic surveillance (which is based on clinical symptoms rather than pathogen detection), would be a few examples of logical further steps. Particularly if additional healthcare data – such as drugs administered, vital signs, or clinical chemical laboratory results – were to be incorporated into the system, a multitude of possible further use cases would emerge. The technical scalability and portability of the platform have been demonstrated, for example, in the development of a tool for the surveillance of COVID-19 infections in healthcare facilities as part of the B-FAST project (<https://www.umg.eu/forschung/corona-forschung/num/b-fast/>) of the German Network of University Medicine (NUM).

Use Case Infection Control Study Group:

University Medical Center Göttingen (UMG):

Simone Scheithauer, Martin Kaase, Patrick Fehling, Sabine Rey, Markus Suhr

Heidelberg University Hospital:

Vanessa M. Eichel, Nico T. Mutters, Klaus Heeg, Angela Merzweiler

Hannover Medical School:

Dirk Schlüter, Claas Baier, Michael Marscholke, Antje Wulff, Pascal Biermann

Charité – Universitätsmedizin Berlin:

Petra Gastmeier, Michael Behnke, Luis Alberto Peña Diaz, Sylvia Thun, Roland Eils

University of Münster (WWU)/

University Medical Centre of Münster (UKM):

Alexander Mellmann, Hauke Tönnies, Martin Dugas, Michael Storck

Technical University of Darmstadt:

Tatiana von Landesberger, Tom Baumgartl

University Hospital Schleswig-Holstein (UKSH):

Benjamin Gebel, Cora Drenkhahn

Helmholtz Centre for Infection Research (HZI):

Thorsten Klingen, Stephan Glöckner

Robert Koch Institute:

Benedikt Zacher, Tim Eckmanns

NEC Laboratories Europe GmbH:

Timo Sztyler, Brandon Malone

About the research project

The HiGHmed consortium (www.highmed.org), which is funded by the German Federal Ministry of Education and Research as part of the Medical Informatics Initiative, is developing interoperable solutions in medical informatics with the goal of making patient data available in digital form for clinical research and education. This should also strengthen clinical patient care in a long-term sustainable manner. By establishing secure data integration centers at the participating university sites, the project aims to provide scientific and medical staff with precisely the digital resources they need for making decisions. In addition to the university sites, many other public and private institutions are participating in the HiGHmed consortium. The Infection Control use case focuses on developing and studying a digital system for the early detection of outbreaks.

References:

David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, *et al.* Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. *Nat Microbiol.* 2019;4:1919–29. doi:10.1038/s41564-019-0492-8.

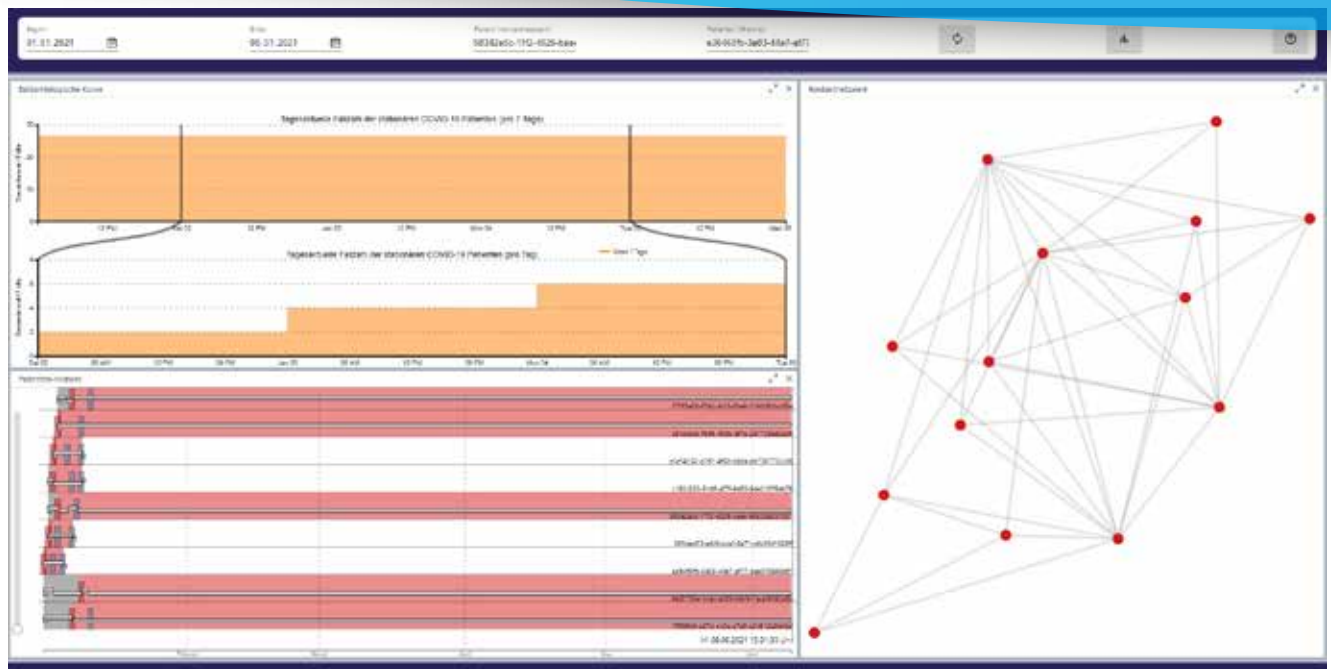


Figure 1: Knowledge of patient movement, microbiological data on pathogens, and contact network information are central to better understanding cluster situations. SmICS provides critical assistance with visualization (shown here on the example of COVID-19) (Source: HIGHmed, Infection Control use case).

Baumgartl T, Petzold M, Wunderlich M, Hohn M, Archambault D, Lieser M, *et al.* In Search of Patient Zero: Visual Analytics of Pathogen Transmission Pathways in Hospitals. *IEEE Trans Vis Comput Graph.* 2021;27:711–21. doi:10.1109/TVCG.2020.3030437.

Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, *et al.* HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. *Methods Inf Med.* 2018;57 S 01:e66–81. doi:10.3414/ME18-02-0002.

Beale T. Archetypes: Constraint-based Domain Models for Future-proof Information Systems, Revision: 2.2.1. In: Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer. Seattle, Washington, USA. Northeastern University, Boston; 2002. p. 16–32.

Wulff A, Baier C, Ballout S, Tute E, Sommer KK, Kaase M, *et al.* Transformation of microbiology data into a standardised data representation using OpenEHR. *Sci Rep.* 2021;11:10556. doi:10.1038/s41598-021-89796-y.

Contact:



Dr. med. Claas Baier
Institute for Medical Microbiology and
Hospital Epidemiology
Hannover Medical School
baier.claas@mh-hannover.de



Dr. rer. nat. Antje Wulff
Peter L. Reichertz Institute for Medical
Informatics of TU Braunschweig and
Hannover Medical School
wulff.antje@mh-hannover.de



Univ.-Prof. Michael Marschollek
Executive Director, Director Campus
Hannover
Peter L. Reichertz Institute for Medical
Informatics of TU Braunschweig and
Hannover Medical School, Germany
Co-Leader of Use Case Infection Control
within the HiGHmed Project
marschollek.michael@mh-hannover.de



Univ.-Prof. Simone Scheithauer
Head of the Department for Infection
Control & Infectious Diseases
University Medical Center Göttingen
(UMG), Institute of Infection Control and
Infectious Diseases
simone.scheithauer@med.uni-goettingen.de

www.highmed.org

national research data infrastructure NFDI

Working together to build a FAIR research data management in Germany

by Nathalie Hartl

Every day, new research data is being amassed to help us gain a better understanding of the world, solve problems, and develop innovations. A wealth of insights also lies dormant in existing data sets, waiting to be accessed through new connections or research questions. All too often, however, data are currently not sufficiently standardized and only available with limitations or on a temporary basis. This is due in part to the lack of sustainable data management integration. As a result, the potential for subsequent data use remains untapped. The German National Research Data Infrastructure (NFDI) addresses this issue and aims to make research data management in Germany more effective.

NFDI pursues the vision of making relevant research data available in line with the FAIR Guiding Principles for scientific data management and stewardship (Wilkinson *et al.*, 2016). (The acronym FAIR stands for Findable, Accessible, Interoperable, and Reusable.) To meet this objective as well as the multiple requirements of diverse disciplines, NFDI is set up as a nationwide, distributed, and growing network. Currently, it consists of nineteen consortia – associations of universities and research institutions within one discipline – funded by the German federal and state governments.

From recommendation to association

The nonprofit association National Research Data Infrastructure (NFDI) e.V., based in Karlsruhe, was established as a joint body that allows the consortia to align their efforts and contribute to the development of a national research data infrastructure. The association is where everything comes together. It coordinates the consortia's interactions and addresses thematic and strategic issues that concern both the future of research data management in Germany and international connections.

In a paper published in 2016 (<https://rfii.de/?p=2075>), the German Council for Scientific Information Infrastructures (RfII) advocated establishing a coordinated research data infrastructure. To build a digital knowledge repository, the authors suggested, existing structures need to be interconnected more effectively and new potentials tapped.

The RfII proposal met with wide acceptance in German research communities. Both the federal and state governments endorsed the recommendation, and in November 2018 agreed to establish and fund NFDI. The long-term goal is to strengthen Germany as a science hub by providing better research data management, and to reap societal benefits from new insights gained from existing data resources.

Up to 30 consortia are to be funded under NFDI. The German Research Foundation (DFG) manages the selection process, with the DFG's NFDI Expert Committee evaluating proposals and providing recommendations to the Joint Science Conference (GWK), which then decides on the funding.



Opening of NFDI office on October 15, 2020, in Karlsruhe

Pictured, from left to right: Sabine Brünger-Weilandt (FIZ Karlsruhe), Holger Hanselka (KIT), Eva Lübke (NFDI Directorate), York Sure-Vetter (NFDI Directorate), Frank Mentrup (Mayor of Karlsruhe) (Photo: Cynthia Ruf/KIT).

Effective research data management as the key to success

The consortia are united by their common goal of building an optimized data infrastructure that will allow for better processing of specific research questions. The following three examples of NFDI consortia serve to illustrate NFDI's significance for health research.

NFDI4Health, the National Research Data Infrastructure for Personal Health Data, focuses on data generated in clinical trials as well as epidemiological and public health studies. The consortium aims to increase the effectiveness and quality of medical research and contribute to improving public health. Projects include increasing the effectiveness of COVID-19 pandemic research. Participating institutions are working on establishing a process that will improve the way that health data from different studies and databases are linked, leading to accelerated insights on infection pathways and disease progression. The consortium also addresses privacy and data protection issues in connection with the exchange and linking of sensitive personal data.

The GHGA (German Human Genome-Phenome Archive) consortium is another example of the relevance of a function-

ing research data infrastructure. As more and more genome data are being collected to enable improved diagnostics and personalized treatments, linking as many local data sets as possible and mining them with the help of AI can contribute significantly to research on rare genetic diseases. In the future, oncologists should be able to compare observed tumor-specific genetic changes with similar cases in the GHGA database. To access the most comprehensive collection of data sets possible, GHGA is networking with the European Genome Archive (EGA).

A third of many possible examples is a consortium in the field of chemistry. NFDI4Chem is leveraging effective research data management in combination with machine learning methods in order to accelerate the search for therapeutic agents and facilitate meta-studies across families

Challenges of making data available

Improving data findability, accessibility, and long-term archiving, as envisaged by NFDI, does not mean that data will be available to everyone at all times, without access restrictions. Strong data protection is particularly relevant for sensitive, personal data (as collected in medicine and social sciences); proprietary and business data (e.g., in the context of

technical innovations); and intellectual property data (e.g., in the arts and cultural sciences). NFDI therefore also addresses legal issues affecting all consortia. The goal is open access to metadata and persistent identifiers that contain, among other things, information about the requirements for data retrieval, provenance, and user licenses. Efficient sharing of research data requires the use of uniform (machine-readable) formats, commonly defined underlying terminologies and ontologies, and the establishment of standards. This is precisely why NFDI follows a community approach.

Only if enough scientists and institutions integrate effective data management into their research process, applying shared standards, can data be reused in a sustainable way, and the potential of existing data sets be exploited. Establishing a reputation system may motivate researchers to address the issue of research data management and to share their data – always in compliance with legal requirements, of course. Sharing research data should be rewarded just as publishing articles is.

Strengthening data literacy among (early career) scientists is another approach. Relevant qualification and education pro-

grams should raise awareness of the enormous importance of high-quality research data management and strengthen skills in the area of data literacy.

The challenges shared by all NFDI consortia are to be addressed as cross-cutting topics in so-called cross-consortia sections. As a first step, four sections on “(Meta)data, Terminologies and Provenance”, “Common Infrastructures”, “Training & Education” and “Ethical, Legal & Social Aspects” were established in October 2021.

Projects at the European level

Big Data in research poses great challenges but offers even greater opportunities. NFDI wants to seize these opportunities and, together with hundreds of scientists, create standards for optimized research data management in Germany and beyond. Similar efforts are also underway at the European level. The European Union has launched the European Open Science Cloud (EOSC) project, in which various countries are working together to establish a cloud infrastructure. NFDI is involved in this undertaking.

THE ASSOCIATION

The NFDI Association was founded in October 2020. The **NFDI Directorate** is the Executive Board of the Association and is supported by the staff of the Karlsruhe-based managing office. The Association bodies include the **Board of Trustees** as an administrative-strategic supervisory body; the **Scientific Senate** as the substantive and strategic body; the **Consortia Assembly**, which determines the substantive and technical principles for the work of the consortia in a cross-disciplinary way; and the **Members Assembly**. (<https://www.nfdi.de/membership/?lang=en>)

Currently, the Association has 188 members (as of September 10, 2021), including the Federal Republic of Germany and the 16 federal states as founding members as well as numerous universities and scientific institutions. The extensive exchange among the consortia enables synergy effects to be achieved and new cooperation opportunities to be realized. Up to €90 million will be provided annually until 2028 to develop the infrastructure.



More and more research data are being collected. NFDI was created to make use of this vast treasure trove of data.
(Source: iStock © in future).

The joint project FAIR Data Spaces combines the European initiative Gaia-X, which involves both science and industry, and NFDI. Funded by the German Federal Ministry of Education and Research (BMBF), the project aims to establish a cloud-based FAIR Data Space for knowledge transfer between industry and science.

NFDI is active on the national and international levels to create solutions that are compatible with the global science system and provide synergies that drive new developments.

References:

Wilkinson, M. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In: Sci Data 3, Article number: 160018. <https://www.nature.com/articles/sdata201618> (last visited: 05/03/2022).

German Council for Scientific Information Infrastructures: Enhancing Research Data Management: Performance through Diversity. Recommendations regarding structures, processes, and financing for research data management in Germany, Göttingen 2016, 90 p. <https://rfii.de/download/rfii-recommendations-2016-performance-through-diversity/> (last visited: 05/03/2022)

Association statutes of the Association German National Data Infrastructure (Nationale Forschungsdateninfrastruktur (NFDI) e.V.: <https://www.nfdi.de/membership/?lang=en> (last visited 05/03/2022)

Contact:



Nathalie Hartl

Scientific Officer PR/Communication

Tel.: +49 721 988 994 20

[nathalie.hartl\[at\]nfdi.de](mailto:nathalie.hartl[at]nfdi.de)

www.nfdi.de/?lang=en

Photo: NFDI

THE CONSORTIA

The **nine consortia** DataPLANT, GHGA, KonsortSWD, NFDI4Bio-Diversity, NFDI4Cat, NFDI4Chem, NFDI4Culture, NFDI4Health and NFDI4Ing have been receiving funding since October 2020. In a second round, it was decided to fund a further **ten consortia** starting October 2021. BERD@NFDI, DAPHNE4NFDI, FAIRmat, MaRDI, NFDI4DataScience, NFDI4Earth, NFDI4Microbiota, NFDI-MatWerk, PUNCH4NFDI and Text+ will be integrated into the existing structures in the coming months.

GAIA-X as an enabler for a health data space

Launched in 2019, the Franco-German project GAIA-X is entering its third year. Here is a snapshot.

by Harald Wagener

GAIA-X stands for the strategy of building competitive digital ecosystems in Europe that enable secure and privacy-protecting applications based on the principles of cooperation and transparency. The aim is to provide a counterweight to the large hyperscalers from the United States and Asia and to lay the foundation for distributed and federated infrastructures that allow many independent participants across this ecosystem to create innovative and new ways of exploiting data without giving up sovereignty over that data.

How is GAIA-X organized?

Gaia-X is organized into several groups that overlap and complement each other: The **GAIA-X Association**, the official body, is a nonprofit organization under Belgian law. It has definitional authority over the official GAIA-X architecture and over the procedures and rules for participating in the GAIA-X ecosystem. The Association also provides an open-source reference implementation of the GAIA-X Federation Services, which acts as an interface between infrastructure and service providers on the one hand and users on the other.

The interests of the member states and the companies and institutions domiciled in them are bundled together in the **GAIA-X Hubs**. Here, the requirements for **Gaia-X** are consolidated on a regional level. Representatives of the **Gaia-X Hubs** comprise the Representative Board, which is in direct contact with the GAIA-X Association and thus ensures that the interests and requirements of the Hubs are taken into account in

the decisions and publications of the Association. There are also **Domain Working Groups** at the regional and European level, which in addition to the requirements for a federated infrastructure common to all domains, gather and forward specific requirements, such as in the area of health or mobility, to the Association based on concrete use cases.

And serving as the connecting link between these Groups is the **GAIA-X Community**, which complements the Association's subgroups by preparing proposals for technical and regulatory standards in work packages that form the basis for developing the GAIA-X ecosystem.

GAIA-X and the healthcare sector

In the healthcare sector, too, digital progress will only be feasible through cloud technologies. It will be necessary in the future to be able to link and process health data from various sources. On the one hand, high security and data protection requirements must be met, while on the other hand, solutions that have been largely isolated so far must be freed from their silo systems to enable better diagnostic procedures and individualized therapeutic approaches with integrative methods.

