Challenges and Opportunities of Digital Transformation in Fundamental Research on Universe and Matter

Recommendations of the ErUM Committees

[ErUM - Exploration of the Universe and Matter]

29 April 2019

Impressum

This document contains strategic recommendations
for the digital transformation in the research field
'Exploration of the Universe and Matter'. It was written
on behalf of and with input and support from the eight
committees listed below. The document was strongly
endorsed by all committees on 10 April 2019.

a. Komitee für Astroteilchenphysik (KAT)
b. Komitee für Elementarteilchenphysik (KET)
c. Komitee für Beschleunigerphysik (KfB)
d. Komitee für Forschung mit Neutronen (KFN)
e. Komitee für Forschung mit Synchrotronstrahlung (KFS)
f. Komitee für Forschung mit nuklearen Sonden und Ionenstrahlen (KFSI)
g. Komitee für Hadronen- und Kernphysik (KHuK)
h. Rat Deutscher Sternwarten (RDS)

Editorial board:
Prof. Dr. Martin Erdmann 1 (Lead Editor)
Prof. Dr. Christian Gutt 2e
Dr. Andreas Haungs 3a
Dr. Klaudia Hradil 4d
Prof. Dr. Thomas Kuhr 5b
Dr. habil. Marcel Kunze 6g
Prof. Dr. Anke-Susanne Müller 3c
Prof. Dr. Günter Quast 3b
Prof. Dr. Matthias Steinmetz 7h

Editor affiliations:
1 RWTH Aachen
2 Universität Siegen
3 Karlsruher Institut für Technologie
4 Technische Universität Wien
5 Ludwig-Maximilians-Universität München
6 Universität Heidelberg
7 Leibniz-Institut für Astrophysik Potsdam

Design and Layout: Britta von Heintze
Print service: DESY

Sources:
Fig. 1,3,4,5,6,7,8: Erdmann
Fig. 2: Haungs

CONTENTS

Executive Summary

Scope: The coming decade will witness new and transformative discoveries in the field of Universe and Matter ErUM, in particular thanks to the advent of large research infrastructures of the next generation. These facilities will cause an order of magnitude increase in the volume of measurement data that must be stored, managed and analyzed. Preparing for this data avalanche will therefore be the key to facilitating the scientific exploitation of the data and thus enabling fundamentally new insights. The pace of discovery is likely to even accelerate thanks to new approaches in machine learning and related technologies in Big Data Analytics.

Research in the field of ErUM brings together about 8,400 scientists from German universities and research centers working on astrophysics, astroparticle physics, hadron and nuclear physics, particle physics, accelerator research, as well as research with photons, neutrons and ion beams. Their scientific background originates in different disciplines, ranging from physics, chemistry, material sciences, biology to medicine. What unites ErUM scientists is research at leading national and international research facilities such as accelerators with their experiments and observatories. Researchers in Germany made numerous contributions to these infrastructures in particular also in terms of their computing and data analysis needs.

Scientific progress is closely linked to the accessibility, optimal use and further development of these large research infrastructures. The facilities have always been designed at the limits of what is technologically feasible. They thus serve both scientific advancement and the investment in the future through the training of young scientists, engineers and technicians. Together, the participating scientists shape the developments in science and technology, and at the same time the future perspectives of tomorrow's technological society.

Challenges and Opportunities: Beside the engineering challenges faced by these infrastructures, more and more data challenges are emerging as a limiting technological frontier. Using the facilities of the next decade will ultimately be determined by our capacity to read, reduce, analyze, process and mine the exabyte-scale data (1 exabyte = 1 million terabyte) readily and regularly produced by these facilities. Indeed, the data rates and associated computing needs of the next generation of national and international research facilities will outgrow the anticipated performance increase in the Information Technology sector, commonly dubbed as Moore's law (the performance of a computer doubles every 18 months), even considering that software developments contribute a similar improvement factor.

Big data and the ability to manage the data avalanche will thus be a prerequisite for the scientific exploitation of the next generation of research facilities, and thus for transformative or even disruptive new scientific discoveries. The development of innovative archiving and data curation methods combined with data-driven algorithms will play a decisive role in this endeavor. The latter form the language with which new scientific questions can be formulated and which are subsequently applied to data. These algorithms will be the key for acquiring new knowledge. Theoretical work that flanks all experimental activities with fundamental calculations, modeling, and simulations will be of similar importance. Within this general context, Big Data Analytics will form a new paradigm of scientific research very much like the computational astrophysics opened up new research avenues 60 years ago.

The transformational opportunities for new discoveries provided by Big Data and Big Data Analytics at large research facilities can only be exploited with transformational changes in the scientific landscape in Germany. These changes must occur in addition to the already existing framework and mesh-in on top of larger scale and more ground layer developments such as the overall National Research Data Infrastructure (NFDI) or the European Open Science Cloud (EOSC). New funding schemes are necessary to enable their development, thus enabling the exploitation of the full potential of data provided by the next generation facilities.

Measures: In a concerted effort by the research communities in ErUM research as represented by their respective eight committees, a coordinated action plan for the next 10 years detailing additional action areas has been devised over the past 18 months. We recommend a portfolio of measures that will be crucial to secure and further expand Germany's strong position in ErUM's research areas in a highly competitive

international environment and that will allow researchers in Germany to make maximum use of national and international research infrastructure, in particular also those on the European Strategy Forum on Research Infrastructures (ESFRI).

These measures include substantial ('beyond Moore') upgrades in computing power, storage and network capabilities. Furthermore, long-term personnel needs to be trained and financed that will have considerable expertise in computer science as well as technical and domain knowledge in the ErUM fundamental research areas. In particular we see the need for action in the following areas:

- **Federated infrastructures:** Expansion of the federated infrastructure beyond Moore to enable the repeated processing, mining and archiving of data volumes in the exabyte range including the required high-throughput network infrastructure.

- **Integration of workflows to exploit infrastructures:** Development of services and science-ready tools for an expanded federated infrastructure in order to support all demands specifically required for experiments in ErUM.

- **Comprehensive management of research data:** Design of the next-generation research data management, both for experimental and simulated data, enabling immediate usage and long-term, experiment-independent use cases.

- **Modern Big Data Analytics in fundamental research:** Integration of all aspects of modern Big Data Analytics and data mining into the successfully established ErUM research groups addressing experienced as well as young emerging scientists.

- **Scientists' integrated web working environment:** Set-up of innovative web-based interfaces to allow scientists to fully concentrate on the discovery process and the full exploitation of experiments and the data delivered by them.

- **Tenure-track programme knowledge in Digitization:** Launch of a tenure-track programme to promote research and education by leadership in the areas of algorithms, data mining and computing models.

and finally,

- **Partnership for Innovative Digitization:** Establishment of a grassroots-developed umbrella-partnership of Innovative Digitization that will foster the exchange between scientists across all disciplines, prioritization of measures, and exploration of new avenues for pioneering projects. This also includes training for and education of young scientists as well as regular workshops for specialists.

This portfolio of recommended measures is designed for broad impact and long-term success and will thus strengthen many areas in research and economy. Germany, like many other countries, faces the major challenge of efficiently adapting to a world driven by the new possibilities and opportunities of digitization. The success of this transition in research and development will be a prerequisite for the efficient training of highly qualified personnel not only for the research sector but also for personnel that will later engage in the economic world spanning the full range from start-ups to established globally renowned companies. In other words: a successful transition in ErUM-data research will be a relevant factor for ensuring Germany's continued economic success.

1 The digitization era in fundamental research

1.1 Impact of basic research

Fundamental research in the field of ErUM[1] has a long and on-going history of groundbreaking successes. We discover new particles yielding deep insights into the fundamental laws of physics, we observe and explore distant galaxies, stars and new planets, we investigate the structure and function of proteins, drugs and viruses, and we discover new materials and observe chemical reactions in real-time. With the continuous development and application of state-of-the-art technologies, the limits of what is feasible and what is still unknown have been pushed ever further.