The health domain has been strongly represented in all areas of the GAIA-X project from the very beginning. The initial papers on GAIA-X were substantially co-conceived by representatives of the health domain. Moreover, the health domain was significantly involved in the differentiation of concrete requirements for a GAIA-X ecosystem, formulating more than 20 use cases in the early phase of GAIA-X. Here, it was possible to

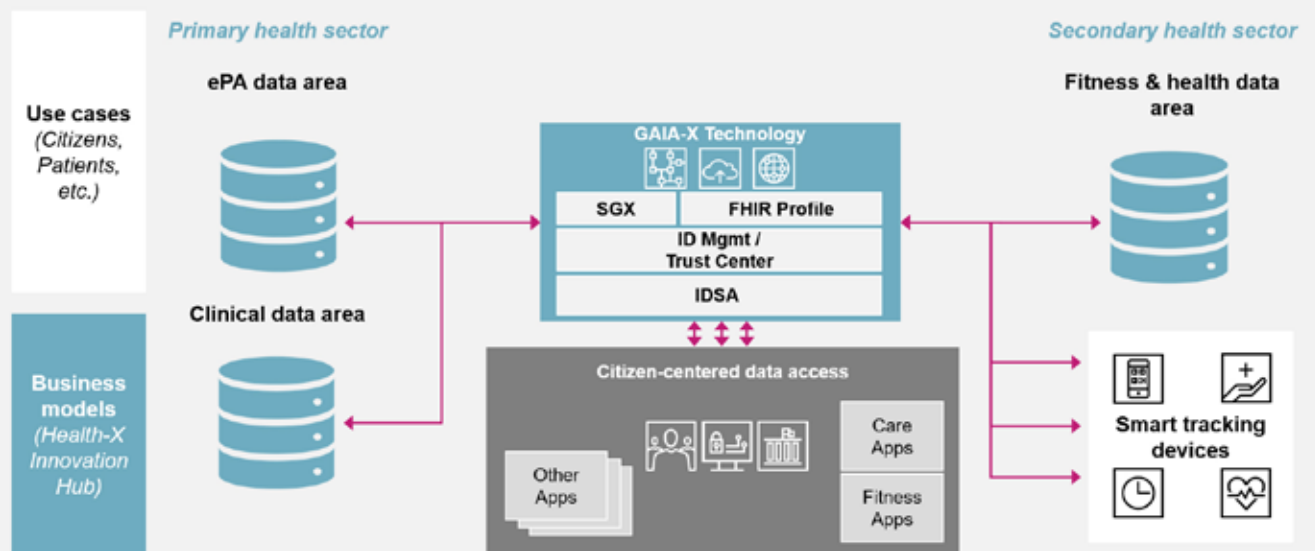


Figure 1: HEALTH-X dataLOFT's technical realization. (Source: BIH, LANGEundPFLANZ)

incorporate real-life experience with federated cloud systems like the de.NBI Cloud, ELIXIR, and the European Open Science Cloud (EOSC), could already practically demonstrate many aspects of possible implementation in the life sciences at the national and European levels. Today, the GAIA-X Association and the EOSC Association are in close exchange to enable even better collaboration going forward and to determine which interfaces are necessary to ensure interoperability between EOSC and GAIA-X.

GAIA-X in practice

With a few exceptions, the use cases of GAIA-X have so far been considered theoretically. In the MVG Working Group of the GAIA-X Community, sub-areas of the existing GAIA-X specifications are now being tested for practical applicability in pilots organized as hackathons. There are also initial smaller funded projects in the context of GAIA-X, such as Smart Health Connect, which addresses the real-time collection of health data from patients with heart disease. These data are evaluated using AI methods to help healthcare providers select individualized therapy steps. This project was also part of MVG's Pilot 003 Hackathon, which was held in August 2021 and focused on the distributed AI analysis of data.

In spring 2021, the German Federal Ministry for Economic Affairs and Climate Action (BMWK) held a funding competition aimed at identifying projects that are worthy of support as "beacon" projects for implementing Gaia-X in the various domains. One of the eleven winners of this competition is the HEALTH-X dataLOFT project under the consortium leadership of Charité, whose goal is to drive the transformation of the healthcare sector with citizens as active partners instead of passive recipients of services. Sovereignty concerning data usage lies in the hands of citizens via a data wallet. This ensures the secure and trustworthy use of data in the European Health Data Space. A second essential cornerstone of the HEALTH-X dataLOFT idea is the linking together of data from the primary and secondary healthcare market, for example, through a data donation mechanism. This generates ecosystems that offer modular platform services like AI analysis and enables the connection of self-tracking data to new diagnostic and therapeutic approaches. In short, the idea of data-based real-world evidence research becomes a usable reality, providing immediate benefits to citizens in terms of prevention and care.

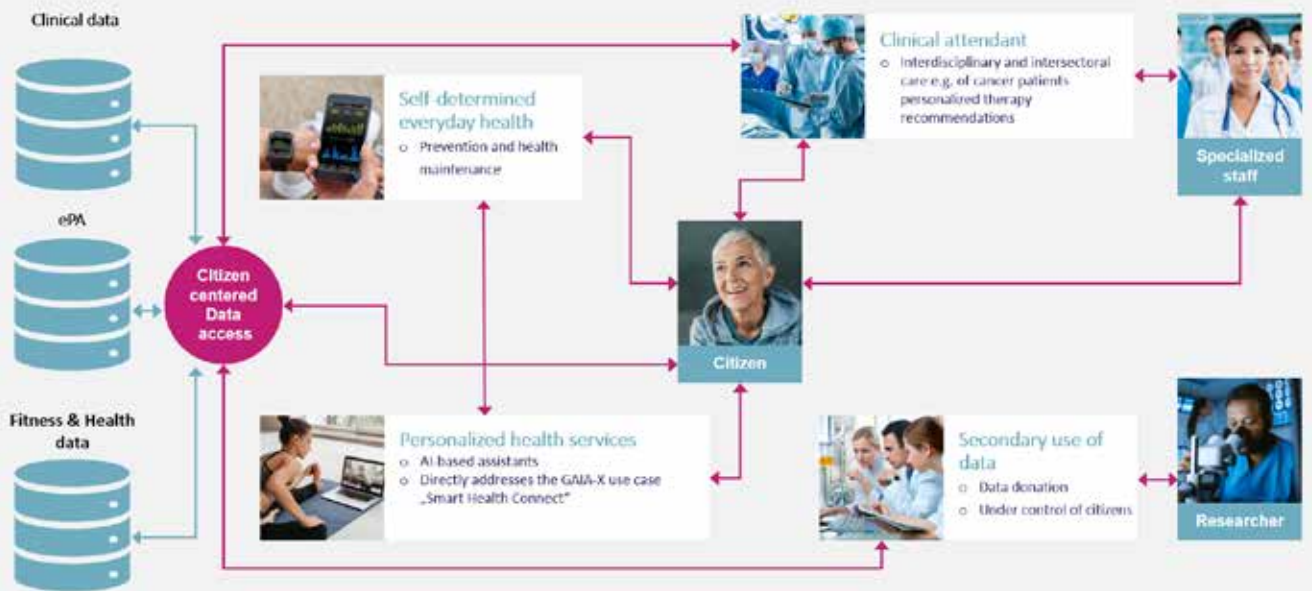


Figure 2: HEALTH-X dataLOFT provides a variety of use cases based on unified data sources, technologies, and processes.

(Source: BIH, LANGEundPFLANZ, shutterstock © Alexander Rath, shutterstock © Andrey Popov, shutterstock © Arsenii Palivoda, shutterstock © ESB Professional, shutterstock © Gorodenkoff, iStock © Ijubaphoto)

This will be demonstrated through various use cases from the thematic areas of Self-Determined Everyday Health, Clinical Companions, Personalized Health Services, and Secondary Use of Data. These use cases bring together data from clinical care, electronic patient records (ePA), and self-collected fitness and health data via citizen-centered data access, so that citizens can receive the best possible data-based healthcare in every phase of life in a self-determined way according to the use cases.

Contact:



Harald Wagener

Center for digital Health
Group leader Cloud
BIH at Charité, Berlin
Harald.Wagener@charite.de

www.hidih.org/research/health-data

GAIA-X IN PRACTICE: FEDERATION SERVICES AND COMMUNITY

The Gaia-X Association has produced a short video introducing the Federation Services:

https://www.youtube.com/watch?v=68tZh_wkeYY

and detailed information is available at <https://www.gxfs.eu>

Community driven development of Gaia-X on GitLab:

<https://gitlab.com/gaia-x/gaia-x-community/>

HEALTH-X dataLOFT website:

<https://www.health-x.org/>

FAIR data: experiences from collaborative research projects

Requirements and step-by-step implementation for data-driven collaborations

by Matthias Ganzinger and Matthieu-P. Schapranow

Once funding for a new collaborative research project has been approved, researchers are eager to get started. However, making data available for joint use requires the right approach to prevent pitfalls. Based on real-world observations, the following overview covers the most important aspects to consider when beginning a data-driven collaborative project.

Sharing data is a prerequisite for many research projects, e.g. because partner institutions contribute their respective existing data or new data are collected for the project and made available to all partners and the scientific community. Typically, data management starts at the same time as all project activities. Consequently, numerous questions have to be answered already in the early project stages to ensure that data are made available in a proper way to all involved partners.

This is precisely the focus of the project group “Data Management and Bioinformatics”, which was founded in 2014 as part of the e:Med research program to establish systems medicine. The project group offers workshops in which designated data managers can share their experiences and approaches to data management in collaborative projects. Over the course of these workshops, members of the project group have identified the challenges presented in Figure 1. Since these were relevant across numerous projects, it seemed sensible to address them together.

In the following, we share details about the integration of multiple data sources and the application of FAIR principles to exploit them. Further details can be found in the chapter on “Biomedical and Clinical Research Data Management” in the text book “Systems Medicine: Integrative, Qualitative and Computational Approaches”, which provides a more in-depth discussion of the group’s work.

Integrating heterogeneous data sources

Real-world data build the foundation for biomedical research, as they comprehensively reflect reality and often deviate from synthetic or simulated data. However, the use of real data requires that privacy concerns are addressed. Therefore, it is advisable to obtain approval for the research project from relevant ethics committees and institutional review boards prior to use of data. In case you plan to use prospective data, i.e., data that needs to be acquired in future, it is mandatory to be compliant with the EU’s General Data Protection Regulation (GDPR) and Good Clinical Practices (GCP) from the beginning on. For example, inform individuals – legally also called data subjects – about their rights and obtain their consent for the acquisition and use of their data.

Occasionally, additional data use scenarios for already collected data arise in the course of a project. In this case, it might neither be possible to inform data subjects nor to acquire their

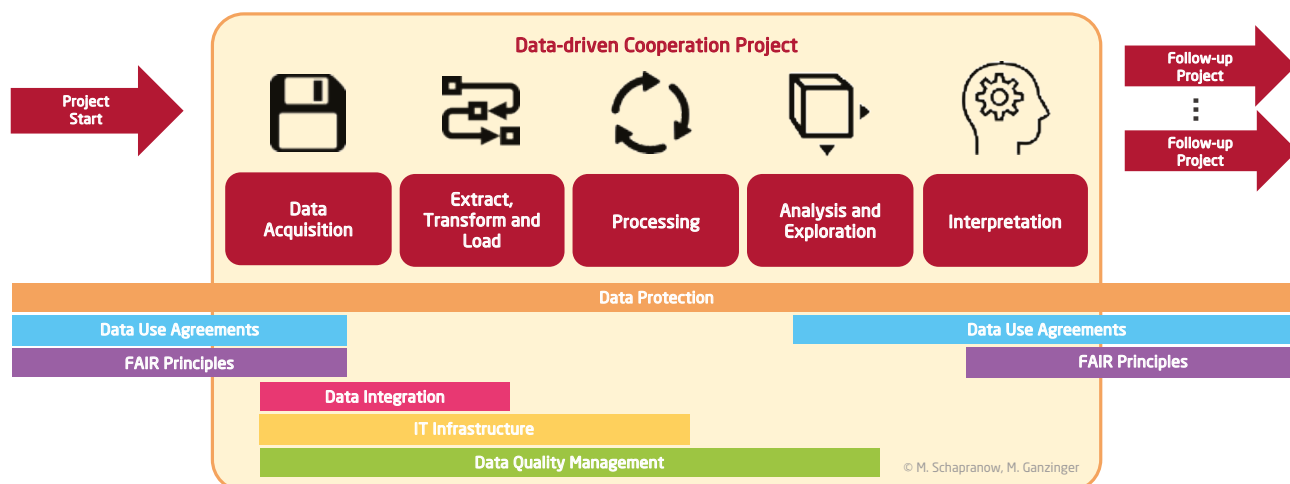


Figure 1: Challenges for data-driven collaborative projects from the project idea to the use of the results by follow-up projects: data protection, data usage rules, FAIR principles, data integration, project internal IT infrastructure, and data quality management (Source: M.-P. Schapranow / M. Ganzinger).

consent. If it is not possible to obtain a new consent for the particular use (in the case of retrospective data from another study, e.g.), data anonymization methods need be applied. They are used to randomize or remove personal identifiers from the data sets so that it is no longer possible to link them to individuals. For most research projects, the anonymized data set will be sufficient as they typically consider patient cohorts instead of individual patients. As a rule, researchers should focus on minimal data use, i.e., only collect or extract the data necessary for your specific research question. So, when in doubt, the less data the better.

Data sharing and FAIR principles

For numerous publicly funded projects, making results available to follow-up projects is already a requirement. Over the past few years, the FAIR guiding principles for scientific data management and stewardship have become a de facto standard in the scientific community. The acronym **FAIR** stands for:

- **F**indable: metadata and data should be findable, e.g., via search engines or data repositories, and uniquely identifiable, e.g., via Digital Object Identifiers (DOI),

- **A**ccessible: access to metadata and data should be possible without barriers,
- **I**nteroperable: data and metadata should conform to standardized definitions in order to be interoperable and able to be merged with data from other sources, and
- **R**eusable: data should be accompanied by extensive, well-described metadata.

Data management step by step

Real-world experiences have shown that the integration of different heterogeneous data sources represents a major challenge in collaborative research. Figure 1 shows the individual process steps for data management in data-driven collaborative projects.

The first step is **data acquisition**. This includes both the identification of existing data sources when already available data are being used and the definition of a data acquisition process whenever additional or new data need to be collected. Amongst others, it might include the selection of relevant sensors and measuring devices as well as the definition of quality

criteria, e.g., measurement accuracy. In biomedical research, processing of personal data is often unavoidable. All processing of personal data must follow required data protection measures as prescribed by corresponding privacy laws such as the EU's GDPR.

The second step involves **extracting** data from their source or the measuring device respectively, **transforming** them into well-defined data formats and data models, and **loading** the data into a data repository for the collaborative project. This so-called ETL process is a well-established concept for data warehousing and enterprise software. For data from heterogeneous sources without a uniform data model, the ETL process also includes data harmonization. The effort involved in data harmonization is often underestimated; it requires all participating disciplines to develop a common data model for the joint research partnership. Data from relational models, which can be easily represented in database tables, should be integrated into a shared database management system as this will not only ensure the integrity of the data but also provide standardized interfaces for data retrieval, e.g., Structured Query Language (SQL).

In the next step, data are **processed**. First, they are checked using suitable algorithms and prepared for analysis. If necessary, this may include steps to clean the data or interpolate missing data, depending on how the data are to be used specifically. Next, the data are **analyzed** and **explored** using data analysis methods to gain insights from the data. Examples are machine learning methods used to classify data semi-automatically or calculate missing entries from existing data. Finally, the analysis results are **interpreted** within the context of the intended use case. This should include the consultation of experts from the respective discipline, such as biology or medicine, to contextualize the findings within the present state of scientific knowledge.

Outlook: Federated learning

Federated IT infrastructures are an alternative option for data-driven collaborative research projects. In these federated architectures, services are provided by multiple research partners in a decentralized network instead of a single central system providing services to all the partners. However, the underlying complexity is hidden and users can access and use these services like in a centralized architecture model. Decentralized IT systems are often available already. Setting up and managing a new federated infrastructure offers potential savings in time and resources compared to central IT infrastructures. Then, data can remain at their point of origin in federated infrastructures, a major advantage especially in the case of sensitive personal data. In contrast to centralized approaches, data analysis algorithms are transferred to the data and executed locally on systems belonging to the data owners. Only aggregated results of the analysis are returned, making the transfer of potentially identifying data obsolete.

Let us consider a federated prognostic model based on AI to illustrate how a federated infrastructure could be used. The model's objective should be to identify high-risk patients after kidney transplantations in order to prevent tissue rejection or even organ failure. For this purpose, a clinical prognosis model V_1 is trained at clinic site A using local training data. Next, the trained model is sent to another clinic (site B), where data of another transplant-patient cohort is available. If the model V_1 is tested on site B data, the prognosis quality might be lower compared to site A because the model does not know all the special cases present at site B, yet. If you were to use transfer learning and optimize the model V_1 with data from site B, you will obtain a new model version V_2 . This procedure can be repeated with an arbitrary number of sites and model versions in the hope that the forecast quality will improve continuously thanks to the additional training data available.

Conclusion

Today, a number of data management approaches for data-driven collaboration projects are available. However, none of them is a one-size-fits-all solution. The selection of the right approach should be guided by the specific project requirements. Regardless of the chosen method, collaboration beyond one's own project network offers valuable insights regarding both the use of one's own results and data management in follow-up projects.

Project profile:

The project group (PG) **Data Management and Bioinformatics** was founded in 2014 as part of the e:Med research program aimed at establishing systems medicine. The PG facilitates cross-project exchange on the topic of data management and on the joint use of standardized procedures to optimize the utilization of data both during the project and after its completion.

References:

Ganzinger M, Glaab E, Kerssemakers J, Nahnsen S, Sax U, Schaad NS, Schapranow M-P, Tiede T. Biomedical and Clinical Research Data Management. In: Wolkenhauer O, editor. Systems Medicine: Integrative, qualitative and computational approaches. Elsevier Academic Press; 2021. p. 532–43.