Knowledge, knowledge leadership and exploitation are the forces that drive progress. Our modern society and life is based on this knowledge and the technology generated in basic research. The implementation of knowledge in industrial processing and services provides a basis for economic success and prosperity. As a result of training and research in ErUM, numerous young scientists, company founders and highly qualified employees originate from our field.

1.2 Prerequisites for new discoveries

Fundamental research is linked to scientific instruments which enable us to look deeply into matter, materials and the universe. In order to discover new rare, small or faint structures (new materials, new particles, galaxies, planets, etc.) and to investigate their dynamic processes, we need to take many more images than previously available - resulting in a much higher data rate. Here, images stand figuratively for all types of measurement instances obtained in ErUM research.

In the years to come all important experimental facilities will be geared towards these much higher data rates. In doing so the data volumes produced by new or upgraded scientific instruments will increase dramatically, in some cases by far exceeding the region of exabytes in the next five to ten years. Typically the computing power required scales by an even larger factor than the data volumes. Thus, the challenge to find new structures results in the demand to store, analyze, process, curate and archive huge amounts of data. To understand and analyze the experimental data, theoretical work is required simultaneously with instrumental advancements. Much of the calculation and simulation work is performed on large-scale computing infrastructures. Their requirements are also expected to increase accordingly.

1 Exploration of the Universe and Matter - ErUM; Framework programme of the Federal Ministry of Education and Research

1.3 Big data science in basic research

These massive data volumes and exploitation thereof, known in short as *Big Data* and *Big Data Analytics*, constitute a key aspect of knowledge generation. To exploit the full potential of these large amounts of data, fast and efficient algorithms are required. They form the language used by scientists to formulate their questions about data in order to gain in-depth knowledge, e.g., about unknown structures or extremely rare processes. This data-driven multiplication of knowledge depends on an efficient transformation from Big Data to Smart Data which increasingly relies on methods from artificial intelligence in conjunction with in-depth domain expertise. Such methods are further developed by scientists to meet their specific requirements and may lead to innovations not only in basic research, but also in artificial intelligence.

1.4 Scientists' institutions and working environments

In Germany, research in the field of ErUM is performed primarily by scientists at universities and research centers of the Helmholtz Association, the Leibniz Association and the Max Planck Society.

Here, classic ErUM research has been carried out by an individual or a small group of scientists, planning and executing an experiment, recording measurements in a log book, analyzing them, and using the thus obtained knowledge to repeat the experiment with refined parameters. With the high data rates and volumes delivered by modern experimental devices and the required short feedback time to control them, this way of performing research is no longer possible. User-oriented digitization is mandatory to obtain and maintain a leading role in a highly competitive international context. Nowadays, the vast majority of scien-

tists work in teams, both within their own institution and with - usually international - partner institutions. Depending on the disciplines, the number of collaborating researchers varies between a few partners and very large collaborations with up to several thousand colleagues.

What they all have in common is the need for straightforward access to smart research data, computing resources, methodologies, publications, expert knowledge and communication with research partners (Fig. 1). Modern digitization offers the opportunity to establish the necessary technology and to put it into action.

1.5 Scientific computing in ErUM

Scientific research in ErUM spans a wide range of very different domains and the supporting scientific computing is powered by a large number of actors and institutions.

Computing for synchrotron radiation and neutron research is essentially served by the universities and infrastructures where the instruments and accelerators are located and the experiments are conducted, e.g. at centers of the Helmholtz Association (HGF).

The typically large data analytics requirements in experimental particle and nuclear physics are handled by a distributed high throughput computing model in a concerted action of HGF computing centers, and centers at universities and the Max Planck Society (MPG). The astroparticle physics community leverages computing centers of the HGF, MPG, and the Leibniz Association.

Figure 1: In the 2020's, the scientists' research environment will look and feel like working in a local environment, allowing access to large-scale data volumes and massive computing resources to process the data, as well as interconnecting with colleagues.

For most areas of experimental ErUM research, simulations are essential to understand the measurements at the required depth. They are performed in all above-mentioned centers. Computing power furthermore plays an important role at the laboratory sites in accomplishing the acquisition of data. For theory and special simulations, high-performance computing resources are indispensable to perform cutting edge research: the corresponding infrastructures additionally include the centers of the Gauß-Allianz.

1.6 Survey of areas with development needs

The development of innovative algorithms and the coherent use of originally unconnected data have sparked a highly competitive international race. Far-reaching advancements in the German research landscape are indispensable in order to be prepared for the upcoming changes that will increasingly result from such data-driven methods and that will offer a multitude of new opportunities.

In order to handle the huge amounts of data and their exploitation (Fig. 2), a crucial factor is the infrastructure, which simultaneously enables the processing of data volumes up to the exabyte range with very high throughput and network requirements. Another challenging area is mastering multiple usages of experimental data, cross-experiment and cross-disciplinary data exploitation which have the potential to generate new knowledge. Moreover, the creation of core competencies in dealing with the modern possibilities of Big Data Analytics will be indispensable.

Scientist with in-depth domain knowledge of ErUM research will be needed to harness the infrastructures, interconnect their research data and develop algorithms for Big Data Analytics. All of this work is necessary to achieve the transition to increased use of data-driven research and to enable new insights into their own research questions and those of other scientists. It will be essential to spread these scientists with their knowledge in order to cope with the digital

evolution and to reach ErUM groups at all universities and research centers.

As practically all fundamental research is today performed in an international setting, any national effort needs to be considered in a global context. Also, as many of the research techniques have a strong link to computer science, information technology expertise, and mathematics, the measures to be taken will have to be linked to these disciplines in the form of close cooperations. Because Big Data and their analysis are global challenges, cooperations with the private/economic sector and with small and medium-sized enterprises (SME) may also arise.

1.7 Workshop and report of the outcome

With the workshop 'Challenges and Opportunities of Digital Transformation in Fundamental Research' (4 and 5 October 2018), the BMBF invited representatives of basic research from universe and matter (ErUM) to come together with representatives from computer science and business.

The participants represented the different fields of ErUM research. They were drawn from the eight standing committees in the respective ErUM research field and joined forces to advance the required developments. Together, these eight research fields employ around 8,400 active scientists with a doctoral degree. The workshop participants are listed in Appendix B. A short list of the committees and the fields they represent can be found in Appendix A.

The ErUM programme has excellent potential not only to prepare for the changes triggered by modern digitization, but also to initiate key measures that will enable a healthy bottom-up development of the research landscape in this era of digitization. This report summarizes our findings and presents the portfolio of recommended measures.

We will first report about the discussions and results arising from the three working groups:

1. Working Group on Federated Infrastructures
2. Working Group on Big Data Analytics
3. Working Group on Research Data Management

We will then summarize the recommended measures together with estimations of the costs associated with the programme as a whole.

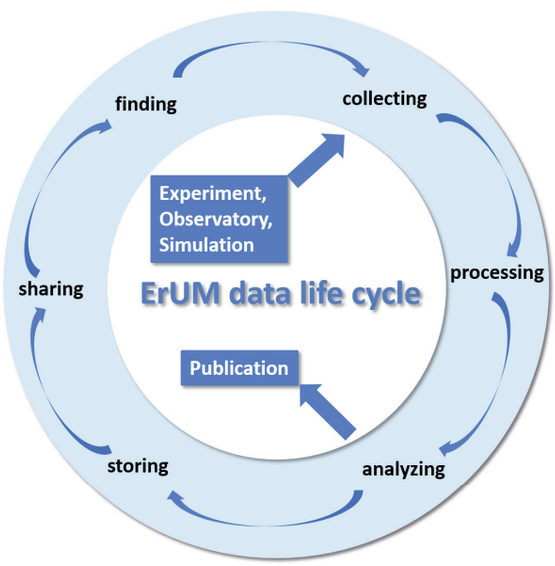

Figure 2: The 'ErUM data life cycle' describes all phases from data collection in experiments to data processing, analysis, storage, sharing and finding.