E:MED PROJECT GROUPS – CROSS-SECTIONAL TOPICS IN FOCUS



The e:Med PGs on substantive and methodological cross-sectional topics in systems medicine offer the opportunity for content-driven networking and an open, direct exchange of ideas. The PGs are independently formed by the scientific community following calls issued by the e:Med Project Committee. Each PG is led by scientists, usually as a team, and participation is open to all e:Med scientists.

The aims of the thematically focused PGs are to:

- 1 Discuss current scientific content and technologies from the field of systems medicine across research partnerships and, if necessary, launch specific initiatives, and
- 2 Enable content-driven networking and interaction among members and ensure access to and exchange of scientifically relevant information.

Thematic focuses of the PGs:

Data management and bioinformatics, data security and ethics, image processing, multiscale sequence-based analysis tools in systems medicine, MS- and NMR-based omics, modeling of disease processes, *in vitro* and *in vivo* model systems

The PGs are an example of the added value of networked interdisciplinary research made possible by e:Med, the BMBF's research and funding initiative to establish systems medicine in Germany.

www.sys-med.de/en



Sharing data is a prerequisite in many research projects, whether because partner institutions contribute existing data or because new data are collected for the project. Symbolic photo of a group of scientists jointly analyzing project data (Photo: shutterstock © puhhha).

Schapranow M-P, Perscheid C, Wachsmann A, Siegert A, Bock C, Horschig F, Liedke F, Brauer J, Plattner H. A Federated In-memory Database System for Life Sciences. In: Castellanos M, Chrysanthos PK, Pelechrinis K, editors. Real-Time Business Intelligence and Analytics, Springer International Publishing, 2019, ISBN: 978-3-030-24123-0

Contact:



Dr. Matthias Ganzinger
Institute of Medical Informatics
Heidelberg University Hospital
matthias.ganzinger@med.uni-heidelberg.de

<https://www.klinikum.uni-heidelberg.de/personen/dr-sc-hum-matthias-ganzinger-3732>



Dr.-Ing. Matthieu-P. Schapranow
Dr.-Ing. Matthieu-P. Schapranow
Working Group Leader
“In-Memory Computing for Digital Health”
Scientific Manager Digital Health Innovations
HPI Digital Health Center,
Hasso Plattner Institute for Digital
Engineering gGmbH,
University of Potsdam
schapranow@hpi.de

<https://hpi.de/digital-health-center/members/working-group-in-memory-computing-for-digital-health/dr-ing-matthieu-p-schapranow.html>



News from the BMBF

Faster detection of antibiotic resistance

More than 50,000 people are infected with antibiotic-resistant germs in Germany every year. And resistance is on the rise. Antibiotics, which have been one of the strongest weapons in the fight against infectious diseases, are therefore on the verge of becoming ineffective. The World Health Organization (WHO) estimates that ten million people will be dying due to resistant germs every year in 30 years if this trend continues unimpeded. However, it is sometimes not easy to detect antibiotic resistance. That is why patients are often treated with ineffective antibiotics in the first instance. "With severe cases in particular, we can lose precious time and the dangerous germs infect even more people", explains Professor Alexander Goesmann of Justus Liebig University Giessen.

He and his team want to use machine learning to detect resistance more rapidly and to identify targets for new antibiotics in the long term. The Federal Ministry of Education and Research (BMBF) is funding the interdisciplinary team from (the universities of) Giessen and Marburg with approx. one million euros under its "Computational Life Sciences" funding measure.

Resistances are complex

A rapid and secure identification of multiresistant pathogens is essential when it comes to helping infected patients effectively and preventing their spread

to more people. Modern genetic analyses can quickly provide clarity if a single gene conveys the resistance. "Yet resistances are often much more complex. There are different gradations, and often further genes are involved", explains Goesmann. Individual mutations can therefore be harmless but become dangerous in combination with other mutations. The research team relies on machine learning methods to identify these complex interrelationships. "The algorithms can analyse huge data sets and detect inherent relationships", says Professor Dominik Heider of Philipps University Marburg. The ultimate aim of the scientists is to enable the development of a rapid and simple test for such complex resistance mechanisms.

For the purpose of collecting data, Professor Goesmann's team collaborates with the Department for Medical Microbiology in Giessen under its director Professor Trinad Chakraborty. His team regularly takes samples from patients populated with multiresistant pathogens and studies them using molecular biology methods. Besides the genome sequences of bacteria, his research team is mainly interested in the reaction of bacteria to antibiotics. In order to study this, they examine which genes the bacteria switch on and off after the administration of antibiotics, but also how bacterial growth changes under given concentrations of different antibiotics. To do this, researchers let bacteria grow under standardized conditions in the lab, treat them with a predetermined dose of antibiotics at set points in time and analyse how bacterial growth changes.

Approaches to new antibiotics

With the help of algorithms, a computer programme then uses this data to learn what genes must be present and switched on for bacteria to continue to



**Antibiotic resistance is a major threat.
In future, AI will help hospitals to detect
it more rapidly and to take the appropriate
measures.**

Source: Adobe Stock / eremit08

grow despite the administration of a certain group of antibiotics. Despite the complex patterns it is possible to rapidly classify the bacterial strain at hand with regard to its resistance to certain antibiotics. For later practical use the bioinformaticians also want to understand why the programme takes a specific decision. Ultimately, the research is about finding what has changed at molecular and genetic level in the bacteria. "How dangerous bacteria are is not only determined by resistance mechanisms but also by other properties, such as their potential to spread."

In future, computer-aided analysis could enable faster and more reliable diagnosis than at present. Shortly after a patient's admission to hospital the attending physicians will then receive an assessment if one or more of the pathogens are resistant, if effective antibiotics are available and if the patient needs to be isolated. The research team will have to further validate and certify the procedure to facilitate safe application for patients. "We assume that we will be ready in five to ten years", Goesmann says. "In addition we hope that our analyses will provide starting points for novel antibiotics and that we can contribute to the development of new weapons in the struggle against multiresistant pathogens."

Further information is available at:

YouTube video on the project:
youtube.com/watch?v=UCv8QEKwXts



Better care based on digital innovation

The digital future of medicine has started already: Wearable sensors, for example, can record vital data on the heart health of patients with cardiac insufficiency even in their home environment and can communicate the data directly to their hospital. Indications of impending critical developments can thus be recognized at an early stage by physicians who can then take targeted countermeasures. Intelligent smartphone apps have long supported physicians in taking the best possible therapeutic decisions including in emergencies. An important pacemaker for such innovation in Germany is the medical informatics initiative (MII) of the Federal Ministry of Education and Research (BMBF). The MII has set up pathbreaking data infrastructure in four consortia which involve

all German university hospitals and has developed innovative IT solutions which have already improved patient care in many university hospitals.

Transfer models for the medical care of the future

Most people first consult their family doctor or a local hospital rather than a university hospital for their medical care. The pioneering work of university hospitals in the digital transformation of medicine must therefore become part of as many areas of the health system as possible, from out-patient care in GP practices to in-patient care in local hospitals to care in nursing and rehab facilities. This is a challenging task. Model solutions will have to be developed and optimized for everyone to benefit in future – which is the task of the six BMBF funded Digital Progress Hubs for the Health Sector. The BMBF is providing about 50 million euros for this flagship initiative of its Digital Strategy until 2025. The Progress Hubs also contribute to the Federal Government's Artificial Intelligence Strategy.

In Focus: Innovation that benefits people in many diverse ways

The Digital Progress Hubs for the Health Sector focus on applications in which data exchange between regional players, university hospitals and research institutions benefits a large number of people. Several Progress Hubs for instance address cancer treatment and pursue the objectives of the National Decade against Cancer which the BMBF initiated in 2019. The management of pandemics is also part of the range of topics covered by the Progress Hubs. The solutions they develop and test are expected to contribute to preparing the health system even better for future crises. As diverse as the medical topics are the technical solutions that are being used. This is highlighted by the following examples:

Platforms for the exchange of data are aimed at improving the study, diagnostics and therapy of diseases by means of intelligent networking and exchange of information – namely between hospitals, out-patient care, rehab and nursing care. At the same time the communication between therapists in different professions – including medicine, psychology and physiotherapy – and between therapists and their patients is to be promoted. The exchange and availability of data is an important precondition for the personalized treatment of patients and also – for instance in the case of cancer – for the best possible long-term care from the hospital to rehabilitation and beyond.

Artificial Intelligence is expected to help physicians for example in critical time-sensitive situations to rapidly take the right measures and save lives – e.g. when tending to a stroke victim during an emergency call-out.

Telemedicine and mobile sensors which transmit health data to treating physicians will also improve health care in rural areas. The data recorded will also help to better understand common diseases such as mental illness and cancer and to optimize therapies.

How Digital Progress Hubs work

The starting point for a Hub is the data integration centre of a university hospital. These centres are currently being set up as part of the IT infrastructure at almost all university hospitals in Germany under the medical informatics initiative. They network with regional partners – including hospitals, medical practices, rehab centres, care institutions and also emergency services. Research institutions and health insurance providers are also partners of the hubs. All partners share and make joint use of their data. This involves close collaboration between science, IT, health care providers, physicians, nursing staff and patients.

Further information is available at:

medizininformatik-initiative.de



AI improves COPD therapy

People often ignore early symptoms of COPD like coughing and shortness of breath when climbing stairs. However, Chronic Obstructive Pulmonary Disease is a serious illness. If it is not treated, irreparable lung damage results and the disease progressively worsens. During the later stages, many patients suffer from a shortage of breath even if they are just sitting on the sofa. The symptoms can be reduced by medication and the worsening of the condition can at least be decelerated.

However, there is still no treatment to stop the progression of the disease or even reverse it. A major problem is that it develops differently with different patients. Further impairments such as diabetes or osteoporosis play a significant role. Depending on the comorbidities and the health status of patients, different therapeutic approaches are needed. This is the starting point for the team of Marburg pulmonary specialist Professor Bernd Schmeck. They want to better classify patients by means of methods based on machine learning and thus enable personalized treatment. The Federal Ministry of Education and Research (BMBF) supports them under its ERACoSys-Med funding measure with almost 600,000 euros.

Schmeck's team uses data from several thousand patients. His Marburg colleague Professor Claus Vogelmeier was able to build up a patient study group with data from over 2,700 patients over a period of more than five years. Besides clinical data on disease progression, the medical experts also record more specific data from laboratory tests such as certain proteins and molecular markers in order to learn more about the disease. "We need an AI approach because of the sheer volume of data – people can no longer process these vast amounts of data without AI", explains Schmeck.

Keeping an eye on comorbidities

Based on thousands of combinations of lab data and disease progressions, the algorithm employed learns



Data helping to cure people. Digital innovation is expected to help not only university hospitals but also neighbourhood medical practices, local hospitals and care and rehabilitation institutions to improve patient care.

Source: Adobe Stock/Ilenetsnikolai

to recognize matches between patients and to classify them into groups with similar symptoms and progressions. These results are then compared by researchers with data from a patient study group in the Netherlands under a European collaboration. Researchers want to find out in particular what parameters and data are or are not relevant to disease progression. The aim is to be able to identify which group a patient belongs to as early as possible in order to enable optimal treatment.

It is important to keep an eye on all aspects of a disease like COPD which primarily concerns people aged over 50 and often goes hand in hand with additional health issues such as cardiovascular diseases, diabetes, osteoporosis or even psychological disorders. Once patients are hospitalized the treatment generally aims at what appears to be the biggest problem and focuses initially on improving the pulmonary function. Yet in some cases it is necessary to treat comorbidities first in order to improve the overall situation of the patient. It might be that patients have pain with every movement due to osteoporosis and therefore inhale less. The team is hoping that the new software will also help them with these cases. Once the available data is evaluated, the system could for instance suggest measuring bone density and draw the attention of the treating physicians to the above-mentioned osteoporosis-related issue.

First software tests in a year

The researchers are already collaborating with a company which develops and uses software for hospitals so that their new software can be rapidly translated into application. The software should be as simple, uncomplicated and non-time-consuming as possible in order to provide concrete support in daily clinical practice. The software should ideally be installed as part of the regular hospital IT and have access to existing patient data in order to be able to make on-the-spot suggestions for diagnosis and therapy to the treating physicians. "The software is expected to be ready for preliminary hospital trials in just a year's time", says Schmeck. In the longer term, the team is planning for a larger test phase followed by certification as a medical product.

The division of patients into different groups might not only help in therapy decisions but also provide for further possibilities, such as the trialing of new drugs. Medical drugs which help one patient do not necessarily improve the health of patients with dif-



Laboratory data is among the information used for the complex computer algorithm to better classify COPD patients.

Source: Philipps-Universität Marburg

ferent comorbidities. The researchers therefore hope that the findings generated by the software will not only improve the understanding of the disease but also lead to new approaches in the development of new therapies.

Further information is available at:

gesundheitsforschung-bmbf.de/de/systemmedizin-9458.php



BMBF-Newsletter:



A monthly overview of important news from the BMBF
bmbf.de/newsletter

Stay up to date with the BMBF on Twitter, Facebook and Instagram:

 twitter.com/BMBF_Bund

 facebook.com/bmbf.de

 instagram.com/bmbf.bund

news from the BIH

€13 million goes to the HEALTH-X dataLOFT health data space in Gaia-X

The German Federal Ministry for Economic Affairs and Climate Action has officially handed over the funding award notices for the projects in the European digital ecosystem initiative Gaia-X. Professor Roland Eils, founding director of the Digital Health Center at the Berlin Institute of Health at Charité (BIH), accepted a check for €13 million euros on behalf of the consortium for the HEALTH-X dataLOFT project. The project aims to create a “citizen-centered health data space” that places citizens at the center of its focus, transforming them from passive recipients of services to active partners. In addition to the BIH, numerous other research institutions as well as companies and institutions in the healthcare sector and patient associations are involved in the project.

The Bundesnetzagentur has approved the first eleven projects for the Gaia-X funding competition “Innovative and practical

applications and data spaces in the Gaia-X digital ecosystem.” Eleven projects were selected from over 130 submitted project outlines to receive funding in the first tranche. A total of around €117.4 million in funding will be awarded by the end of 2024.

In his video message at the presentation of the funding award notices for the Gaia-X funding competition, Germany’s Economics Minister, Dr. Robert Habeck, highlighted the importance and innovation of the Gaia-X projects in the context of digital transformation: “The launch of Gaia-X can only be described as a success. Now it is a matter of quickly paving the way for broader application.”

Transformation in the healthcare sector

“HEALTH-X dataLOFT has set itself the ambitious goal of achieving a transformation in the healthcare sector,” says Roland Eils, spokesperson for HEALTH-X and founding director of the BIH Center for Digital Health.

HEALTH-X involves a total of 14 partners from across Germany. They include research institutions such as the BIH, the Hasso Plattner Institute, Fraunhofer-Gesellschaft institutes, numerous companies from the medical technology sector, start-ups and SMEs from the healthcare sector, as well as cloud and IT companies from Europe and the United States.

www.health-x.org



Professor Roland Eils, founding director of the Digital Health Center at the Berlin Institute of Health at Charité (BIH), accepted a check for €13 million euros on behalf of the consortium for the HEALTH-X dataLOFT project. Source: BIH



Sylvia Thun receives German Cross of Merit

German President Frank-Walter Steinmeier has awarded Professor Sylvia Thun the Cross of Merit on Ribbon of the Order of Merit of the Federal Republic of Germany in recognition of her services in the field of healthcare digitalization. Both a physician and engineer by training, Thun is the Director of the Core Facility Digital Medicine and Interoperability at the Berlin Institute of Health at Charité (BIH). She has made special contributions to medical data standardization, with the aim of pulling together health data from different sources in order to facilitate better diagnostic and therapeutic decisions.

“We have a vast treasure trove of data. It would be unethical not to use these data.”

The healthcare sector generates countless data every day – information about diagnoses, treatments and disease progression as well as information about molecular details and metabolic processes. “We have a vast treasure trove of data,” Professor Sylvia Thun says. “It would be unethical not to use these data.”

German President Frank-Walter Steinmeier has awarded Professor Sylvia Thun the Cross of Merit on Ribbon of the Order of Merit of the Federal Republic of Germany in recognition of her services in the field of healthcare digitalization. Source: BIH/Thomas Rafalzyk

Yet data from research and healthcare are recorded differently in every laboratory and every hospital – they are formulated differently, formatted differently and stored in different software systems, if not on paper. That makes the data difficult to use. Interoperability, the ability to exchange and share data from a variety of sources, is now being addressed by a new expert panel.

Communication standards in healthcare

“We need communication standards in the healthcare system,” says Thun, a physician and medical IT specialist. But her real goal is to improve the care patients receive. “There are excellent projects that have shown how well things can be done when everybody works together. Many projects have clearly demonstrated that digitalization and interoperability can make healthcare better. And that’s what drives me: a desire to make healthcare better.”





Bianca Lemke (JWI) and Lutz Kloke (Cellbricks) with the 3D printer on the parabolic flight. The test takes place in zero gravity.
Source: Cellbricks

biological bandages from a 3D printer, for use in outer space

Scientists at the Berlin Institute of Health at Charité (BIH) have teamed up with a company called Cellbricks to develop a 3D printer that can produce a biological wound closure: Using a mix of gelatin and skin cells, the device prints a perfectly fitting bandage that can be used to close large wounds. This could be a good alternative to autologous skin grafting, and not just for burn victims on Earth: Astronauts could also be treated individually far away from any hospital. The researchers have conducted a parabolic flight experiment to prove that the printer even works in zero gravity.

When it comes to burns or abrasions that measure several square centimeters in total area, the body's self-healing powers are overwhelmed and often only an autologous skin graft can help. But problems arise both during the removal and transplantation of skin, explains Professor Georg Duda, director of the Julius Wolff Institute of Biomechanics and Musculoskeletal Regeneration at the BIH and spokesperson for the BIH Center for Regenerative Therapies (BCRT). "Unfortunately, skin grafts often lead to scarring, which is not a satisfactory outcome, both medically and cosmetically, for either the physician or the patient." In their search for an alternative, Duda's team of scientists came across Cellbricks, a company that has developed a 3D printer capable of producing custom skin patches of various sizes and shapes.

Biological printing ink

"In our case the printing ink is biological," explains Bianca Lemke, a PhD student working with Professor Duda. "It consists of a special form of gelatin with methacrylate ends that hardens when exposed to UV light. We mix that with skin cells that ideally come from the graft recipient. And if we go to great technical lengths, we can even print tiny tubes into the mixture, which we then populate with blood vessel cells to ensure that the skin patch contains blood vessels." The shape and size of the required wound closure can be adjusted on the printer, and printing can take one or several hours depending on how challenging the particular bandage is.

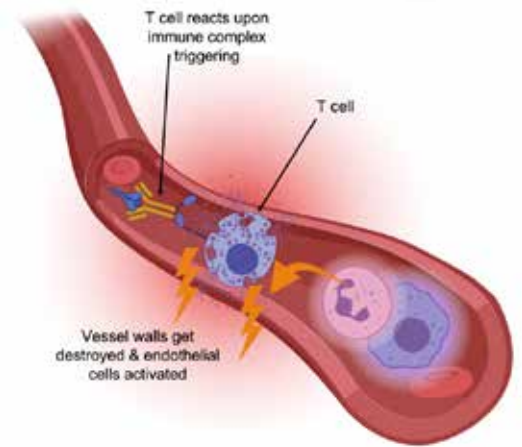
Parabolic flight experiment

During the five-hour flight, 31 parabolas were flown, which means the plane flew upward on a parabolic arc and then descended steeply before flying upward again. During the upper part of the parabola, zero gravity was experienced for about 20 seconds; during the ascent and descent, multiple gravitational forces were produced for a short time. To check how the Earth's gravitational force affects the printing result, Lemke had integrated microbeads into the printing ink.

Professor Duda stayed on earth but was happy to learn the results: "We really hoped the printing results were stable at least during weightlessness, and they were! That does mean we will hopefully be able to offer astronauts personalized wound care one day. Even though there is still a long way to go until then."

The CD16 positive T cells have an increased cytotoxic effect, especially on the inner cell layer of blood vessels.

Source: Translational Immunology Department at the BIH



the fatal role of T cells in COVID-19

Scientists from the Berlin Institute of Health at Charité (BIH), together with colleagues from Charité – Universitätsmedizin Berlin and the university hospitals in Bonn and Aachen, have found a type of immune cells that is particularly active in severely ill COVID-19 patients. The CD16 positive T cells have an increased cytotoxic effect, especially on the inner cell layer of blood vessels. Their presence, along with complement system factors, is associated with a highly fatal outcome of the disease. The scientists published their findings in the scientific journal *Cell*.

It is now virtually certain that a dysfunctional immune system plays a key role in severe COVID-19: Overactive immune cells attack and destroy the body's own tissue, even if the actual viral infection has already been contained or even overcome. Professor Birgit Sawitzki, head of the Translational Immunology Department at the BIH, is particularly interested in the role of T cells in SARS-CoV-2 infection. "T cells are the conductors of the whole orchestra of immune cells and signaling molecules," she explains. "T helper cells make it possible to develop a targeted defense using tailored antibodies, T killer cells specifically destroy infected or malignant cells in the body, and regulatory T cells make sure everything stays in sync. Unfortunately, certain T cells are responsible for a particularly severe course of COVID-19."

Safety mechanism is put out of action

To find out what role T cells play in COVID-19, the scientists used single-cell analyses to examine the blood of patients with mild or severe COVID-19 disease and compared it with the blood of healthy subjects and of patients with other viral infec-

tions. They discovered T cells carrying the CD16 molecule on their surface in individuals with severe COVID-19. "This was a surprise to the immunology community," Sawitzki reports, "because CD16 is actually expected on cells of the innate immune system like natural killer cells or monocytes, but not on T cells that belong to the acquired or specific immune system."

CD16 helps cells of the innate immune system recognize and kill virus-infected cells in the body. The molecule detects antibodies bound to virus-infected cells and then stimulates immune cells to release cytolytic enzymes, which destroy the virus-infected cells. Yet, T cells are not in need of such help.

The complement system is also involved

The researchers observed in laboratory experiments that upon contact with antibodies, the CD16 positive T cells released cytotoxic molecules and damaged pulmonary vascular cells. In collaboration with researchers from Aachen, they also discovered CD16-positive T cells in the lungs of deceased COVID-19 patients. While searching for the origin of CD16 positive T cells, the scientists came across the so-called complement system: This encompasses over 30 proteins that are dissolved in the blood plasma to help fend off microorganisms. They are activated in the course of the immune response by various mechanisms, such as by bound antibodies, and lead to the elimination of infected cells. "We have found that certain components of this system are abundantly produced in patients with a severe course of COVID-19 and contribute to the emergence of CD16 positive T cells. Here we seem to have discovered an important new link," Sawitzki suspects. "If this link is confirmed, inhibiting the complement system could potentially help reduce severe courses to a minimum."

mTOMADY

A digital health wallet for people without insurance

Company profile: mTOMADY

by Stefanie Seltmann

More than one billion people in low- and middle-income countries lack access to essential health care services, mainly due to inadequate safety networks. Physicians and global health researchers from Charité – Universitätsmedizin Berlin developed a digital solution to this problem and launched the social enterprise mTOMADY in late 2020. mTOMADY users can send, save, and spend funds for medical services securely and efficiently via a mobile phone-based system. The BIH Charité Digital Clinician Scientist Program and the Digital Health Accelerator Program of the Berlin Institute of Health (BIH) at Charité have supported this initiative. The mTOMADY health platform was initially rolled out in Madagascar, with other African countries to follow.



For many years now, physicians in the Department of Neurology and Experimental Neurology at the Charité – Universitätsmedizin Berlin, Dr. Julius Emmrich and Dr. Samuel Knauß, have been volunteering in developing countries such as Madagascar. Some 93 % of the population of this island nation lives below the poverty line and less than 5 % of Malagasy citizens have a bank account. As a result, the doctors have witnessed their fair share of distressing moments during their volunteer work. “For example, a patient who had been seriously injured in a

car accident died because his family could not afford to pay for his treatment in advance.” Knauß recounts. Likewise Julius Emmrich recalls: “We often saw seriously ill or injured patients refuse to go to the hospital due to fear of the high costs of treatment.”

Mobile money for healthcare

In recent years, Madagascar has made significant progress in terms of digitalization and large parts of the country have been equipped with cell towers, providing mobile network coverage even in the most remote areas. More than half of Madagascar’s population has a mobile phone, and more than one million new users are added every year. This increasing digitalization has spurred an enormous boom in mobile money, through which money can be transferred securely via mobile networks from the sender to the recipient, much like text messaging.

The challenging circumstances that Emmrich and Knauß encountered in their work, along with developments in mobile and digital technologies in Madagascar, led to the idea to create a health platform called mTOMADY, which means “strong and healthy” in Malagasy. mTOMADY works like a digital health wallet. To use the service, all you need is a SIM card that connects you to a local mobile network. Cash can be converted into mobile money at mobile cash points, which are available on almost every street corner in Madagascar. Users can efficiently and securely transfer their funds via the mobile money system to an account reserved for medical expenses.

A special feature of mTOMADY is that the money loaded onto the health account can only be used to pay for healthcare services. This ensures that users build up a financial cushion



Healthy and happy children in Ejeda, a town and commune in southwest Madagascar (Photos: © Doctors for Madagascar).

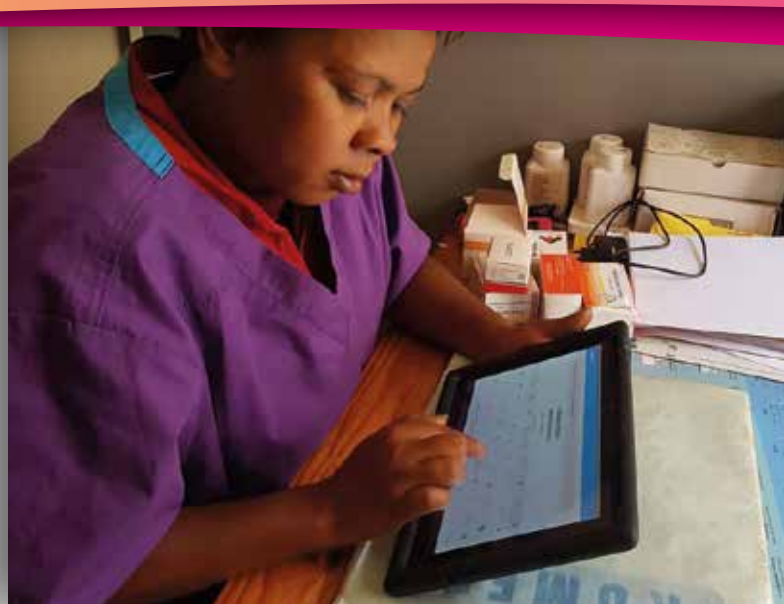
for emergencies or unforeseen expenses due to illness or accidents, which can save lives in critical situations. In addition, users can enroll in health insurance through the platform. Donations can also be made to a digital donation platform from which all mTOMADY members benefit.

BIH supported development and spin-off

Emmrich and Knauss have been participants in the BIH Charité Digital Clinician Scientist Program since 2019, which has allowed them to develop mTOMADY during their residency as neurologists. The BIH's Digital Health Accelerator provided around €1 million to the project from July 2018 to December 2020. The team received mentoring from experienced experts in areas such as mobile technology, software development, product development, and health insurance, as well as help in planning and preparing for the spin-off, which was completed in December 2020 with the founding of mTOMADY gGmbH. "We are excited that through our support of mTOMADY, a digital health solution is now available for developing countries and that we can contribute to social value creation in local communities," says Thomas Gazlig, head of BIH Charité Innovation, the joint technology transfer office of Charité and the BIH.

In January 2020, mTOMADY was an award winner in the Global Health Hub Germany's "New Ideas for Global Health" innova-





Medical personnel in Madagascar: (left) a training course and (right) mTOMADY in use at a clinic (Photos: © Doctors for Madagascar).

tion contest. The Malagasy Ministry of Health is currently working with mTOMADY to make the mTOMADY health wallet an integral part of the country's healthcare system.

Initial focus was on helping pregnant women

mTOMADY began by supporting pregnant women in Madagascar's central highlands. "Madagascar has a high maternal and infant mortality rate, and few births are attended by skilled health personnel," says Emmrich in explaining the choice. Knauß adds: "Starting with support for pregnant women first lays an important cornerstone for the country's future." mTOMADY has since opened up its service to other patient groups as well. Emmrich and Knauß are also in the process of integrating several local insurance companies into the health platform, to improve the efficiency of the service and offer patients more comprehensive support. The mTOMADY team, which now includes staff from nine different countries, plans to expand the service to Ghana and Uganda soon.

In a BIH Podcast, Samuel Knauss and Julius Emmrich talk about their experiences in Madagascar and the launching of mTOMADY (German only): <https://www.bihealth.org/de/aktuell/wie-bezahlt-man-den-arzt-in-madagaskar>

About the company:

- 🚀 **Founded by a German-Malagasy team in 2020**
- 🚀 **25 employees in Antananarivo and Berlin**
- 🚀 **Over 200,000 registered users**
- 🚀 **Over 45,000 treatments**

www.mtomady.de

If anyone would like to support the project, the organizers are always happy to receive donations. These can be made out to the nonprofit organization Doctors for Madagascar: <https://www.doctorsformadagascar.com>

Contact:

Dr. Samuel Knauß and Dr. Julius Emmrich
Department of Neurology and Experimental Neurology
Charité – Universitätsmedizin Berlin
samuel.knauss@charite.de
julius.emmrich@charite.de

Samuel Knauß (middle) and Julius Emmrich (right) with local medical personnel in front of a healthcare center in Antananarivo, Madagascar (Photo: © Doctors for Madagascar).



“Interdisciplinarity is the only way to answer complex questions”

Interview with Dana Westphal and Michael Seifert

Interdisciplinary junior research alliance is investigating new therapeutic approaches to melanoma brain metastases

Dana Westphal and Michael Seifert are leading one of the seven junior research alliances that began receiving support in 2020 as part of the BMBF's e:Med funding concept. In this interview with *gesundhyte.de*, Dr. Dana Westphal (a biologist) and Dr. Michael Seifert (a bioinformatician) talk about their interdisciplinary collaboration as well as goals and successes of their research alliance.

gesundhyte.de: Melanoma is the most dangerous form of skin cancer. What makes it such a threat to our health?

Dr. Dana Westphal: Every year, about 21 in 100,000 people will develop melanoma in Germany alone. Unfortunately, the incidence is rising, especially among young people. Melanomas are especially dangerous because they metastasize to different parts of the body much more frequently than non-melanoma skin cancers. In addition, roughly half of all patients with malignant melanoma develop brain metastases, which drastically reduces their chances of survival. Early and regular screenings are therefore necessary to detect and remove malignant precursors before metastases develop.

gesundhyte.de: You have been conducting research on melanoma for six years. What do you hope to achieve with the e:Med junior research alliance MelBrainSys?

Dr. Dana Westphal: Thankfully, effective therapies for malignant melanoma have been available since 2011. However, brain metastases remain a therapeutic challenge, especially when they become symptomatic. Our project aims at examining the differences between brain metastases and metastases in other organs and identifying targets for new brain-specific therapies.

Dr. Michael Seifert: Our research builds on a preceding project in which we established a data set of DNA-methylation and gene expression profiles from local melanoma patients. This data set contains one metastases pair for each patient, made up of its brain and non-brain metastasis. These patient-specific pairs allow us to determine differences between brain and non-brain metastases in much greater detail. Previously, researchers mainly compared brain and non-brain metastases from different patients. Our data set can contribute to overcome this. Hopefully, it will help us to identify candidate genes and altered signaling pathways for new types of therapies for brain metastases.



Dana Westphal and Michael Seifert are jointly coordinating MelBrainSys, an e:Med junior research alliance investigating new therapeutic approaches to melanoma brain metastases (Source: University Hospital Dresden/I. Starke, D. Westphal and M. Seifert/private).

gesundhyte.de: Did the first year of your project go well?

Dr. Dana Westphal: Like many other scientists, we have been affected by the coronavirus pandemic. Our laboratory had to shut down for three months. We used the time to characterize the patient samples we just talked about and prepare them for future analyses.

gesundhyte.de: What have you achieved in your first year?

Dr. Michael Seifert: We can now analyze the omics data of the individual metastases pairs in detail using the models we developed. In addition, we were able to prepare a large public data set in such a way that we can use the expression and methylation data to learn a genome-wide melanoma-specific gene regulatory network. Through this network and with the help of our patient data, we can now start to determine how specific gene alterations affect signaling pathways or immune signatures.

gesundhyte.de: What do you hope to learn from your research?

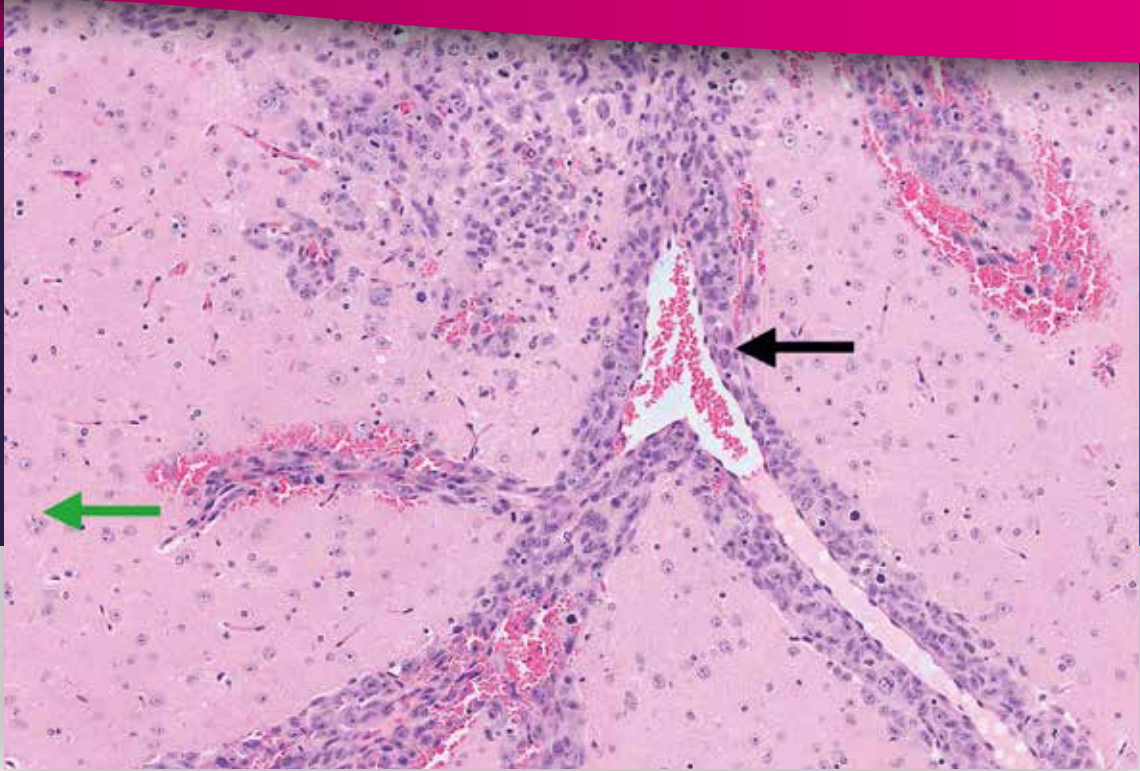
Dr. Dana Westphal: We are hoping to uncover cellular mechanisms that differ between brain and non-brain metastases. This could provide us with a starting point for new therapies that act specific against brain metastases. It would be wonderful if we would be successful with that.

Dr. Michael Seifert: Another joint long term goal of our work is to support physicians in their treatment decisions with the help of a classification system established on the basis of biomarkers that we hope to identify in our molecular data.

gesundhyte.de: How is your alliance structured?