2 Working Group 1: Federated Infrastructures

2.1 Introduction

Fundamental research in the field of Universe and Matter requires massive computing resources to store, process, distribute, analyze and curate present and increasing data volumes and to perform simulation tasks of increasing complexity. New and powerful analysis techniques like deep neural networks further increase resource demands and benefit from special processor architectures. The need for computing and storage resources increases much more rapidly than technological progress alone can satisfy. Therefore, it is clear that now and in the future, both the further development and expansion of infrastructures and the investment in the development of the appropriate software to use the hardware must go hand in hand.

The computing landscape is presently undergoing a drastic change: dedicated, community-specific resources are complemented by common infrastructures, and techniques are being developed to make new types of resources available, ranging from high performance computing centers over publicly financed research clouds to commercial cloud providers. Efficient utilization of such heterogeneous, federated infrastructures poses a significant challenge to the research communities.

The goal of this working group is to identify needs, propose concepts and concrete steps towards the provisioning and effective usage of federated and demand-oriented infrastructure(s) for the analysis and management of research data in the field of ErUM. This concerns the future availability and efficient use of hardware resources, such as computing capacities, storage space and fast network connections, including intelligent database solutions and cloud services. The dedicated development of operating models and software is equally important to use the different hardware resources easily, transparently and efficiently. In this context, both national and international (European) developments and initiatives need to be considered.

2.2 Status and requirements

Input from the communities was requested to assess the type of computing infrastructures and resources being used presently, the near and long-term future demands, the existing experience with cloud technologies and the required development work.

This detailed survey paints a rather diverse picture, including very specific as well as more general, common requirements. Improved instruments and increasing depths of analyses will lead to a major increase in computing needs in many international research collaborations in the coming years, that go beyond the performance gains expected from technological progress (‘Moore’s law’). Better algorithms for data processing, simulation and analysis (see working group 2) and the adaptation of software for the better utilization of modern processor hardware will have a mitigating effect on demands for hardware.

Nevertheless, the current computing capacity consisting of the well-established centers for High-Throughput Computing (HTC), High-Performance Computing (HPC) and infrastructures located directly at the (raw) data sources need to grow significantly. They should also be augmented by additional hardware resources, such as those provided by computer centers at universities and science institutions or by commercial cloud providers, to ultimately meet the demands.

Such a heterogeneous environment poses significant challenges for the design of future computing models, software development and workflow and data management. For some use cases, successful mapping onto common, federated infrastructures has already been demonstrated. Specialized computer architectures, e.g. equipped with accelerating co-processors, are already being exploited by many communities, mostly in installations in proximity to the scientific working groups. In the future, such resources need to be provided more centrally in shared computing centers.

Resources on science clouds are of high relevance for many applications within ErUM. However, substantial R&D work is required to ensure seamless integration into the processing and data management workflows (see working group 3). Encapsulation of computing requests in virtual environments or containers is an important technique that is already common in some communities, while others need support to catch up. It should be noted, however, that existing HPC systems and cloud services typically do not provide

cost-efficient permanent storage space for large data volumes. Bringing data to the computing resources and distributing processed data to the places where scientists work requires a nationwide fast network connecting data providers with universities and research centers.

The following general list of goals and requirements is derived from the input of all ErUM communities:

- Fast access of analysts to data relevant for analysis is of prime importance.
- Sufficient resources of adequate type for data processing (i.e. the derivation of high-level objects for analysis from the raw data) and for simulation tasks as well as Big Data Analytics and theory calculations must be provided.
- Open access to research data is a concern that was voiced by the entire research community (see also working group 3). This requires high-performance storage systems to ensure efficient access.
- Data archive(s) to permanently store valuable scientific data and metadata, accompanied by the archiving of virtual or containerized software environments to ensure mining and analyzability (data preservation) at a later point in time.
- Education in data analysis and data management is essential to enable young researchers to efficiently work in necessarily more complex environments involving large federated infrastructures.

The advancement of information and communication infrastructures requires a longer-term (> ten-year) commitment. A combination of improvements in software, algorithms, workflow and data management and increased efficiencies will, however, only partly mitigate the lack of hardware resources resulting from growing data volumes.

The complementation of community-specific resources by federated infrastructures clearly is an important ingredient to optimize benefits in relation to efforts and costs. We therefore recommend the following elements to be considered to enable the usage of federated infrastructures by research groups in the field of Universe and Matter:

- Appropriate financing for the necessary expansion of computing and storage resources at large-scale research centers and universities.
- R&D work to establish community-overarching workflows and processes ('Scientific IT Services') in close coordination with the partners involved.
- Development of efficient (cloud-based) operating models and enhancements of software frameworks to maximize the range of usable infrastructures.
- Creation of a federation of computer centers within the framework of a national or dedicated 'ErUM Science Cloud'.
- Integration into national concepts and international endeavors such as EOSC.
- Funding of development and implementation of new computing models for future experiments that support easy access to federated, heterogeneous resources.

2.3 Conclusion

The future success of research in all ErUM communities depends on efficient and sustainable federated information and communication infrastructures. These should also serve as building blocks for and benefit from national and international initiatives. Based on the above considerations and on the lively discussions in the working group we summarize as follows:

- Secured, long-term funding of an agile information and communication technology (ICT) in terms of a federated ErUM-ICT infrastructure is required. The infrastructure(s) must meet the future challenges in terms of storage (German share of data volumes of exabyte per year), network (high-bandwidth network with 100 gigabit per second in the short-term and terabit per second in the long-term) connecting all ErUM centers, computing power (specialized as well as federated resources) and services.

- Long-term funding for software development and competence for the efficient utilization of the federated infrastructures is needed. After initial development work to enable the ErUM research groups to efficiently use federated resources, funding for long-term or permanent positions is needed to ensure support of tools and services and a sustainable knowledge transfer to young scientists. These efforts should result in the installation of a national and international competence network in Big Data management.

3 Working Group 2: Big Data Analytics

3.1 Introduction

Data is valuable because we can derive scientific insights from it. This is achieved by processing the data with sophisticated algorithms. For a few years, the set of available data processing tools has been rapidly increasing. Modern Big Data Analytics methods, like machine learning and others, provide new opportunities to obtain scientific results from large datasets. ErUM scientists have to learn how to exploit these new methods for their research.

Big Data Analytics methods are also indispensible to solve the challenge of the ever-increasing data rates resulting from the quest for studying structures with much higher spatial and temporal resolutions. There is a growing need of the research community for new tools for an efficient analysis of Big Data sets. This comprises both on-line analysis during data acquisition to control the experiments and reduce the data rate, as well as the in-depth analysis of huge data sets later on. Big Data Analytics is a common challenge for all communities in ErUM and calls for a collaboration with other disciplines such as mathematics and computer science in order to develop and implement new analytic tools.

The challenge is also recognized and addressed at the international level. For example, the particle physics community has formed a High Energy Physics (HEP) Software Foundation, and in the US the National Science Foundation has funded the Institute for Research and Innovation in Software in High Energy Physics (IRIS-HEP) with 25 M€ for five years to develop sustainable, experiment-overarching solutions.

Expertise in modern software development techniques is required to implement algorithms that optimally exploit and manage the hardware resources provided by the federated infrastructure. This includes not only data processing algorithms, but also simulations and theoretical calculations. New architectures, like GPUs or FPGAs, and new methods, such as machine learning, at the same time represent a challenge and an opportunity for Big Data Analytics and the federated infrastructure (see working group 1). Only with sophisticated data reduction and processing methods will we be able to handle the huge data volumes and rates without losing information that is relevant to the science we want to explore (see working group 3). This can only be achieved if Big Data Analytics competence is combined with knowledge about the scientific merit of the data. As the evolution of technologies is expected to continue rapidly, it requires a continuous effort to fully exploit the state-of-the-art Big Data Analytics methods.

There are numerous examples in the ErUM-related literature of boosted scientific output by the application of modern data analysis techniques, like the doubling of the sensitivity of several analyses of data collected by the Belle experiment by an improved particle reconstruction algorithm. Another example is the extraction of chemical abundance from measured spectra by deep neural networks, which performs better than a classic fit and led to reduced signal-to-noise requirements for future surveys. In the analysis of X-ray protein crystallization images, deep neural networks supported the effective automatic screening of protein solutions for successful crystallization. However, these generally are achievements of individuals or groups that have acquired Big Data Analytics expertise by self-education in addition to their competence in their scientific field.

While Big Data is a challenge, it also offers opportunities, as the large datasets provide sufficient input for the training of modern machine learning algorithms. ErUM offers many challenges to Big Data Analytics that allow the training of young researchers on real problems and may lead to applications in other areas or innovations in the field of Big Data Analytics in general.

The goal of working group 2 is to identify the communities' status and requirements in terms of Big Data Analytics. Potential common approaches and overlaps between the committees in terms of Big Data challenges are to be identified and efficient funding schemes for addressing the needs have to be discussed. Cross-correlated action in terms of Big Data Analytics including the training and education of the next generation of scientists needs to be addressed.

3.2 Status and requirements

We asked the communities about the status and requirements of Big Data Analytics using a catalogue of questions on the following topics:

- Questions to be solved and expectations for Big Data Analytics
- Data level for Big Data Analytics
- Data volume for Big Data Analytics
- Analysis methods for Big Data Analytics
- Existing and expected collaborations in Big Data Analytics

The outcome of this questionnaire is that the research questions to be addressed vary considerably between the communities, ranging from questions relating to fundamental physics in astrophysics and particle physics to questions relating to material science and biology in neutron and synchrotron research. Equally diverse are the past experiences and expectations for the future with methods of Big Data Analytics. Some communities have extensive experience in methods such as machine and deep learning, and expect these methods to become even more important in the future. Other communities, on the other hand, rely heavily on inverse methods based on data models. However, the common challenge for all communities is to further develop tools for Big Data Analytics.

While data levels, structures and formats very much differ as well, all communities stressed the need for quick online data analysis and visualization of experimental data. In some research areas the non-capability to quickly analyze large amounts of data soon develops into a serious bottleneck when executing experiments.

Data volumes to be handled reach from terabytes to petabytes (foreseeable exabytes), and data rates are approaching values between 10s and 100s of petabytes per year. Also, methods of Big Data Analytics vary between the communities.

Some of the communities have already established cooperations within ErUM and also with mathematics and computer science, while others see the potential of such collaborative efforts but have not yet engaged in it.

In most of the groups, data analytics is organized on a classical doctoral student university level with a foreseeable short-term rotation of personnel leading to a loss of knowledge and making long-term developments and curation of tools difficult. Lack of sustainability is identified as a common problem that can impede progress. Collaboration with mathematics/computer science is rare, occasional and presently not coordinated on a larger scale. The experts from mathematics stressed the need for specific scientific projects/questions to be tackled in common collaborations.

3.3 Conclusion

In conclusion, the research questions, approaches and Big Data Analytics tools needed for the different communities are rather diverse. Given the fact that ErUM data deals with questions from particle physics, hadron and nuclear physics via condensed matter science to astrophysics, this does not come as a surprise. However, working group 2 clearly identified the following common needs:

- Development and implementations of tools for Big Data Analytics and data management, aiming for common solutions wherever possible and exploiting existing structures and competences.
- Need for a collaborative effort in terms of Big Data Analytics including users, facilities, mathematics and computer science; it was suggested that the collaboration should be organized in a bottom-up approach.
- A platform for sharing Big Data Analytics solutions within and even across communities.
- Integration with data management (e.g. for efficient data access or mining archived data) and federated infrastructure (e.g. for utilizing resources optimized for Big Data Analytics tasks).
- Training and education of the next generation of scientists in Big Data Analytics; a tenure-track programme would enable this in a time frame that is appropriate for the upcoming challenges.
- Ensure sustainable development and curation of algorithms and tools.

These needs partially overlap with the results from working group 1 and working group 3 and shall be addressed in a common funding scheme.

4 Working Group 3: Research Data Management

4.1 Introduction

Upcoming challenges for research data management will encompass metadata findability and data curation to ensure the long-term use of science data with an archiving method that allows for reasonably fast access. Sustainable and systematic access to research data will increase the potential for break-through science by allowing for cross-experiment as well as cross-disciplinary approaches. Accordingly, Research Data Management plays an important role in shaping the future ErUM landscape.

Sophisticated Big Data Analytics tools will be employed to extract science results from large data sets, requiring complex simulation and modeling of experimental setups and physics processes (see working group 2). Intelligent data management strategies will be required to facilitate efficient data processing/reduction and to allow for seamless data access by scientists using the federated digital infrastructure (see working group 1). Hence, it will open a window for the reusability of research data (data mining). Doing this correctly may allow us to answer scientific questions that transcend the initial scope of the experiment in the future.

Commonly accepted standards for data policies and regulations for ownership and access are a prerequisite for managing data on shared infrastructures. Guided by the FAIR (Findability, Accessibility, Interoperability, Reusability)[2] principles for scientific data management and stewardship, the different communities within ErUM need to develop effective procedures to address their data management challenges. The definition of research data management procedures and the development of software to implement these procedures on a federated digital infrastructure are considered key elements to current and future ErUM research. This process must be included within other on-going initiatives, e.g. the set-up of the National Research Data Infrastructure (NFDI) funded by Germany's federal and state governments. The inclusion of large-scale facilities in international projects calls for an international setting for this process. First cross-experimental agreements already exist for some areas, e.g. in astronomy within the framework of GAVO (German Astrophysical Virtual Observatory) and the IVOA (International Virtual Observatory Alliance). Likewise the synchrotron and neutron communities received funding for a Photon and Neutron open science cloud (PaNOSC) from the European H2020 program for contributing to the European Open Science Cloud (EOSC).

2 M. D. Wilkinson et al., Scientific DATA: The FAIR Guiding Principles for scientific data management and stewardship, 3: 160018, 2016; DOI: 10.1038/sdata.2016.18

4.2 Status and requirements

The discussion of the data types across the communities revealed different data types. The large-scale experiments at international facilities often have well-established computing models that detail the data processing and management strategies for their large-scale data sets. Small or medium-sized experiments from multiple disciplines at national and international facilities produce medium-sized data volumes. And dedicated small-scale single use-case experiments often yield highly specialized data-sets. In addition, interoperability and reusability of data requires the preservation of all relevant metadata, including not only the data and experiment or simulation description, but also the analysis tools and libraries used. Thoroughly prepared and documented data schemata and metadata are necessary to answer future scientific questions and to enable the exchange of open data across communities. Activities to achieve this aim is an important contribution to the setup of a national research data infrastructure (NFDI).

In preparation for the workshop we developed a questionnaire about the following subjects to assess the status and requirement concerning research data management in the different communities:

- Existence of metadata in consideration of the FAIR Guiding Principles for scientific Data Management
- Work flows for data management
- Disposition to transfer data to an inter- or disciplinary research data center for archiving and publication
- Standard procedures for data reduction/handling of raw data

As anticipated due to the wide variety of research fields in the communities, the outcome of the survey was varied, in particular with respect to the level

of established workflows and data policies. Large international communities in astro-, particle or nuclear physics traditionally have established computing models that detail their strategies for handling the Big Data problem. They have experience with distributed data management procedures, sophisticated processing chains from the detector to the analysis stage, and with operating distributed storage and archive systems. The more heterogeneous communities with mostly dedicated single use-case experiments carried out by scientists from various research fields (i.e. physics, chemistry, biology, geoscience or even heritage science), where the challenges lie in the complexity rather than in the size, are only now entering the era of large-scale scientific data management. With the expected increase in data rates from experiments at current and new facilities, all communities face the need to further develop their data management procedures.