Dr. Dana Westphal: In addition to Dr. Seifert and myself, Dr. Rebekka Wehner and Dr. Matthias Karreman are part of the project team. Dr. Wehner is an immunologist at the Institute of Immunology here in Dresden, and Dr. Karreman is a neuro-oncologist at the German Cancer Research Center in Heidelberg. The collaboration is going very well. We each bring our expertise to the project.

Dr. Michael Seifert: Interdisciplinarity is one of the most important features of our junior research alliance. As a bio-informatician, I could not perform my work if I do not get the molecular data from my project partners. On the other hand,



Hematoxylin-eosin staining of tumor tissue: melanoma cells (purple) grow along blood vessels (black arrow) and invade the surrounding brain tissue (green arrow) (Source: Dr. med. Matthias Meinhardt, Institute of Pathology, University Hospital Dresden).

my three project partners would not really be able to do a targeted investigation of any signaling pathways or candidate genes without the results of my subproject. We are working on a very complex research question that can only be tackled by an interdisciplinary approach, because none of us can be an expert in each discipline.

gesundhyte.de: Did it take a while to get into your stride as a team? Or did you work well together from the start?

Dr. Michael Seifert: Dr. Westphal and I had already worked well together on a previous project. During that time, we realized that we had to look at the molecular data of each patient much more individually and that we had to expand our research systematically. In order to look for new therapeutic approaches, we realized that we require a larger team with more diverse expertise. Already in the beginning of the grant writing process, we discovered that the four of us can work well together. But I also learned that some things that are obvious to me are not really obvious for others. A common language and detailed explanations of basic things in each discipline are needed to avoid misunderstandings.

Dr. Dana Westphal: It is important to talk to each other constantly. Thus, we have regular meetings to discuss the direction of the project and keep it on the right track.

gesundhyte.de: What made you decide to become a scientist?

Dr. Michael Seifert: When I was in school, I was really interested in informatics and biology and topics related to medicine also fascinated me. At the time when I graduated from high school, bioinformatics centers were established in Germany with a lot of funding. This gave bioinformatics a real boost and opened up new career pathways in science or industry. All these things supported my decision to study bioinformatics. As a scientist, I really like that my work can hopefully contribute to the development of better therapies.

Dr. Dana Westphal: I became a scientist because my biology teacher brought this topic across in such a great manner. Eventually, my journey led me to the other side of the world, and I did my PhD in New Zealand. Thereafter, I spent five years in Australia before coming back to Dresden. What I like most about my current position here at the University Hospital Dresden is the patient-oriented research and the feeling that my work is contributing to improving their well-being.

Marco Leuer conducted the interview.



MelBrainSys:

Investigating therapeutic targets for melanoma brain metastases

Melanoma is a form of skin cancer that has increased significantly in recent years. The tumor frequently metastasizes, even at the early stages of the disease. Almost half of all patients with malignant melanoma develop brain metastases during the course of the disease. Compared to metastases in other organs, those in the brain are much harder to treat with existing therapies. So far, little research has been performed to determine the key molecular factors and how they can be used to develop more effective therapeutic approaches in the future.

In addition to the specific genetic mutations in melanoma brain metastases, epigenetic reprogramming may also play an important role in the diminished effectiveness or resistance to existing treatment strategies. The research on the complex interplay of these molecular levels is very difficult and was further hampered in the past, because most of the available brain and organ metastases (e.g. in lung, liver, or kidney) were from different melanoma patients.

This is where the BMBF-funded e:Med junior research alliance MelBrainSys comes in. The project provides access to a special data set containing gene expression and methylation profiles of brain and organ metastases from the same patient. This allows us to perform patient-specific analyses that provide deeper insights into molecular changes of melanoma brain metastases. Using an innovative systems medicine approach, which includes both bioinformatics and wet lab experiments, we want

to identify candidate genes and altered signaling pathways that differ between brain and non-brain metastases. These comprehensive analyses will then be used to identify potential targets for new brain-specific therapies.

The interdisciplinary collaboration between dermatooncology, immunology, neurooncology, and bioinformatics is crucial to the success of this junior research alliance. It allows the team to meaningfully connect laboratory experiments and computer-based analyses to identify candidate genes and signaling pathways for further experiments in cell cultures and *in vivo* models.

The results obtained by the alliance's work have the potential to lay important foundations for future clinical studies to improve the treatment of melanoma brain metastases.

Contact:

Dana Westphal, PhD

Junior Group Leader WG Brain Metastases

Clinic and Polyclinic for Dermatology

Faculty of Medicine and University Hospital Carl Gustav Carus
TU Dresden

dana.westphal@uniklinikum-dresden.de

Dr. Michael Seifert

Group Leader Bioinformatics Core Unit

Institute for Medical Informatics and Biometry (IMB)

Faculty of Medicine Carl Gustav Carus

TU Dresden

michael.seifert@tu-dresden.de

INCOME – integrative, collaborative modeling in systems medicine

Hackathons for a culture of data and model sharing

by Nina Fischer, Wolfgang Müller, Dagmar Waltemath, Olaf Wolkenhauer and Jan Hasenauer

Systems medicine is an interdisciplinary approach in which physicians and clinical researchers collaborate with experts from the fields of biology, biostatistics, informatics and mathematics to improve the diagnosis, prevention and treatment of diseases. For this, large-scale, heterogeneous data sets must be collected experimentally and merged using holistic models. In reality, however, most projects are limited to individual cellular processes and develop customized models for these. In the e:Med project INCOME, funded by the German Federal Ministry of Education and Research (BMBF), we were able to promote collaboration and exchange between research groups through targeted networking activities. In numerous INCOME meetings, researchers and developers have jointly created a culture of data and model sharing, developing methods to improve the technical links and reusability of existing simulation models. In this way, the meetings provided a platform for strengthening the community.

Mathematical models are used in systems medicine to study disease-related biochemical processes behind diagnostic, prognostic and therapeutic decisions. These models can be formulated in open standards such as the Systems Biology Markup Language (SBML) and stored in model databases such as BioModels. Standards such as the Simulation Experiment Description Language (SED-ML) aim to unambiguously link data and

models and describe various experimental conditions. The development and establishment of standards and databases have had a huge impact on improving the availability and reusability of models and the reproducibility of research results. Nevertheless, many research groups and even widely used software tools have not yet adapted these standards. Instead, they use their own, less standardized formats, which often do not allow rigorous annotation. Correspondingly, according to recent surveys, only about 50% of published modeling studies are still reproducible (Tiwari *et al.*, 2021).

In addition to the models themselves, the data sets used for parameterization and validation are often difficult to access. They are not published, sufficiently annotated, or clearly linked to the model. The aforementioned factors prevent the reuse and expansion of existing models and data sets, thus compromising the success of research projects.

The severely limited reusability and the time required to develop high-quality models are two reasons why many mechanistic models represent only single signaling pathways and often ignore interactions with others. It is high time to leverage existing infrastructures, familiarize researchers with available resources, and make software developers explicitly aware of available libraries and tools for standard-compliant modeling.



Figure 1: Impressions from two INCOME events.
INCOME 2019 in Berlin and INCOME 2021 as a virtual event. (Photos: Jan Hasenauer)

Main objectives of the INCOME project

The objectives of the INCOME alliance project were to:

1. improve networking between research groups and to promote the use of standards
2. establish an integrated database for models and data
3. facilitate collaborative development of large-scale models

Work toward implementing these goals was carried out at several locations in Germany. INCOME (2017–2021) was coordinated by Jan Hasenauer (Helmholtz Zentrum München and University of Bonn). Wolfgang Müller (HITS Heidelberg), Olaf Wolkenhauer (University of Rostock) and Dagmar Waltemath (University Medicine Greifswald, formerly University of Rostock) also worked on the project.

Training, networking and community-building

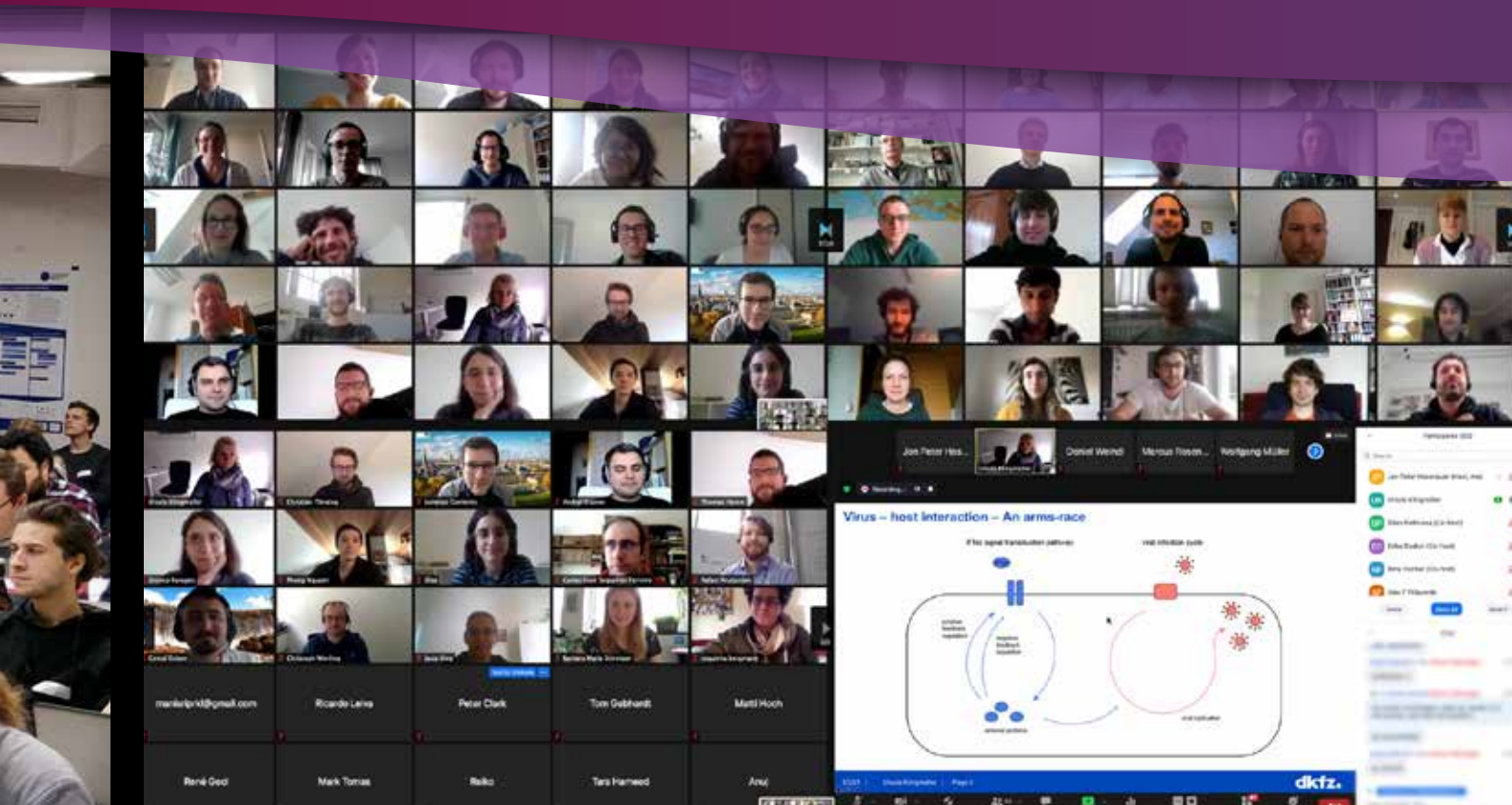
INCOME held a total of five events: three conferences with integrated hackathons and two separate hackathons (see Figure 1).

The conferences provided information on recent results in the areas of model building, method development and application. A diverse range of presentations also introduced standards

to a new generation of systems biologists and systems physicians and demonstrated their utility. Panel discussions, as well as group and individual meetings, were used to address open problems and develop solution approaches. The conferences offered extensive networking opportunities, enabling young scientists to exchange ideas with experienced scientists in the community.

Complementing the conferences, the hackathons offered enough space and time for working on joint projects. To significantly simplify the use of standards and improve both reproducibility and reusability, support for standards was provided in many available software tools (e.g., Copasi, Data2Dynamics, Dmod and pyPESTO). In addition, developers took time to demonstrate new software and advise users on any problems. In return, they received direct feedback on their own tools and were able to align further development with user needs.

The events attracted participants from over 20 countries. This internationality and the interdisciplinarity contributed significantly to the development of a strong community at the intersection of modeling, method development and standardization. The increasing number of participants in the events reflects



the growing interest in this area, as well a general interest in collaborating more intensively (see Figure 2). Many aspects of the INCOME events have been directly incorporated into the development of the global standardization network COMBINE (www.combine.org).

Closing gaps

In addition to implementing standards in existing workflows, INCOME also identified and addressed open issues. For example, at the first event in October 2018, it already became clear that there was no suitable solution to formulate parameterization problems. To address this gap, the Parameter Estimation table (PETab) format was developed during the hackathons, but also between meetings (see Figure 3). PETab allows a new level of standardization, improving the reusability of data and models for new studies, such as the development of integrated models.

A large number of research groups have collaborated in the development of PETab, and as a result, the new standard has already been made accessible to thousands of users in eight software tools in a very short time (Schmiester *et al.*, 2021). This would not have been possible without a strong, thematically focused community and the joint hackathons.

Creating resources

INCOME has helped create community resources for advancing systems biology and systems medicine as research disciplines. For example, a previously established repository for data sets and models (Hass *et al.*, 2019) was transferred to PETab and has been continuously expanded. This not only facilitates reuse, but also enables a realistic benchmarking of methods. A study on the scalability and robustness of existing numerical solution methods has already been completed (Städter *et al.*, 2021), and several more studies on optimization and sample generation methods should be available soon. In addition, as part of a large global community led by Marek Ostaszewski (Luxembourg Centre for Systems Biomedicine), the INCOME consortium has been working on a model for SARS-CoV-2 infections (Ostaszewski *et al.*, 2020). The resulting COVID-19 Disease Map is probably the most comprehensive description of molecular and cellular processes during infection.

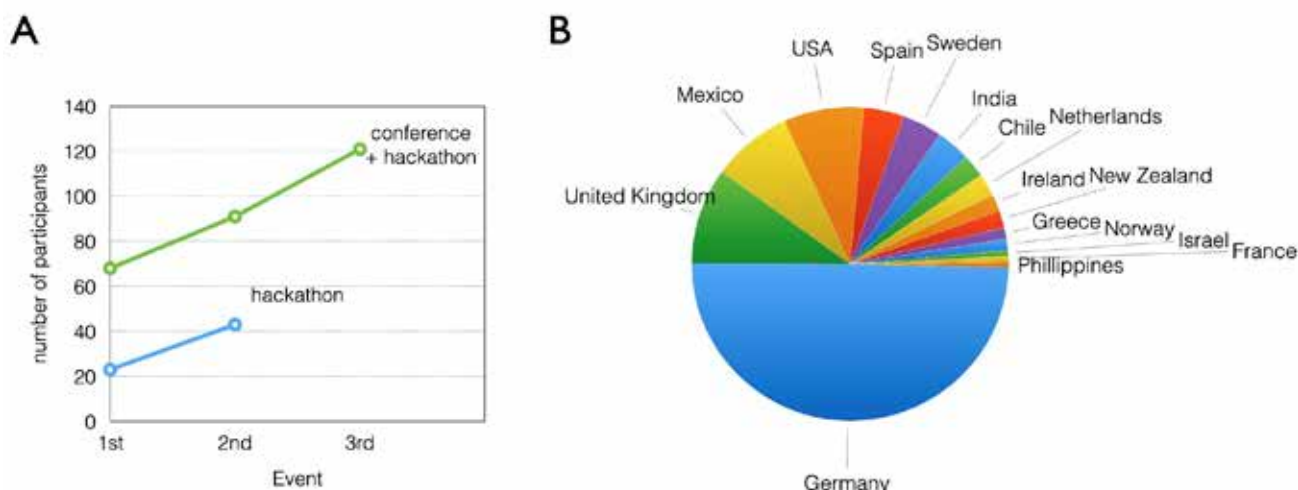


Figure 2: A shows statistics from registration data (number of participants) of the five events held from 2018–2021, while B shows statistics from registration data of the virtual event held in 2021 (Source: Jan Hasenauer).

All resources that have been created are freely available and can be easily used in further projects due to their high degree of standardization. This facilitates knowledge transfer and simplifies a holistic view of diseases.

But where do we go from here?

INCOME has shown all consortium members how much more can be achieved when they are part of a strong community, and also how important community-building is. The training sessions during the conferences, exchange and collaborative work during the hackathons, and numerous discussions have all helped create a closer connection between model, software and standards development. This has been a process. Individual meetings would not have been sufficient, and there was substantial added value gained from repeated exchange between returning participants, who in turn promoted the meetings in their research circles. We firmly believe that the knowledge that has been and continues to be carried into individual research institutions through INCOME events has significantly improved and will continue to increase the reusability of models. Our goal is to reach many more

researchers in the future. We therefore plan to continue the event series beyond the end of the project (and are working on the feasibility).

We thank all the participants from the conferences and hackathons who have contributed to the success of INCOME.

References:

- Tiwari *et al.* (2021). Reproducibility in systems biology modeling. *Molecular Systems Biology*, 17:e9982.
- Schmiester *et al.* (2021). PETab - Interoperable specification of parameter estimation problems in systems biology. *PLOS Computational Biology*, 17(1):e1008646.
- Hass *et al.* (2019). Benchmark problems for dynamic modeling of intracellular processes. *Bioinformatics*, 35(17):3073-3082.
- Städter *et al.* (2021). Benchmarking of numerical integration methods for ODE models of biological systems. *Scientific Reports*, 11:2696.