The 'ErUM Data Life Cycle' in Fig. 2 describes all stages, from the data collection during experiments or observatories, to data processing, analysis, storage and sharing.

Research scientists need to collaborate with qualified computer scientists, software developers, mathematicians and data stewards to develop thoroughly adapted software systems for data management and metadata handling. Common approaches within and across communities would provide an environment for the efficient interchange of data and algorithmic approaches.

All communities agree on the FAIR Guiding Principles for scientific data management and curation, which aim at generating open data that is re-usable and inter-operable, but follow different approaches to achieve these goals. Different communities with a large range of instruments and small scientific teams strive to follow these principles. Best practice examples might be a step towards this goal. Observatories and other astronomical instruments publish their processed data for open access after a short period of time. Large-scale experiments share their data world-wide through international collaborations, before making them fully accessible to all. All communities require more refined data management procedures to address their future needs, which at the same time would facilitate the implementation of FAIR principles.

Likewise, the current approaches for handling metadata differ between communities due to the variety of experimental instruments and methods. The metadata content ranges from experiment/simulation description, to data provenance and information for analysis preservation. Solving the metadata problem across our communities requires careful consideration.
A 'one-size-fits-all' approach does not seem appropriate. However, the identification of common issues within and across communities will be beneficial. The development of metadata systems to store all relevant information necessary for data re-use and exchange needs to consider low-level metadata (untreated, raw data) as well as reduced (treated) and high-level metadata ('smart data'). For the more heterogeneous communities, common standards for metadata need to be developed and established to allow for increased interoperability.

4.3 Conclusion

Even though the survey revealed a wide range of current practices for data management and approaches to metadata collection, all communities face similar challenges to address their future data management needs. The development of improved data management procedures and of systems for handling metadata is a requirement for all communities across ErUM to manage the data life cycle on a future federated computing infrastructure in the area of exa-scale data volumes.

- Some of the more heterogeneous communities would benefit from a more developed university - large-scale facility collaboration in research data management.
- Where possible, common standards should be established to promote interoperability.
- Future developments will benefit not only from the experience from established large-scale data management systems within our communities, but also a close collaboration with computer scientists, software developers, mathematicians and scientists from other fields.
- A comprehensive web working environment is needed to enable seamless data access for researchers to carry out their scientific analyses.

- The working group underlines the importance of ‘data stewards’ to manage the data life cycle and to act as curators for metadata. People with such specialist expertise, either at large facilities or within the experiment collaborations, are needed to support the science data management in their communities and to develop the systems for implementing the data management procedures.
- The complexity of managing future exa-scale data volumes and the data of heterogeneous communities, as well as the needs for data preservation, reusability and preparing for open access, require a long-term funding commitment to keep such experts in scientific position. The training and education of specialists in data management as well as of scientists working with data management infrastructures has to be taken into account.

5 Recommended measures and cost estimates

The discovery of new rare, small or faint structures and their dynamic processes require much higher data rates. Over the next five years, all major experimental infrastructures will therefore drastically increase their data rates.

Based on the surveys conducted within the participating communities and the subsequent analysis during the ErUM workshop, we synthesized and propose a portfolio of actions that we consider crucial to secure and to further expand Germany’s strong position in the research fields of ErUM in an internationally competitive environment. Simultaneously, these measures will also help strengthen and broaden Germany’s position in key areas of the modern information technology and digitization era.

Our proposal brings together the entire ErUM research community of 8,400 scientists and addresses common challenges of Big Data in a coordinated effort. By bridging the gaps between various disciplines in the sciences, spanning the entire range from physics at accelerators to deep- and all-sky observations, a critical mass is overcome, thus allowing for a sustainable push into the next decade of data analysis. This holistic approach is a unique feature of this initiative.

We propose a coherent and interlinked portfolio of measures. We should like to emphasize that all measures rely on an already working science environment provided by the many stakeholders in ErUM. The measures are required in addition to all current resources. In the following we classify and describe the individual measures separately and emphasize that only the concerted action of all measures will lead to an overall success.

- **Federated infrastructures:** Owing to the rapid increase in data volumes in the years to come, the demand for computing power, storage space and networks will increase substantially. We see an immediate need for action in the expansion of the existing infrastructures.
- **Integration of workflows to exploit infrastructures**: This hardware needs to be made usable according to our scientific computing requirements. To be able to use the capacities of the infrastructures, we recommend the creation of additional positions for scientists with ErUM domain knowledge.
- **Comprehensive management of research data:** Research data contain a high potential for added value, especially for multiple exploitation, for cross-experiment and cross-disciplinary usages. This requires numerous ErUM-specific measures (metadata, curation, etc.), for which we recommend additional scientific positions.
- **Modern Big Data Analytics in fundamental research:** The large amounts of data open up new possibilities for Big Data Analytics, which can extract possibly decisive new information. To enable a large number of experiments to be covered, we recommend additional positions for scientists with ErUM domain knowledge and expertise in Big Data Analytics.
- **Scientists’ integrated web working environment:** To increase the efficiency in scientific work, a new integrated web working environment is needed that

looks and feels like local work but enables access to globally distributed resources. To facilitate this development, we recommend additional scientist positions from various disciplines.

- **Tenure-Track programme knowledge in digitization:** Scientific leadership positions are urgently needed for cutting edge ErUM research related to algorithms, computing and data models as well as for the distribution of knowledge of modern digital methods. We recommend the establishment of a tenure-track programme.
- **Partnership for Innovative Digitization:** The different research directions and specializations in the ErUM field also bear the challenge of cross-discipline communication and interaction. We propose the creation of an umbrella organization which will form the core institution for the exchange between the scientists.

These measures will be explained in more detail in the following sections.

5.1 Federated infrastructures

We recommend an increased investment in computing, storage and networking resources to meet the needs of basic research in the field of Universe and Matter to process, distribute, analyze, curate and archive increasing amounts of data and to perform complex simulation tasks. Even if the corresponding technology is progressing rapidly, this alone is not sufficient to cover the ErUM communities' requirements arising from the more complex data provided by the experimental facilities[3].

Challenges: The coming years will see steeply increasing data acquisition rates and more complex data from facilities in many areas of basic research that relate to ErUM. This applies in particular to particle physics at the LHC (Large Hadron Collider at CERN) and Belle II (Japan), astroparticle physics at CTA (Cherenkov Telescope Array, Chile and Canary Islands) and IceCube (South Pole), astrophysics at the facilities of ESO (ELT, VLT and ALMA, Chile), LSST (Large Synoptic Survey Telescope Chile), MeerKAT (South Africa) and SKA (Square Kilometer Array, South Africa and Australia), as well as hadron and nuclear physics at the LHC and at FAIR (GSI in Darmstadt). At the same time, the requirements for almost simultaneously running data analyses on large amounts of data recorded at photon and neutron sources (e.g. European XFEL, FLASH, PETRA III in Hamburg, BESSY II Berlin, ESRF Grenoble, ESS in Lund, Sweden) or during the commissioning of accelerators (e.g. KARA at KIT, MESA in Mainz, the Helmholtz ATHENA facilities) are also increasing. This data avalanche and the corresponding need for extended simulations can only be mastered by sustainable, distributed computing systems.

Goal and resource providers: The goal is to establish a cost-effective, sustainable and efficient federated information and communication infrastructure (ICT) for research in the field of ErUM. The infrastructure also delivers building blocks of broader international (EOSC, European Open Science Cloud) initiatives. It will offer future-oriented solutions for world-class research.

The ErUM communities presently use computing and storage capacity at a variety of computing centers, ranging from computing resources directly located at the (raw) data sources to infrastructures at High-Performance (HPC) or High-Throughput (HTC) centers provided mainly by the Helmholtz Association and the Gauß Alliance of German HPC centers. Additional resources are provided by universities and the German federal states, the Max Planck Society or by computer clusters funded by the German Research Community (DFG) or directly by the German Ministry of Education and Research. As forerunner, ErUM scientists successfully demonstrated the exploitation of resources in commercial cloud environments to handle peak loads. Computing infrastructures in Germany should support a larger variety of user communities in the future and develop into federated infrastructures. New resources in science clouds should be made available, and ErUM scientist must be enabled to easily access these resources and to use them efficiently.

Experimental and theoretical research groups have access to computing installations (Tier 3 analysis clusters) funded by their home institutions. For university groups, shared funding by the DFG and the universities and computing resources provided by HPC centers

3 We use the term *detector* or *facility* synonymously for large-scale experiments and instrumentation (e.g. detectors, telescopes, cameras, sensors, spectrographs, receivers) at large research infrastructures.

of the Gauß-Allianz are important for local end-user analysis and theory calculations. Requirements for these resources are expected to increase by an amount which might be slightly more than technological progress alone will provide. Any additional investment in these infrastructures that may be required is not included in the following cost model.

Cost model: Improved algorithms, adjustment of computing models and increased efficiency in using existing resources are just one ingredient needed to face the challenges ahead. However, significant extra funding is necessary to ensure the optimal scientific harvest from future research data and novel analysis techniques. No solid figures for expected additional needs for investments are currently available for all communities.

However, as one example, the approved roadmap for the high-luminosity upgrade of the Large Hadron Collider at CERN provides a rather precise prediction for the required data storage, which will increase by a factor of more than ten compared to the present volume after the year 2025. At present, the investment in annual hardware upgrades for the German contribution to the computing resources integrated in the central workflow systems (so-called Tier 1 and Tier 2) of the particle physics experiments amounts to 4.1 M€ per year. The operational costs of this infrastructure, not including the 40 site administrator positions, amount to an additional 2.7 M€ per year. A further example is the ALICE experiment for research on hadrons and nuclei, for which the annual invest and operational cost at Tier 1/Tier 2 sites are about 1 M€ and 0.7 M€. All these Tier 1/Tier 2 costs are complemented by hardware and operating costs for the analysis clusters at universities (Tier 3) at the same level.

Taking into account technological progress, advancements in algorithms and the adaption of computing models, funding for hardware at Tier 1 and Tier 2 sites will still have to increase substantially, by an estimated factor of two to three, compared to present levels. Pessimistically assuming that a factor of three will be needed, additional funding for hardware and its operation of 14 M€ per year on top of the current level of 6.8 M€ will be needed for the particle physics experiments after the year 2025.

An example of emerging investment needs is astroparticle physics, where a cross-observatory analysis and data center is to be set up in the coming years. Structurally, this would be conceivable as an extension of a Tier 1 center with a hardware requirement of approximately 1 M€ per year.

The cost model does not include the funding of the university Tier 3 analysis centers. Neither included is additional funding beyond the current level for the network component of the federated infrastructure (the cost for a connection to the DFN network with 100 gigabit per second currently amounts to 0.5 M€ per year per site). Should additional technological progress not cover the future demands, additional costs - possibly several M€ per years - need to be secured. Neither considered are the costs for computing installations at the experimental facilities which operate the scientific instruments.

Ensuring sufficient computing support for all further German ErUM communities will also require additional hardware resources beyond the current level of financing. They need services to be delivered that correspond to those of the federated centers of e.g. particle physics. Assuming that the requirements of all ErUM communities but particle physics will be in a similar order of magnitude to those of particle physics of 14 M€ per year, we apply - as a very rough estimate - a factor of about two to the numbers derived for the additional needs of the particle physics community. This results in a steady increase of additional annual investment in hardware and operation reaching about 30 M€ after the year 2025.

5.2 Integration of workflows to exploit infrastructures

We propose an adequate number of personnel to support the many user communities with complex, domain-specific data processing chains and workflows, and to develop community-specific services.

Challenges: Efficient access to future large-scale federated infrastructures, locating research data for processing and analysis as well as methods for data management and data curation require new software

tools and services. Sustainable development work on these services, both by experienced resource providers and on the part of scientists, is vitally important for the operation and exploitation of the infrastructures. For the contributing scientists, both in-depth knowledge of the scientific demands and the computing infrastructures and services are indispensable to find and deploy solutions to optimally use the facilities. Together they need to develop the necessary software tools to make the data accessible, to control and optimally distribute the often complex workflows, and to visualize the results. On the one hand, such solutions have to be developed for standard analysis situations such as strictly predefined procedures of data processing, while, on the other hand, individual, unpredictable workflows driven by new ideas of individual scientists have to be made possible and realized.

Responsibilities: Concrete tasks include the strategic development, deployment and maintenance of software supporting access to and efficient usage of the computing infrastructure. This includes tools for data access and data management. Furthermore, these tools must support processing of complex workflows operating on the data. Account must be taken of domain-specific software libraries. State-of-the-art tools like virtualization and container technologies should be applied where possible to avoid unnecessary dependencies on operating systems or computer architectures. This includes support of modern cloud technologies, which offer the possibility to extend the range of usable infrastructures towards cloud installations at research institutions or commercial cloud providers. Methods for managing scientific workflows and scheduling them for execution on the resulting heterogeneous infrastructures in the best-possible manner must also be developed. Communication between these scientists and their exchange of experience is vital and will be fostered by the activities of the umbrella partnership organization described below.

Scientific headcount: Personnel is required for development work, for the support of community-specific services, and for dedicated community support. The availability of domain experts working in close collaboration with the resource providers is of utmost importance to harness the potential of rapidly developing computer architectures and infrastructures.

To set the scale for additional investments in personnel, the extensive experience from the operation of the German contributions to the Worldwide LHC Computing Grid may be helpful. For computing resources integrated in the central workflow systems (Tier 1 and Tier 2 sites), there are presently 20 scientist positions dedicated to the development and operation of experiment-specific services and user support. Recently, additional funding for five scientist positions has been granted for the development of innovative digital technologies to enable scientists to efficiently access and use heterogeneous computing resources.

As the complexity of the computing landscape, the amount of resources, and the number of services and communities to be supported by large federated centers will increase, dedicated personnel is needed in this area in the future. For the research areas hadrons and nuclei, astroparticle physics and astrophysics, the situation is comparable to particle physics. An even greater number of positions may be required to support the large and diverse condensed matter community. Usage of common infrastructures and tools is expected to lead to synergies, and we therefore recommend the allocation of 100 new scientist positions for all ErUM communities to ensure the operation of community-specific services, the seamless integration of community-specific data processing into the workflow management systems, as well as domain-specific user support for the efficient exploitation of resources at large federated infrastructures. As these are sustained tasks, the new positions should ideally have a long-term perspective.

5.3 Comprehensive management of research data

We propose to support the ErUM communities with the creation of dedicated data scientist and data steward positions to foster the cross-community and cross-disciplinary exploitation of data through adequate data curation and management.

Challenges: Comprehensive management of research data for a multitude of applications in the data life cycle (Fig. 2) constitutes a major challenge. Over the past decades, the awareness for measurement data from the numerous facilities has increased considerably. This is

due to the successful processing of large amounts of data with modern methods of Big Data Analytics, in which new, valuable information could be extracted from the data itself. In addition, by combining data from different detectors, e.g., in astronomy or astroparticle physics or neutron and synchrotron radiation etc., new knowledge emerges.

While the different communities deal with very different issues, they all share common challenges concerning their data management processes. To analyze the common facts and to attempt similar possible developments will be a major challenge within the cross-community commutation. Furthermore, today's experiments are so expensive that repeating them at a later time appears unrealistic. Hence, the creation of data repository platforms will be an essential element.

Sustainable data management procedures for experimental data and simulated data are thus indispensable for long-term success. This applies to the data itself, the data structures and metadata, long-term data storage, data access and iteratively the further data processing as well as data storage and access. All these data will be on distributed, common science data infrastructures which are connected by fast networks.

Responsibilities: The scientists will be responsible for reliably and comprehensibly compiling the various levels of data, from raw data to calibrated data to very elaborate data and reconstructions. The majority of the work will lie in the development and production of qualified data infrastructure software with which data processing can be carried out in a transparent procedure taking into account the FAIR data basic principles.

To be able to use the data for cross-experiment data analyses, these scientists will coordinate their work in committees to design the data of the research field. Here, the partnership umbrella organization described below will play an important role in the exchange of experience. The tasks to be solved by these scientists also include the elaboration and design of suitable and uniform data policies on access rights and embargo periods. In addition, they will advise and support other scientists in all aspects of the use regarding the research data.

Scientific headcount: In view of the many very different experiments in the ErUM programme and the complexity of each of these experiments, a significant increase in personnel will be required for this new level of research data management. Using the examples of the international large-scale particle physics experiments, as well as the large experiments on hadrons and nuclei, we expect a minimum of two scientists per experiment to cover the German participation for these tasks. Similarly, to be able to handle the ten largest experiments for astrophysical all-sky observations with all messengers[4] at all energies, a minimum of 20 scientists is anticipated. The synchrotron and neutron communities perform thousands of individual experiments each year, resulting in a high demand for data management and data curation. Cross-correlations between and the combination of the different experimental data-sets are necessary. Here, a well-balanced measure of support at the individual research centers and universities has to be kept in mind. Moreover, evaluating common challenges of the different communities to benefit from common solutions, and without naming all areas of the ErUM programme in detail here, we estimate that this new area of tasks will require 100 new scientist positions in the years to come. Due to the long-term nature of the responsibilities, these positions should ideally have long-term perspectives.

5.4 Modern Big Data Analytics in fundamental research

We propose the creation of positions for ErUM scientists with in-depth Big Data Analytics skills to establish modern analysis methods in the fundamental research teams.

Challenges: At all universities and research centers in Germany there are many established ErUM research groups with very different scientific questions and participation in numerous, very different experiments. In all these groups, new opportunities are currently arising due to the large amounts of data provided by the current and future generation of large research facilities.

4 Observations include: NIR/optical, sub-mm and radio, X-rays, gamma rays, cosmic rays, neutrinos, and gravitational waves.

With the advent of Big Data Analytics, a new chapter has been opened in which measurement data can be exploited to a new depth, making far-reaching experimental successes more likely. Thus, investments in the construction and operation of experiments will be used even more effectively. Algorithms are the language with which scientific data questions are formulated. The development of a new algorithm is thus a scientific achievement that can lead to groundbreaking discoveries.

The expected improvement potential through Big Data Analytics has many themes, e.g. real-time decisions on the usability of data and thus storage for subsequent evaluations. Further examples are ultra-fast initial data evaluations during data acquisition in order to make decisions about the immediately following measurement program. Furthermore, it concerns in-depth evaluations of the measurement data after data acquisition, simulations, and theoretical calculations, leading to journal publications with a higher scientific value.

Integration of modern Big Data Analytics into research on a very broad scale is essential for keeping Germany's competitive pace of progress. To be able to use Big Data Analytics in the ErUM groups at a level that enables substantial improvements in the outcomes, scientists are needed who have both the domain knowledge about experiments and in-depth knowledge of Big Data Analytics.

Responsibilities: The task of these scientists will be to transfer the expertise in Big Data Analytics into the experiments and to implement the corresponding evaluation programs according to the respective scientific questions. This implementation also includes research on the methods themselves in order to ensure scientific reproducibility, uncover causes for predictions in the use of machine learning techniques, and verify the stability of the methods. Communication between these scientists, mathematicians, and computer scientists and the exchange of experience in the application of the methods will be of key importance and will be strengthened by the measures of a partnership umbrella organization detailed below.

Scientific headcount: Due to the wide range of experiments in the various disciplines and their specific requirements, an expansion of personnel for Big Data Analytics is indispensable. The qualification of these scientists requires a combination of both in-depth domain knowledge and expertise in Big Data Analytics. While some synergies can and should be exploited, the required combination of qualifications to answer a specific scientific question is often only available in very few or even single groups. ErUM research groups are located at the 50 large universities and at the research centers in Germany. For a sustainable and coherent measure we recommend establishing at least 200 new scientific positions, ideally with a long-term perspective. These positions will strengthen all user communities in ErUM disciplines including physics, chemistry, biology, medicine, earth science and possibly more. Their expertise in both their scientific domain discipline and in data science, together with strong ties to the respective scientific fields, provide an ideal background for advancing new data analysis methods. They can also provide a link to scientists from industry who strive to apply Big Data Analytics at large research facilities to solve specific problems.

5.5 Scientists' integrated web working environment

We propose the development of a scientist-oriented web system that offers the full functionality of modern scientific working with globally distributed collaborators, resources and research data. The system should provide the comfortable look and feel of a local working environment.

Challenges: Scientists should be able to concentrate fully on answering their scientific questions and therefore need comprehensive technical support. Technology must simplify the scientific process without restricting it, i.e. the creativity and feasibility of new ideas should be supported to the maximum.

Scientists in ErUM need a working environment in which they can access research data and proven algorithms, develop and eventually include their own new algorithms, and apply these tools to perform data analyses on scientific data. A similarity to search engines such as Google is noticeable, but the decisive differences are that scientists need to extract novel

information from their own or archival data. In order to do so, they need to develop new algorithms for data processing or modify and adapt existing algorithms such that they comply with their research questions.

Modern scientific work requires many more advanced functionalities in such environments. Examples are software portals, code management systems, team instruments for the continuous development of software projects, design systems for the creation and execution of highly complex data analyses, methods of analysis preservation, as well as integrated access to metadata systems and data portals. Also, a comprehensive communication system for collaborative work is needed which enables all levels of information exchange, reaching from a meeting of a handful of scientists to mastering conferences with the exchange between several hundred scientists.

For several years now, prototype infrastructures have been built for scientists' web access to research data, computing resources for processing them, and development options for individual algorithms that are suitable for answering the respective scientific questions[5].

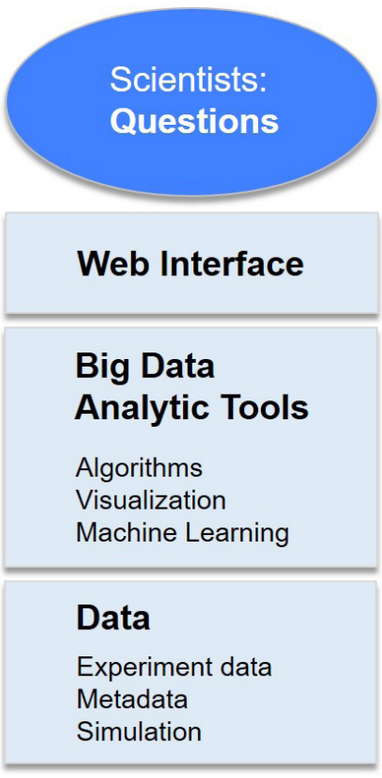

Figure 3: Web working environment with tentative labeling *Scioogle* enabling scientists to answer their scientific questions by accessing research data, developing and applying Big Data Analytic Tools, having the computing power available for data processing, and visualizing the results.

The development of a large-scale service with tentative labeling *Scioogle* (Fig. 3) serving scientists of various disciplines in Germany is vital and should be realized through the cooperation of various disciplines (ErUM scientists, computer scientists, infrastructure-providers). In the portfolio of recommended measures presented here, *Scioogle* is tightly connected to successful developments in all measures detailed in the previous sections, namely data management, utilization of the federated infrastructures for mastering highly complex workflows, etc. It will be hosted in the federated infrastructure described above.

Scioogle is in accordance with the ideas of EOSC and is expected to expand in the scientific field in the international context where Germany should take a leading role. The networked access will enable numerous new analyses and the formation of new ideas in ErUM. Furthermore, it can be used by numerous other disciplines to realize their data analyses and also to enable the interested public to participate in Big Data Analytics. Its openness can also be valued as a contribution to transparent sciences which resonate across society.