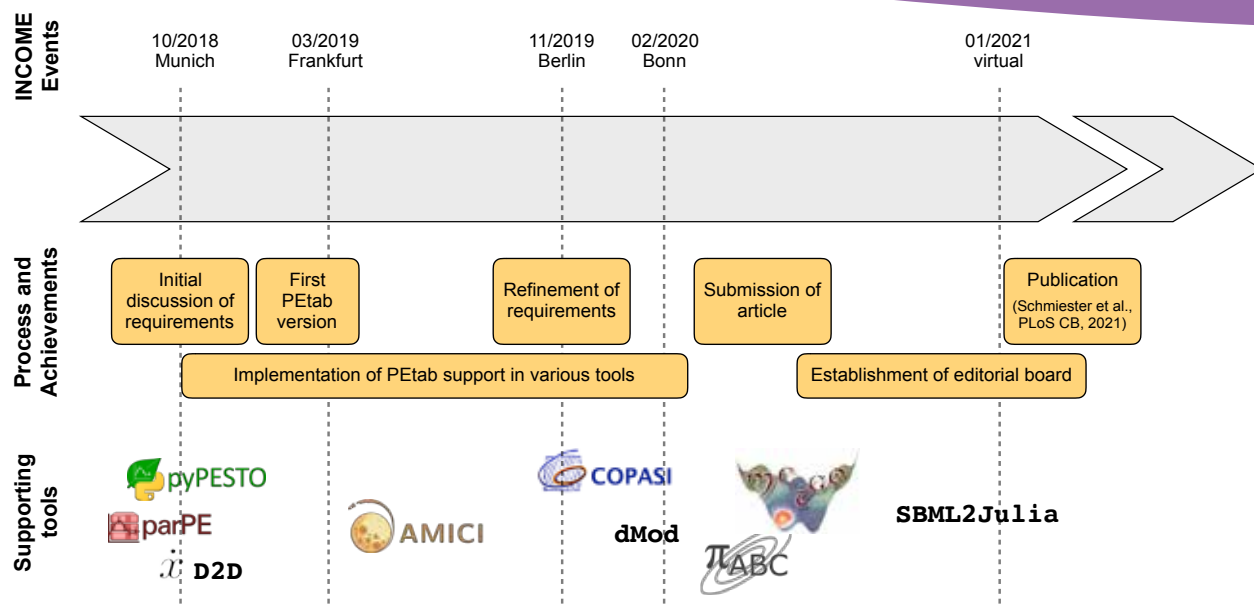


Figure 3: Development of the PETab standard within the framework of the INCOME conferences (Source: Jan Hasenauer).

Ostaszewski *et al.* (2020). COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms. *Scientific Data*, 1:136.

Contact:



Prof. Dr. Jan Hasenauer
Life and Medical Sciences (LIMES) Institute
University of Bonn
jan.hasenauer@uni-bonn.de

www.mathematics-and-life-sciences.uni-bonn.de



E:MED NETWORKING FUNDS

The **projects** supported by the **e:Med Networking Funds** show how valuable resources like software tools, repositories and standards can be developed in hackathons, workshops and summer schools. One success of these projects is specific translational applications in clinical trials on diseases like gastric cancer, breast cancer and inflammatory bowel disorders. Six research alliances link together several e:Med projects in a targeted manner. The scientists contribute their expertise on relevant cross-cutting topics, thereby generating crucial added value through their interdisciplinary collaboration on clinical research, basic research, bioinformatics and modeling. This initiative has been made possible by the BMBF as part of its e:Med funding scheme through resources from the Networking Funds.

www.sys-med.de/en

e:Med junior research alliance LeukoSyStem

Interview with Simon Haas

Single-cell multi-omics leukemia researcher;
e:Med alliance leader; BIH, Charité and MDC in Berlin;
DKFZ and HI-STEM in Heidelberg

Simon Haas researches like he jogs: he constantly explores new directions. He was one of the first scientists in the world to examine hematopoiesis using novel single-cell technologies – and in doing so he turned conventional knowledge about blood formation on its head. Where a method is lacking, he develops it: most recently spatial single-cell multi-omics to detect leukemia stem cells and target them with therapies.

gesundhyte.de: You are leading the **junior research alliance LeukoSyStem** within e:Med, a major funding scheme for systems medicine set up by the German Federal Ministry of Education and Research (BMBF). This is not the first project you have led, so what is LeukoSyStem aiming to do?

Dr. Simon Haas: Our main aim with the alliance is to better understand leukemias (blood cancers) by focusing on stem cells and using **single-cell methods to identify new therapeutic avenues**. Leukemias are typically treated with traditional chemotherapy. But that's a bit like using a bulldozer. Although it effectively fights cancer cells, it also damages healthy cells and this leads to severe side effects. Moreover, chemotherapy frequently leaves behind cancer cells, and these are often so-called cancer stem cells. They are only a part of the tumor, but these few cells can completely regenerate the entire tumor mass. It is precisely these cancer stem cells that often provide the reservoir for relapse after successful therapy.

So we're aiming to develop therapeutic approaches that specifically kill leukemia cells and especially leukemia stem cells, but leave healthy cells untouched. These therapies should have

fewer side effects and at the same time prevent relapse, which is the main cause of death in most forms of leukemias. To accomplish this, we are using so-called single-cell technologies. We obtain a wealth of information from individual cells, which enable us to understand in particular which cell is healthy and which is diseased. In this way, we can specifically identify target structures on the cancer cells and render these cells harmless while keeping healthy cells alive.

"Leukemias are frequently treated with traditional chemotherapy. But that's a bit like using a bulldozer. Although it effectively fights cancer cells, it also damages healthy cells."

gesundhyte.de: That sounds plausible and very effective. What methods are you using to pursue this? Did you develop them yourself?

The single-cell field is a relatively new research direction that has emerged only in the last few years. LeukoSyStem's members have been involved in the field since its inception and have developed many of the methods themselves. These include a spectrum of methods, from molecular experimentation to data analytics to medicine.

gesundhyte.de: Your publications give the impression of an extraordinary goal-oriented journey. Has that been the case, or has each answer you found opened up a hundred new questions from which to choose?



Simon Haas is developing single-cell multi-omics in order to detect leukemia stem cells – and target them with therapies.
(Source: © Felix Petermann, MDC)

Of course, it's a combination of both: Ideally you have a goal in mind, but you've also got new results coming in daily so you make adjustments according to what they say. You see what's possible with a new, specific method and you do things that weren't possible before. **Science really does thrive on unexpected results.** It's quite wonderful to follow them and perhaps arrive at findings that are even more exciting than those we expected from the initial research question.

gesundhyte.de: *What about the discovery that the microenvironment is so important? Did that come as a surprise, or were you expecting it?*

We've known that the cellular environment plays a key role in cancers for a while now. The immune system in particular is constantly working to fight the cancer. We've developed methods that allow us to examine this microenvironment in more detail. Initially, this was to help us **understand the healthy stem cell system. But now we're also using it to better understand leukemias.**

gesundhyte.de: *What are those methods and why do they work so well?*

Single-cell methods allow us to deconstruct tissue into its individual cells and see which components make up the healthy tissue and which make up the cancerous tissue. Our consortium has developed a variety of **single-cell multi-omics methods** that allow us to gather different types of information from individual cells at the same time. So for instance, we can simultaneously detect the activity of thousands of genes, DNA damage

(such as mutations), and cell surface molecules in individual cells. This is particularly promising because many targeted therapies are directed toward cell surface structures. We're now using the methods we developed to identify very specific molecular targets so that we can accurately attack and eliminate cancer cells.

However, one limitation of the single-cell method is that we lose the spatial information. We understand the individual cells that form a tissue's scaffolding, but we can no longer see where they're located spatially. Traditionally, scientists are using microscopes to see how the cells are arranged spatially in a piece of tissue, but one can only examine a few parameters at a time like that. That's why we and other researchers have spent the past few years developing "**spatial omics.**" These are methods that allow us to visualize thousands of pieces of spatially resolved molecular information. The idea is to join these two worlds – spatial omics and single-cell omics methods – so that we can better understand how an organ or tumor is constructed and organized.

"We're now using the methods we developed to identify very specific molecular targets so that we can accurately attack and eliminate cancer cells."

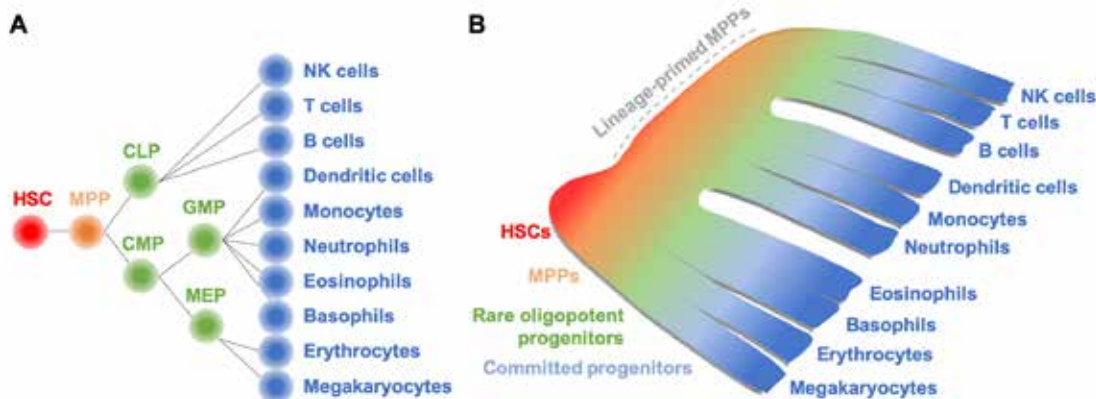


Figure 1: A) Classical model of blood formation (hematopoiesis), B) Revised model of blood formation.

(Source: <https://www.sciencedirect.com/science/article/abs/pii/S0301472X20302678>).

gesundhyte.de: That's fascinating. Are these novel methods also helping you understand physiological processes better? Which methods do you think have the most potential?

We recently showed how blood stem cells develop – these stem cells are responsible for creating new blood and immune cells throughout a person's life. This was the first time these rare cells had ever been characterized using single-cell multi-omics technologies. As well as giving us a detailed understanding of these important processes at a molecular and cellular level, it also provided a new model for hematopoiesis that is now recognized as the standard. Prior to this, the textbook knowledge of how blood and immune cells are formed in the bone marrow was that stem cells develop step-wise and then gradually branch out, like a tree. It turns out that it's not a step-by-step process at all. We demonstrated that a) the stem cells choose a blood cell line very early on, and b) this differentiation occurs in a continuous flow. This model has now essentially replaced the old one.

"Berlin is really the place to be right now. There's so much going on – exciting research, new institutions."

The spatial omics methods that we developed also allowed us to examine at an extremely high resolution the bone marrow environment in which the blood stem cells exist. This led to the discovery of several new cell types that help the stem cells do their job.

I think single-cell multi-omics analyses that allow us to examine hundreds of surface markers have the greatest potential for the future. The methods could bring vast therapeutic benefits. We're currently working on a way to use them in a clinical setting.

gesundhyte.de: That would be a really important step. Does Charité offer especially good conditions for this? You've only been in Berlin for about a year. What were the reasons for your choice?

Berlin is really the place to be right now. There's so much going on – exciting research, new institutions. I'm associated with three institutions here: the BIH, which specializes in medical translation; Charité, one of the largest university hospitals in Europe; and the MDC, which is a leader in developing single-cell technologies. The joint focus area "Single Cell Approaches for Personalized Medicine," which was set up by Professor Angelika Eggert and Professor Nikolaus Rajewsky, gives the institutes a platform where they can develop single-cell methods together. At Charité, I'm part of the Department of Hematology, Oncology and Tumor Immunology. This means I can interact closely with the clinicians and we can integrate the new methods directly into clinical routines.

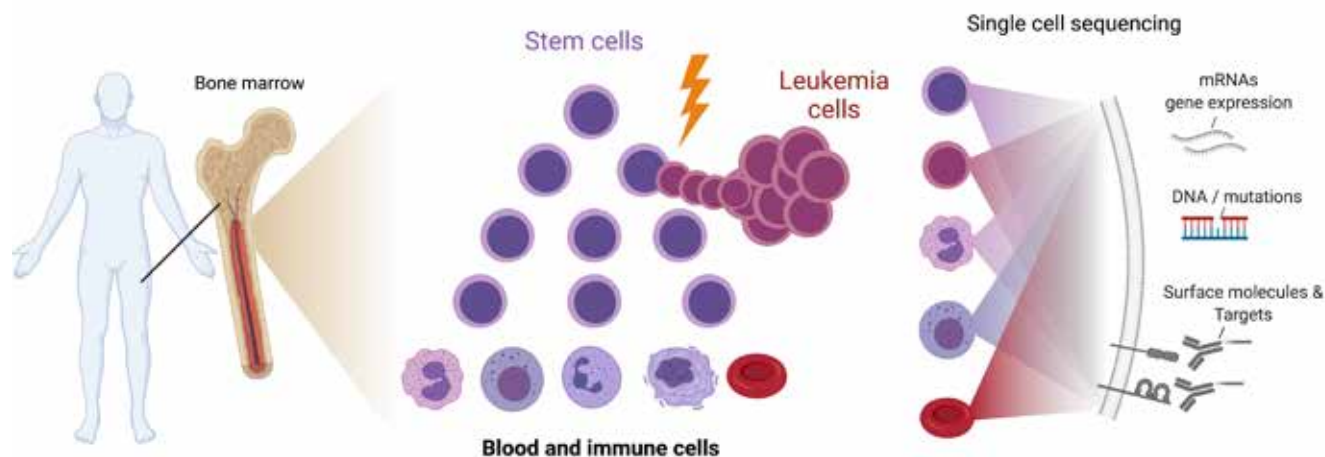


Figure 2: Single-cell sequencing enables the study of leukemias. Healthy hematopoietic stem cells in the bone marrow produce blood and immune cells. The cellular environment in the bone marrow (niche) is essential for hematopoiesis. Leukemias develop from hematopoietic stem cells or progenitor cells through the accumulation of DNA damage (e.g. mutations). Single-cell sequencing provides a detailed understanding of the molecular processes that occur during the transformation of healthy stem cells into leukemia cells. The interaction between leukemia cells and the cellular bone marrow environment can also be studied. (Source: © Simon Haas, Created with BioRender.com)

What I like most about Berlin is **the energy and the willingness to work together** – across different institutes in a really interdisciplinary way – to achieve something meaningful.

The **funding that the BMBF provided through e:Med** has helped get things in Berlin off to a good start. The **support for young scientists** is especially valuable as it carves out space for creative ideas. Systems medicine is going to be important in all areas of modern medicine and will play a significant part in improving medical care.

"It's very gratifying to see your work advance understanding in a field."

gesundhyte.de: You "grew up" in the scientific community in **Heidelberg**. Until recently, you were working at the German Cancer Research Center (DKFZ) and the Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM). How has that shaped you as a scientist?

I've studied and done research in the UK, the USA and Germany. But the Heidelberg research scene definitely had the biggest influence on me. Heidelberg is an incredible research location, especially when it comes to biomedicine and cancer research – and that's primarily thanks to the DKFZ, the European Molecular Biology Laboratory, and the university hospital. I owe an enormous amount to this environment. Even when I was still a student, I was able to participate in research activities, such as the iGEM program for synthetic biology that Professor Roland Eils coordinated. I wrote my doctoral thesis under Dr. Marieke Essers at the DKFZ, and set up my first research group at HI-STEM under the leadership of Professor Andreas Trumpp. I actually still lead a research group there and am involved in a lot of collaborations with Heidelberg.



Simon Haas in his laboratory at BIMS-B-MDC/BIH.
(Source: @ Thomas Rafalzyk, BIH)

gesundhyte.de: Do you see your work being translated into clinical practice?

Absolutely. It's not an easy road, but it's our goal. **We'll keep going until we get it right** – and that means sparing no effort and working in close collaboration with the Departments of Hematology and Oncology at Charité and the Heidelberg University Hospital. I'm very confident that we'll be able to use many aspects of basic research to achieve significantly better prognoses, diagnoses and therapies for patients. Right now, when a cancer patient comes to the hospital, they're examined using a battery of different diagnostic assays. With the single-cell multi-omics analyses, we're trying to develop diagnostic tests that combine everything into one test so that we can predict as accurately as possible which patient will respond best to which targeted therapy. That's still a long way off, but hopefully we'll get there soon.

"I'm very confident that we'll be able to use many aspects of basic research to achieve significantly better prognoses, diagnoses and therapies for patients."

gesundhyte.de: You are very creative and go-getting. What drives you in your daily work?

I love breaking new ground and going down new paths. What's motivating about science is that you face new challenges every day and are always learning something new. You develop yourself along with your own research and discoveries. It's very gratifying to see **your work advance understanding in a field**. The scientist's profession may come with many obstacles, but it's a great advantage to be able to set your own research agenda.

gesundhyte.de: Let's come back to your newly adopted city of Berlin. How do you perceive it, and what do you do when you have free time?

I love Berlin. It's just incredibly diverse, it never stands still and there's always something new. Berlin's history is of course also particularly interesting, and it's very evident in many places. Outside of work, I exercise a lot. From my apartment in Mitte, it's easy to go jogging and explore Berlin in all directions. The COVID-19 pandemic has had us in its grip since I've been in Berlin, so I'm looking forward to being able to immerse myself in Berlin's cultural life more than is currently possible.

Dr. Silke Argo conducted the interview.

Contact:

Dr. Simon Haas

BIH, Charité, MDC Berlin

DKFZ, HI-STEM gGmbH Heidelberg

Simon.haas@bih-charite.de

<https://www.bihealth.org/en/research/research-group/systems-hematology-stem-cells-precision-medicine>

"we love data!"

A portrait of BIH Professor Claudia Langenberg

by Stefanie Seltsmann

Professor Claudia Langenberg has headed the Computational Medicine Group at the Berlin Institute of Health (BIH) at Charité since September 2020. Combining her expertise in genetic epidemiology and public health, she examines the molecular mechanisms of metabolic diseases such as type 2 diabetes using large amounts of data from international patient and population studies. The BIH recruited the German-British researcher from the University of Cambridge, where she had been Programme Lead for Molecular Epidemiology at the Medical Research Council (MRC) Epidemiology Unit.

Claudia Langenberg speaks quickly with a slight British accent in the Zoom interview, even though she was born in Munich. "Twenty years in London leave their mark," laughs the physician, who has set her sights on a weighty subject: she is investigating the causes of obesity, insulin resistance and type 2 diabetes. "These common chronic metabolic diseases have both environmental and genetic causes," says the specialist in public health, a field concerned with the health of the population as a whole. With obesity, for instance, lack of movement and excessive intake of high-calorie foods play a role, but genetic susceptibility also increases our risk of putting on weight or distributing

fat unfavorably. "We need huge studies with tens or hundreds of thousands of participants in order to be able to clearly determine the influence of individual genes, partly because we are testing millions of genetic variants," Langenberg says, explaining the core of her research. "This requires powerful software and computers." Despite the relatively small effects of individual variants, Langenberg points out, they can help us decipher the causes of and mechanisms underlying metabolic diseases, which in turn allows us to assess the chances of success of new or existing therapeutic strategies. "We love data," she says, commenting on the new field of computational or data-driven medicine.

"We need huge studies with tens or hundreds of thousands of participants in order to be able to clearly determine the influence of individual genes."



Photo: Thomas Rätzky/BIH



Berlin's former Governing Mayor Michael Müller opens the exhibition "Berlin – Capital of Women Scientists" with some of the women featured, including Claudia Langenberg, far left (Photo: BIH/Konstantin Börner).

The "architecture of metabolism"

In recent years, Langenberg has been particularly interested in analyzing metabolites and proteins circulating in human blood. Their composition is different in every person, they represent our 'chemical individuality' so to speak, and genetic variation has a big influence on this. *"Today, hundreds and thousands of these molecules can be detected and measured on a massive scale. A few years ago, this would have been impossible to imagine,"* says the scientist enthusiastically. In collaboration with international researchers from other groups, her work has enabled the team to produce an atlas of the "architecture of metabolism". *"It is unusual for an epidemiologist to create knowledge for the next generation of biochemistry books,"* Langenberg smiles.

"Today, hundreds and thousands of these molecules can be detected and measured on a massive scale. A few years ago, this would have been impossible to imagine."

But what the scientist finds particularly fascinating is distinguishing between changes that lead to the development of metabolic diseases and subtle measurement differences in healthy people. *"Many of the responsible genes we found are known to cause severe 'monogenic' diseases, so called 'Inborn Errors of metabolism' that are rare and generally apparent early in life,"* Langenberg reports. *"Evidently, the same genes also lead to large differences in the same metabolites affected in such patients in our studies of the general population. Our task now is to find out whether and how these changes manifest themselves clinically, because this will enable us to identify people at risk, better understand the underlying mechanisms and identify potential new therapeutic strategies."* Her work therefore is fully in line with the BIH's mission of turning research into health.

The role of genes in COVID-19

"I was very happy to come to Berlin," Langenberg says. *"I not only have the enormous support of the Translation Hubs and Core Facilities of the BIH in evaluating large amounts of data, but I also have access to internationally renowned cooperation partners at the BIH and Charité."* For instance, Langenberg has for some time now been working with Professor Markus Ralser, Director of the Institute of Biochemistry at Charité, to, among other things, analyze protein components in the blood of COVID-19 patients.

"We were able to use the results of the Berlin study to identify genetic influences on COVID-19-relevant proteins in our Cambridge studies and quickly make them available to the scientific community," Langenberg explains. For this purpose, she adds, we teamed up with colleagues from Munich to develop an interactive web server (<https://www.omicscience.org>).

Claudia Langenberg is also a big fan of Berlin in general: *"It's a terrific city. My father was born in Berlin, and I think there's no better place for my family to get acquainted with German history. The 'Berliner Schnauze' is pretty much the exact opposite of British politeness and understatement, but such contrasts are what make living a binational life so exciting,"* laughs the mother of twin daughters. That was also one reason for the move to Germany. *"I really wanted my daughters to go to a bilingual German school and grow up among different cultures and languages. The international*

community there is really great, even though of course school got off to a bumpy start in the coronavirus years." Langenberg says. Her British husband, who is also a scientist, is currently working as a visiting scientist at the BIH, in addition to his leading the Institute of Health Informatics at University College London, UK.

Claudia Langenberg continues aspects of her work in the UK and is still leading a team at the University of Cambridge. *"Some large, ongoing studies will not be completed for one or two years, and we are in the process of forming new international collaborations, which will be managed from Berlin. It's a win-win situation for both sides."*

Contact:

Prof. Dr. Claudia Langenberg

Computational Medicine,

Berlin Institute of Health (BIH) at Charité

and MRC Epidemiology Unit, University of Cambridge, UK

Claudia.Langenberg@charite.de

www.bihealth.org/en/research/research-group/computational-medicine

"We were able to use the results of the Berlin study to identify genetic influences on COVID-19-relevant proteins in our Cambridge studies and quickly make them available to the scientific community."

About Claudia Langenberg

Claudia Langenberg was born in Munich and studied medicine in Münster. After completing clinical training in Germany, she moved to the United Kingdom and the United States to obtain her master's and PhD in epidemiology. In 2016, she completed specialist training in public health and the following year was appointed Programme Lead for Molecular Epidemiology at the University of Cambridge, where she had previously uncovered the genetic basis and risk factors of metabolic diseases and was involved in setting up international consortia and meta-analyses. Langenberg has published over 250 scientific papers, many in prestigious journals, and is recognized as a Highly Cited Researcher by Thomson Reuters, a distinction received for being among the most cited researchers in her field. She has received numerous awards for her scientific work and served as editor in chief of "Generation Genome," an annual report of the UK's chief medical officer which contributed significantly to the reorganization of genomic medicine in the country's National Health Service (NHS).

The Digital Hub CAEHR

Optimal & personalized treatment decisions for cardiovascular patients

by Dagmar Krefting

A lack of communication at the interfaces of the healthcare system often leads to suboptimal treatment due to a loss of information and time. With the Digital Hub CAEHR – coordinated by the University Medical Center Göttingen – a research-compatible electronic patient record (ePA) will be developed and implemented to enable the collection of structured patient data at all points of the care system according to uniform standards and used throughout the entire care chain. This will enable more precise prognoses regarding the course of diseases, which in turn can lead to optimized therapies and newly developed care models.

Whether stroke, heart failure, or coronary heart disease – cardiovascular diseases are still the most common cause of death in Germany despite important advances in treatment. Moreover, these diseases often take a chronic course. Both the patients affected and the doctors treating them must be prepared for long-term, individualized and personalized treatment.

What are the goals of CAEHR?

And this is precisely where the Digital Hub CAEHR (CARDiovascular diseases - Enhancing Healthcare through cross-sectoral Routine data integration), funded by the Federal Ministry of Education and Research (BMBF), comes into play. If cardiovascular diseases are diagnosed at an early stage, they can be treated well in many cases.

For precisely this patient-centered treatment, CAEHR standardizes and structures the health data from outpatient and inpatient care and makes it accessible in a research-compatible ePA to all actors along the entire treatment pathway (paramedics, nurses, doctors, and patients) for individual patient care. CAEHR is thus concerned with improved care for patients through the optimized and timely provision of relevant, standardized and structured health information and the establishment of intelligent data-driven services. In three regions of Germany – Hanover/Göttingen, Berlin, and Würzburg/Mainfranken – CAEHR will test digital solutions for better care of people with cardiovascular diseases over the next four years and further develop them for a nationwide use in the future.



The lack of communication at the interfaces of the treatment pathway leads to a loss of information and time. Standardization and digitalization of data enable the intersectoral provision of relevant information at the right place at the right time - and the use of this data for the development of clinical decision support systems, which should lead to a measurable improvement in health care.

(Source: © CAEHR-Consortium)



Prof. Dr. Udo Bavendiek, Clinical Leadership of CAEHR, Hannover Medical School (Source: © HiGHmed)

What are Digital Hubs?

The Digital Hubs, funded by the Federal Ministry of Education and Research (BMBF), aim to incorporate the pioneering work of the Medical Informatics Initiative (MII) on digitalization in medicine from university hospitals – initially in pilot projects – into all areas of the health care system: from outpatient care in the general practitioner's office to inpatient stays at local hospitals and care in rehabilitation and nursing facilities.

Use Cases

At the various points along the treatment pathway, healthcare providers often have incomplete access to patient data. CAEHR therefore, focuses on the flow of information between the different sectors of the health system by means of three use cases: The **use case “Emergency Stroke Care”** is dedicated to stroke patients and the interface between acute inpatient and emergency care. The **use case “Rehabilitation”** focuses on the interface between inpatient care and rehabilitation for patients after aortic valve replacement using catheter technology, and the **use case “Outpatient care”** focuses on the interface between inpatient and outpatient care for patients with chronic heart failure and coronary heart disease.

The resulting added value for the patients, the specialist staff involved, the scientists involved and the healthcare system as a whole is to be demonstrated in the three mentioned use cases.

Use Case 1: Emergency Stroke Care

This use case deals with optimized resource allocation. The relevant patient data already generated in the ambulance is to be transferred directly to the target hospital and the emergency physicians are to be supported by AI-supported systems in making time-critical therapy decisions.

Use Case 2: Rehabilitation after heart surgery

In this use case, individual rehabilitation services are developed. The documentation should be digital and no longer paper-based. In addition, the interprofessional planning of individual therapies is to be strengthened.

Use Case 3: Outpatient care for coronary heart disease

This use case aims to improve data interfaces through IT solutions. Wearable sensor technology should increasingly be used in the home environment. In addition, seamless documentation in diagnostics and therapy should contribute to strengthening prevention.

Project partners

The CAEHR project comprises of a total of 28 collaborative partners and consists of the following nine direct grant partners:

- University Medical Center Göttingen and Georg-August University Göttingen
- Hannover Medical School
- Charité – Universitätsmedizin Berlin
- University Hospital Würzburg
- Osnabrück University of Applied Sciences
- HiGHmed Association
- Vitasystems GmbH
- AOK Lower Saxony
- System Vertrieb Alexander GmbH

In addition, numerous other institutions from science and industry, as well as patient representatives are active partners in the CAEHR project. An overview can be found here:

www.gesundheitsforschung-bmbf.de

Contact:

Prof. Dr. Dagmar Krefting

Head of the Digital Hub CAEHR
University Medical Center Göttingen
dagmar.krefting@med.uni-goettingen.de

Eva König

Press and Public Relations HiGHmed
communications@highmed.org

news

And the winners are...

MTZ® Award for Systems Medicine 2022

This year's winners of the MTZ® Award for Systems Medicine have been announced: **Dr. Paul Stapor** (Technical University of Munich), **Dr. Marcus Rosenblatt** (Albert Ludwig University of Freiburg) and **Dr. Fabian Kern** (Saarland University) convinced the national review panel and the board of the MTZ® Foundation with their excellent doctoral theses.

The MTZ® Foundation awards the prize to young scientists who have achieved groundbreaking and outstanding results in the field of system-oriented health research in their doctoral theses. The focus of the dissertation can be a molecular genetic, clinical, mathematical or informatics approach. However, the work should be part of an interdisciplinary environment, as it is characteristic for a systems medicine approach. Only by sharing expertise, complex biological phenomena can be deciphered and made useful for personalized medicine.

The prize wants to highlight promising young scientists and help them to gain public recognition. The MTZ® Foundation cooperates with the German Federal Ministry of Education and Research (BMBF) and Project Management Jülich (PtJ).



– for a better future –

The MTZ® Prize for Systems Medicine 2022 has been awarded for the first time under this name, but continues the successful series of the MTZ® Award for Medical Systems Biology. That has already 21 prizewinners since 2008. The award consists of a certificate and prize money (€10,000 divided by three prizewinners). The award ceremony was part of the national conference "Systems Biology of Mammalian Cells" in Heidelberg in May 2022.

Further information (in German):

www.mtzstiftung.de

<https://www.mtzstiftung.de/stiftung/mtzfoundation/mtzaward>

www.ptj.de/systembiologie

<https://sbmc2022.de>

Source: Project Management Jülich (PtJ)



Photo: shutterstock © Syda Productions

de.NBI TRAINING

Bioinformatics Training and Competence Building

The German Network for Bioinformatics Infrastructure (de.NBI) organizes training events and provides online training materials to enable bioinformaticians and life scientists to exploit their own and publicly available data more effectively.



Find out more at: www.denbi.de/training

events

“Right at the right time”

Systems Biology of Human Disease (SBHD) conference goes hybrid in 2021 for the first time

by Franziska Müller and Julius Upmeyer zu Belzen
on behalf of SBHD 2021's organizers

If not now, when? Under this motto, the thirteenth annual International Conference on Systems Biology of Human Diseases (SBHD) took place in early July. After being held successfully in Berlin for the first time in 2019, the conference was back at the Kaiserin-Friedrich-Haus for SBHD 2021. The venue, located right next to the famous Charité-tower, has hosted scientific events for over a century.

Despite a quite manageable number of on-site participants and a considerable number of hygienic precautions, the event was a complete success. After more than a year of only virtual meetings, the systems biology community could hardly be restrained in their excitement to see each other face to face again.

“It felt like being at a festival,” was among the many pieces of feedback we heard. It is strange how two years of pandemic change our perceptions: there were barely 60 people in a room designed for 170. But we can only agree with that comment: We also felt like we were at a festival, the energy and atmosphere were just incredible. This was certainly also felt by the 65 online participants, who, although not physically present, equally gave talks, took part in discussions, and presented posters.

However, it was not only the side program, which was visibly enjoyed by all participants, which made the event such a success, but rather the very interesting presentations and discussions.

The current COVID-19 pandemic not only left its mark in terms of the hygiene measures and the hybrid format, but also many of the talks directly addressed COVID-19 research. Various approaches were explored ranging from single-cell transcript-



The winner of last year's Anne Heidenthal Prize, Helena Radbruch, with Roland Eils (SBHD's Chair) to her left and Georg Draude (Chroma Technology Corp.) to her right. (Source: Franziska Müller)

omics to proteomics and fluorescence microscopy to conventional mathematical modeling and machine learning. Besides shedding more light on the novel coronavirus and its molecular mechanisms, these analyses aimed to identify potential therapeutic agents against the virus.

In addition to the study of COVID-19, these methods will of course continue to be used in research into diseases such as cancer, Alzheimer's disease, tuberculosis, and ALS. This means there were exciting findings to report here as well.

Yet not only projects on specific diseases were presented, but also projects such as the Human Lung Cell Atlas, which provide valuable insights into a wide range of medical issues. Alongside these extremely detailed datasets about a small number of individuals, another topic covered was their counterpart, population cohorts. These consist of a few select data points for a large number of individuals, enabling scientists to gain a further perspective on human diseases.

And last but not least, method development was discussed in detail at SBHD 2021 as it plays a fundamental role in current advances in systems biology. The methods covered included improved statistical methods, differential equation systems, classical machine learning, and deep learning. Recurring themes in these diverse methodological approaches were the identification and critique of underlying assumptions, as well as improvements in the evaluation and testing of the possible biases of these models.

A highlight of the conference certainly was the talk by **Helena Radbruch** of the Institute of Neuropathology at Charité on the entry of SARS-Cov-2 into the brain, research for which she was awarded the Anne Heidenthal Prize for Fluorescence Microscopy by Chroma Technology.

Interested in further events of the Berlin Institute of Health at Charité (BIH)?

Then subscribe to our **BIH Event Overview** here:
<https://www.bihealth.org/en/news/events/event-newsletter>

Offering almost 100 events per year, we would like to inform about our work in translational research and medicine.

Our goal is to translate biomedical research findings into novel clinical applications and to use clinical observations to develop new research ideas. Our approximately 400 scientists see it as their mission to bring relevant medical benefits to patients and the general population.

Always up to date...

With our monthly **BIH Newsletter** you will receive the most important information from the Institute. To subscribe to our newsletter, please follow this link:
<https://www.bihealth.org/en/news/newsletter>

BIH Berlin Institute
of Health
@Charité

Two additional prizes, sponsored by the journal Molecular Systems Biology from EMBO Press, were awarded that day for the best short talk and poster. The Best Poster Award, worth €200, was won by **Jens Hansen** of the Mount Sinai Health System in New York for his poster on the pathway network analysis of multi-omics data from the Kidney Precision Medicine Project, and the Best Short Talk Award went to **Brian Orcutt-Jahns** of UCLA for his talk on multivalency as a critical, unexplored axis of regulatory T cell-specific therapies.

In addition to these distinguished speakers, the program was enriched by many other excellent scientists, including among others, **Bree Aldridge** (Tufts School of Medicine), **Bernd Bodenmiller** (University of Zurich), **Monique van der Wijst** (University of Groningen), **Trey Ideker** (UC San Diego) and **Claudia Langenberg** (Berlin Institute of Health at Charité).

With presentations and discussions on method development, on data collection, and on specific research into individual diseases like ALS, tuberculosis, and COVID-19, SBHD 2021 was a fantastic kick-off for many in the community to return to in-person events. It once again became apparent just how important face-to-face events are for networking and scientific exchange.

We, along with the rest of the systems biology community, look forward to another successful edition of this popular conference, which is held June 19-22, 2022, in Nashville Tennessee.

Contact:

Franziska Müller

BIH at Charité – Center for Digital Health
franziska.mueller@charite.de
www.sbhdberlin.org

Roland Eils delivers the opening remarks at the conference dinner.
(Source: Claudia Böckermann)



events

A matter for the court

Do patients have a right to AI-based treatment?

by Matthieu-P. Schapranow, Hannelore Loskill and Klemens Budde

Clinical decision support systems (CDSS) incorporating artificial intelligence (AI) algorithms will be available to assist physicians in finding diagnoses and personal treatment decision in near future. Is there a patient right to consult such learning systems if patients make an explicit request for such assistance? The Plattform Lernende Systeme (PLS) teamed up with Leibniz Universität Hannover to explore this issue in a fictitious lawsuit based on a realistic case, which was held at the Center for Art and Media (ZKM) in Karlsruhe on September 22, 2021.

What is the issue in dispute?