Responsibilities: In order to build and maintain a web working environment for scientists, experienced scientists with substantial knowledge of infrastructures and in the requirements of scientific work with research data are needed. The cooperative task of experts from the various disciplines (ErUM scientists, computer scientists, infrastructure providers) is to build the user functionalities of the web working environment with which, firstly, users can find the research data they are interested in. Secondly, a marketplace for algorithms is to be developed so that users can run proven algorithms on data for their questions, continue to develop and enter their own, better algorithms, and make proven algorithms available on the marketplace. Thirdly, comprehensive integration of all functionalities

5 The VISPA project, https://vispa.physik.rwth-aachen.de; The SWAN project, https://swan.cern.ch; Belle II project, https://stash.desy.de/projects/B2T/repos/b2-starterkit

regarding modern scientific work from the development of software projects, code management systems, design and creation systems of data analyses with integrated workflow management, analysis preservation, software portals, and communication systems is needed for collaborative work. Finally, a strong user support team needs to be established.

Scientific headcount: The number of scientists required to transcend prototype facilities and reach a production environment that is accepted and used by 8,400 scientists and their students is not easy to estimate. In view of the multiple areas of responsibilities from integrating research data, a marketplace for algorithms and many further work management systems all the way to a broad communication system including user support, we recommend that 50-100 scientists from different disciplines be assigned to set up and operate a comprehensive web environment that will accelerate ErUM science in Germany. As these are sustained responsibilities, these new positions should ideally provide a long-term perspective.

5.6 Tenure-Track programme: knowledge in digitization

Education and training in methods of modern digitization play a decisive role in implementing the necessary far-reaching advancements in the German research landscape. We propose a Tenure-Track programme for scientific leaders with special expertise in the areas of computing models, data models and algorithm development.

Challenges: In communicating the basics of science education, universities are the portal for spreading knowledge about modern digitization. Together, universities and research centers establish training of scientific knowledge, and constitute the source from which the vast majority of young scientists, company founders and highly qualified company employees emerge. The measures proposed here will thus have a broad and large-scale impact.

The success of modern experimental research relies on both excellent measurement technology and cleverly designed advanced computing and data models. Traditionally, leading scientist positions in experimental ErUM science have been awarded to scientists with knowledge in the development of detectors and the analysis of experimental data. Both universities and research centers lack a correspondingly strong group of scientific leaders who perform frontier research in the areas of algorithm development, computing and data models for their scientific data analyses in ErUM (Fig. 4). Their goal will be to accelerate scientific progress through improved algorithms and better exploitation of data.

In this context, it should be noted that education and training in modern digitization methods in ErUM differ from currently upcoming programmes in data sciences. In ErUM, teaching in modern digitization is directly related to research applications and is regarded as a necessary expertise in addition to ErUM domain knowledge to successfully advance research in our field.

The initiation of such scientific leaders through the usual path of positions arising from retirements would take decades and would severely risk Germany falling behind. Instead, establishing new leadership through a large-scale Tenure-Track programme would flank the experimental programs with the urgently needed data-oriented research on a much shorter time scale.

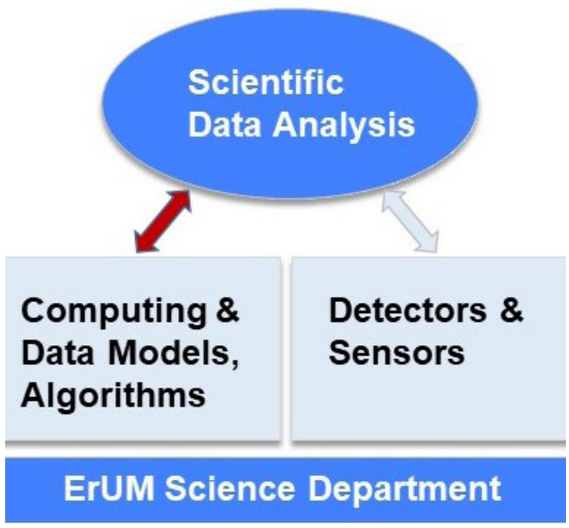

Figure 4: Tenure-Track programme for new scientific leaders focusing on frontier ErUM research related to algorithm development, computing and data models, thereby complementing the experimental and theoretical research in ErUM. Their responsibilities include modernizing curricula in science education and spreading knowledge about current digitization methods.

Universities and research centers can receive funding for a scientific leadership position with the endowment of a scientist position for a period of six years if they ensure the structural prerequisites for continuing as a permanent leader. The programme should be open to all ErUM disciplines including physics, chemistry, biology, medicine, earth science and possibly more.

Outstandingly qualified young scientists are already available for such a tenure-track programme today. In the past, these individuals often left the university/research sector because of a lack of opportunities within the more traditional science programs at universities and better offers in industry. Sustainability of the digital research sector shall be ensured by introducing a digital pillar with new career paths in fundamental sciences. With this, research in the field of ErUM will be accelerated and subject-specific teaching about the new possibilities of digitization will be secured.

Responsibilities: The new scientific leaders will take responsibility in conducting frontier research in the areas of computing and data models and algorithm development, complementing the experimental and theoretical research in ErUM. Beyond their research they will modernize the curricula of natural sciences education in data-driven methods and teach correspondingly. On the one hand, this would support young scientists in improving their training in the methods of Data Analytics, computing and data models; on the other hand, the knowledge and skills of these students would diffuse into practically all groups of institutes as part of their final theses. Thus, the Tenure-Track programme will advance data-driven research and at the same time be a catalyst to distribute and deepen knowledge in digitization through ErUM.

Scientific headcount: In order to achieve such a structural change through the approximately 50 large universities and research centers nationwide and the various departments of basic research on the universe and matter, we propose a correspondingly extensive and broadly diversified Tenure-Track programme with 100 scientific leadership positions, each supported by one research assistant.

5.7 Partnership for Innovative Digitization

We propose the establishment of a self-organized structure that will integrate scientists from the various disciplines associated with ErUM. This will accelerate scientific progress through the transfer of knowledge on modern digital methods and their applications to very different research questions.

Opportunities and challenges: The communities in ErUM work on very different scientific questions, but are facing similar problems regarding modern digitization. Challenges and opportunities lie close together. Only the ErUM scientists themselves can define and develop ideas on how they can all be brought together into a fruitful and efficient exchange. A minimum

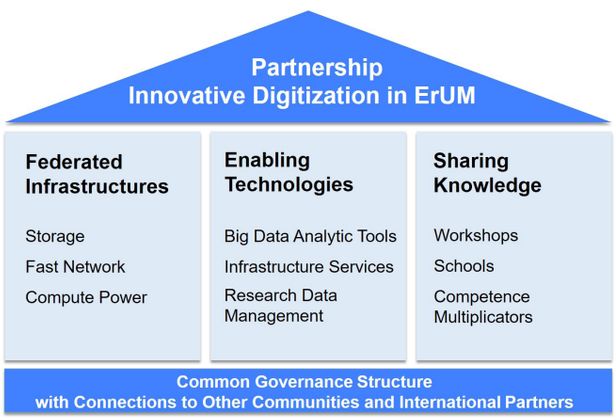

Figure 5: The Partnership for Innovative Digitization in ErUM constitutes the strategic measure across all common activities and interests of the participating scientists. It is based on self-organization and comprises the areas of federal infrastructures, the various activities for enabling modern digital methods, and the joint actions for preserving and transferring acquired knowledge.

measure is to offer a basic structure through which scientists can learn from each other, exchange their problems and discuss the transferability of solutions. This basic structure will generate a new and strong network for responding to the challenges of modern digitization.

Many scientists have already had very good experiences with the so-called 'Forschungsschwerpunkte' initiated by the BMBF and the 'Alliances' initiated by the Helmholtz Association. This BMBF workshop also brought together scientists who would otherwise probably not have communicated with one another. In the past, the self-structured development of communities with similar research interests has greatly favored scientific progress. In simple terms, in many cases