The trial was based on the application scenario entitled “Using artificial intelligence to fight cancer” that PLS experts had developed. It employs a realistic medical case to demonstrate the potentials of learning systems in for optimizing cancer patient treatment. In the specific case a male patient named Anton Merk was recently confronted with the diagnosis lung cancer.

Mr. Merk is now looking for the best therapy for himself. His research led him to a pulmonary clinic specialized for tumor diseases, where he wants to have his tumor treated.

His treating oncologist does not only want to rely on her personal expertise, but also wants to make use of a clinical decision support system. This tool is able to analyze data of many similar lung cancer cases while making use of latest AI algorithms. Thus, CDSS can support the physician’s work by providing additional insights into similar cases. Mr. Merk’s treating oncologists has already experience in working with the AI-based CDSS.

Therefore, Mr. Merk insists that his oncologist is making use of the AI-based CDSS, but the management of the clinic rejects the use of the special software.

The clinic management fears the unclear liability situation in the event of a medical error or damage to Mr. Merk. In addition, the costs for using the AI-based CDSS are not covered by health insurance, yet. Furthermore, the management argues that the clinic would not be equipped adequately to effectively employ the software in medical care so far.

During an surgery, an AI-based navigation system assists Mr. Merk’s physician and reliably warns her, for example, if she gets too close to blood vessels – a big help for surgeons. Mr.



To support its work in Karlsruhe, the court questioned experts via live stream. Among the experts taking part was Dr. Matthieu-P. Schapranow of the Digital Health Center of the Hasso Plattner Institute in Potsdam (Photo: ZKM | Center for Art and Media Karlsruhe / Elias Siebert).

Merk also wants the AI-based CDSS to be used to assist selecting his personal medical therapy.

Mr. Merk is now suing in court to force the clinic to use the AI-based CDSS. The dispute centers on whether patients have a right to AI-based treatment.

Expertise in assessing AI recommendations is a key requirement

PLS experts from the fields of patient rights, artificial intelligence, medicine and ethics serve as expert witnesses to help guide the court in its decision-making.

Hannelore Loskill, the Federal Chairwoman of BAG Selbsthilfe, represented the patient's perspective and stressed that Mr. Merk's wishes should be at the center of any medical decisions so that he can be treated based on latest available medical knowledge.

Dr. Matthieu-P. Schapranow, lecturer and researcher at the Digital Health Center of the Hasso Plattner Institute, informed the court and those watching the trial about the technical capabilities of AI-based CDSS and the explicability of results obtained by such systems.

Prof. Dr. Klemens Budde, senior physician at Charité – Universitätsmedizin Berlin, served as the third expert witness and represented the position of medical professionals. In his opinion, AI-based support systems are an add-on that supplement a physician's knowledge and even provides access to the experiences of other doctors around the world. At the same time, he also pointed out risks posed by incomplete or biased data used for building such systems that might result in a less reliable prognosis. In the end, he said it will be Mr. Merk's oncologist, who will evaluate the recommendations proposed by the software and decide which treatment to pursue. Yet, Budde went on to call attention to the perils of "cook-

book medicine" in which physicians rely too much on AI-based CDSS. That is why it is important, he said, that doctors in the future know how to assess the system's recommendations as well as the probabilities and data on which they are based – and that humans have to make final decisions – not only in cases of doubt.

However, the use of AI-based CDSSs also raises ethical questions. **Dr. Jessica Heesen**, media ethicist at the University of Tübingen, pointed out that these systems could call into question the physician's competence and authority. Ultimately, who do people trust more – technology or their doctor? Who will decide in the future whether a medical decision is right or wrong? These are questions that cannot be answered conclusively by the court.

The verdict was a compromise for everyone

Even though many arguments were made in favor of using the AI-based CDSS, it was not possible to definitively resolve all issues. It also remains to be seen whether patients will have the right to receive treatment with the help of an AI-based CDSS in the future. On the stage of the ZKM, Mr. Merk and his lung clinic agreed to define an individual treatment contract. On the one hand, it limits the clinic's liability and on the other hand specified the costs Mr. Merk has to cover. Thanks to the individual treatment contract, the use of the AI-based CDSS is possible in this particular case.

The fictitious lawsuit format used here is a modern form of science communication. In an exemplary case, facts, arguments and the current legal situation are presented in a serious manner by experts playing the different roles. The event is geared toward interested citizens and requires neither legal nor medical expertise. The aim of the PLS's work is to raise awareness of the possibilities of learning systems, while critically reflecting on their use and clarifying the legal requirements that enable their deployment at the same time.

events

The trial was presided over by **Dr. Jessica Heesen**, who made sure that ethical aspects were not given short shrift by the court. **Prof. Dr. Susanne Beck** placed the possibilities of such a medical procedure in a legal context.

Links:

Further details on the event and the PLS's work as well as a video recording of the trial are available at:

www.plattform-lernende-systeme.de/streitsache.html

Kontakt:

Dr.-Ing. Matthieu-P. Schapranow is member of the Working Group on Health Care, Medical Technology, Care of Plattform Lernende Systeme, Working Group Leader "In-Memory Computing for Digital Health", and Scientific Manager Digital Health Innovations at HPI Digital Health Center, Hasso Plattner Institute, University of Potsdam.
schapranow@hpi.de

<https://hpi.de/digital-health-center/members/working-group-in-memory-computing-for-digital-health/dr-ing-matthieu-p-schapranow.html>

Hannelore Loskill is Federal Chairwoman of the Federal Association of Self-Help for People with Disabilities, Chronic Illnesses and their Relatives e.V. (BAG Selbsthilfe) and member of the Working Group on Health Care, Medical Technology, Care of Plattform Lernende Systeme.
info@bag-selbsthilfe.de

<https://www.bag-selbsthilfe.de/bag-selbsthilfe/ueber-uns/vorstand>

Prof. Dr. med. Klemens Budde is Co-Head of the Working Group on Health Care, Medical Technology, Care of Plattform Lernende Systeme and Deputy Medical Director of the Medical Department, Division of Nephrology and Internal Intensive Care Medicine at Charité – Universitätsmedizin Berlin.
klemens.budde@charite.de

https://nephrologie-intensivmedizin.charite.de/metasperson/person/person/address_detail/budde-1



Interested citizens were invited to participate in the evening event not only on site in Karlsruhe but also via internet livestream. Neither legal nor medical expertise was assumed (Photo: ZKM | Center for Art and Media Karlsruhe / Elias Siebert).

Virtual networking in systems medicine

e:Med Meeting 2021

by Ann-Cathrin Hofer and Silke Argo

The 2021 e:Med Meeting on Systems Medicine was held from September 20–22 and provided a welcome meeting place for the German systems medicine community. New features included community sessions and a wider variety of opportunities to network and communicate, which were expanded in response to feedback from the e:Med community. As has long been standard for e:Med, the 2021 conference once again covered a broad range of topics. A visitor-friendly virtual conference space for the event's 230 participants was provided by the e:Med@scoocs platform.

A live piano recital is not exactly the norm at a digital conference, which made the opening performance by Berlin-based pianist **Michael Nickel** all the more enthusiastically received by those attending the high-profile research event.

Dr. Lorna Moll (DLR-PT Bonn) of the funding body BMBF and **Dr. Matthias Karreman** (DKFZ and Heidelberg University Hospital), then co-spokesperson of the e:Med Project Committee, opened the conference with a welcome address, sharing motivational highlights from the German systems medicine community. Right from the start, e:Med 2021 participants enjoyed exciting presentations not only by e:Med researchers from current and first phase projects, but also by international scientists – all of whom were there to present and discuss their latest research and to expand their network.

Professor Philip Rosenstiel (UKSH Kiel) gave the opening keynote lecture, in which he presented his systems medicine research into the treatment of chronic inflammatory diseases, and provided an outlook on future clinical applications and further developments. On the second day of the conference, **Professor Heribert Schunkert** (DHZ Munich) explained in his keynote how genetic changes can be incorporated

SAVE THE DATE!

e:Med Meeting on Systems Medicine

November 28 - 30, 2022
Heidelberg, DKFZ



- In-person conference
- Poster exhibitions
- International keynote speakers
- Technology sessions
- Community sessions
- Flash talks
- Breaking news talks
- Company exhibitions
- Networking!

Find out more at:
www.sys-med.de/de/meeting



e:Med
SYSTEMS MEDICINE

SPONSORED BY THE



Federal Ministry
of Education
and Research



events



into the assessment of coronary heart disease. The final key-note speaker, **Professor Julie George** (University of Cologne), presented her current work on the molecular landscape of lung tumors with a focus on translational medicine.

Following seven thematic strands spanning methods, modeling, and translation, selected researchers shared their latest scientific findings in over 20 inspiring presentations, all of which prompted lively discussions from the interested audience. COVID-19 also featured – not in the form of restrictions, but as a focus of the presentations in the session “Systems Medicine of COVID-19.”

In the 20 flash talks, meanwhile, researchers had just two minutes to impress an audience with their scientific poster by presenting their work in a concise and creative way. Participants voted digitally for their favorite talks, resulting in **Lea Zillich** (ZI Mannheim) receiving the 2021 Creative Award – a voucher for a NaWik workshop valued at €275.

The virtual poster exhibition featured almost 50 posters offering insight into cutting-edge research from all fields of systems medicine. There was a very wide variety of content and graphic approaches on display – in formats ranging from classic posters to video and audio contributions. Five winners received an e:Med Poster Award 2021, each endowed with €150: **Dr. Ulrike Träger** (DKFZ Heidelberg), **Liza Vinhoven** (UMG Göttingen), **Dr. Raik Otto** (HU Berlin), **Robert Müller** (TU Dresden) and **Zhijian Li** (RWTH Aachen).

As well as enabling participants to watch the presentations and take a virtual stroll through the poster exhibition, the digital e:Med platform also provided networking tables where researchers could sit down with other participants to connect and share ideas. Targeted networking was further facilitated through individual profiles and the option to set up spontaneous video meetings. For the first time, the e:Med Project Committee initiated a community session on the future of systems medicine – a brainstorming event that brought together all participants to discuss their vision of the systems medicine of tomorrow. The conference was brought to a close by **Professor Rainer Spanagel** (ZI Mannheim), co-spokesperson of the e:Med Project Committee, who emphasized the vital importance of the BMBF’s funding and the generous support shown by longstanding sponsors of the e:Med meetings, Life & Brain and Illumina, as well as by partners de.NBI and NaWik.

The 2021 e:Med Meeting was a big success – as the participant survey also shows. However, there is still a great desire to meet in person. So we’re especially looking forward to seeing the e:Med community face to face next year from November 28–30, 2022, in Heidelberg.

Contact:

Dr. Silke Argo, e:Med Management Office
c/o German Cancer Research Center (DKFZ)
Im Neuenheimer Feld 581, D-69120 Heidelberg
s.argo@dkfz.de
www.sys-med.de
www.sys-med.de/de/meeting



Glimpses from the 2021 e:Med Meeting on Systems Medicine, incl. menu, dashboard, program, poster session, and video screenshot of talk by Dr. Matthia Karreman, then spokesperson of the e:Med Project Committee (Source: e:Med/scoocs).

imprint

gesundhyte.de

The magazine for Digital Health in Germany – International Edition Issue 14, June 2022

gesundhyte.de is an annually published magazine with information from German research in the fields of Digital Health and Systems Medicine.

ISSN 2748-3533

This issue is an adapted version of the already published German magazine.

Publisher:

gesundhyte.de is published by the Berlin Institute of Health (BIH) at Charité and the Project Management Jülich (PtJ).

Editors:

Editor-in-Chief: Prof. Dr. Roland Eils
(BIH at Charité)

Editorial Coordination: Franziska Müller
(BIH at Charité)

Editorial Team:

Dr. Judith Albrecht (BIH at Charité), Dr. Silke Argo (e:Med), Melanie Bergs (PtJ), Dr. Stefanie Gehring (DLR-PT), Prof. Dr. Lars Kaderali (University of Greifswald), Katharina Kalhoff (BIH at Charité), Dr. Marco Leuer (DLR-PT), Dr. Yvonne Pfeiffenschneider (PtJ), Dr. Matthieu-P. Schapranow (Hasso-Plattner-Institute), Dr. Thomas Schmidt (Plattform Lernende Systeme), Dr. Stefanie Seltmann (BIH at Charité), and Dr. Gesa Terstiege (PtJ).

Address:

Editorial office gesundhyte.de
c/o BIH at Charité
Center for Digital Health
Kapelle-Ufer-2, 10117 Berlin, Germany

The authors are responsible for the content of bylined articles. Unless otherwise stated, the authors hold the copyright to the accompanying photos and illustrations. The editorial board accepts no further responsibility for the content of URLs cited by the authors in their articles.

Design and layout:

Kai Ludwig, LANGEundPFLANZ Werbeagentur GmbH, Speyer (www.LPsp.de)

Translation:

Todd Brown, Berlin, Germany

Printed by:

Ottweiler Druckerei und Verlag GmbH, Ottweiler; Germany



Subscriptions:

The magazine is funded by the German Federal Ministry of Education and Research (BMBF). It is published as part of the public relations work of the publishers. It is supplied free of charge and must not be sold.

For subscription please visit www.gesundhyte.de or contact:

Editorial office gesundhyte.de
c/o BIH at Charité
Center for Digital Health
Kapelle-Ufer-2; D-10117 Berlin
franziska.mueller@bih-charite.de

systembiologie.de
becomes
gesundhyte.de!



Please visit our homepage if you would like to find out more about the topics of our magazine.

What you can expect:

- Exciting stories from **everyday research** – find out more about ongoing projects
- Interviews and profiles – get to know the **faces** behind the research
- Extensive overview of events – never miss another important **date**
- Information about current **funding** – stay up to date
- Get involved – suggest your own **topic**

We look forward
to your visit
on our homepage!



about us

Presenting the gesundhyte.de editorial team

gesundhyte.de aims to communicate the successes of the German research to a broad audience in a descriptive way. The magazine is created once a year in German and English by a multidisciplinary editorial team from various German research institutions: Berlin Institute of Health at Charité, Hasso Plattner Institute Potsdam, University of Greifswald,

Project Management Jülich, DLR Project Management and representatives of the initiatives: Lernende Systeme – the Platform for Artificial Intelligence in Germany and e:Med Systems Medicine. The magazine is financed by the Berlin Institute of Health (BIH) at Charité and the German Federal Ministry of Education and Research (BMBF).

Front row, left to right: Prof. Dr. Roland Eils (BIH at Charité), Franziska Müller (BIH at Charité), Dr. Judith Albrecht (BIH at Charité), Dr. Silke Argo (e:Med), Melanie Bergs (PtJ),

second row, left to right: Dr. Stefanie Gehring (DLR-PT), Prof. Dr. Lars Kaderali (University of Greifswald), Katharina Kalhoff (BIH at Charité), Dr. Marco Leuer (DLR-PT), Dr. Yvonne Pfeiffenschneider (PtJ),

third row, left to right: Dr. Matthieu-P. Schapranow (Hasso Plattner Institute), Dr. Thomas Schmidt (Plattform Lernende Systeme), Dr. Stefanie Seltmann (BIH at Charité), Dr. Gesa Terstiege (PtJ), and Kai Ludwig (LANGEundPFLANZ)



contact

Berlin Institute of Health at Charité

Editor in Chief: Prof. Dr. Roland Eils

Editorial coordination and subscription services: Franziska Müller

Editorial assistance: Dr. Judith Albrecht

c/o BIH at Charité

Center for Digital Health

Kapelle-Ufer-2; D-10117 Berlin, Germany

E-Mail: franziska.mueller@bih-charite.de

www.hidih.org

Communication and Marketing, Press Office:

Dr. Stefanie Seltmann

Anna-Louisa-Karsch-Str. 2; D-10178 Berlin, Germany

E-Mail: stefanie.seltmann@bih-charite.de

www.bihealth.org

Lernende Systeme – Germany’s Platform for Artificial Intelligence

Contact persons:

Dr. Thomas Schmidt and Dr. Ursula Ohliger

Managing Office: c/o acatech

Pariser Platz 4a; D-10117 Berlin, Germany

E-Mail: t.schmidt@acatech.de, ohliger@acatech.de

www.plattform-lernende-systeme.de

Project Management Jülich

Forschungszentrum Jülich GmbH

Universities, Innovation Structures and Health (HIG)

Contact persons:

Dr. Yvonne Pfeiffenschneider, Dr. Gesa Terstiege and Melanie Bergs

D-52425 Jülich, Germany

E-Mail: y.pfeiffenschneider@fz-juelich.de, g.terstiege@fz-juelich.de, m.bergs@fz-juelich.de

www.ptj.de

German Aerospace Center (DLR)

Project Management Agency

Department Health (OE20)

Contact persons:

Dr. Marco Leuer and Dr. Stefanie Gehring

Heinrich-Konen-Str. 1; D-53227 Bonn, Germany

E-Mail: marco.leuer@dlr.de, stefanie.gehring@dlr.de

www.dlr-pt.de

e:Med Management Office

Managing Director:

Dr. Silke Argo

c/o Deutsches Krebsforschungszentrum (DKFZ) – V025

Im Neuenheimer Feld 581; D-69120 Heidelberg, Germany

E-Mail: s.argo@dkfz.de

www.sys-med.de



INTERNATIONAL CONFERENCE ON

SYSTEMS BIOLOGY OF HUMAN DISEASE

JUNE 19-21, 2023

SAVE THE DATE

SEE DETAILS @:
www.sbhdBerlin.org

KAISERIN FRIEDRICH-HAUS
ROBERT-KOCH-PLATZ 7
10115 BERLIN, GERMANY

ORGANIZED BY:
ROLAND EILS
BERLIN INSTITUTE OF HEALTH (BIH) AT
CHARITÉ – UNIVERSITÄTSMEDIZIN BERLIN, GERMANY
VITO QUARANTA
VANDERBILT UNIVERSITY, NASHVILLE, USA
IRINA LEHMANN
BERLIN INSTITUTE OF HEALTH (BIH) AT
CHARITÉ – UNIVERSITÄTSMEDIZIN BERLIN, GERMANY

SUPPORTED BY: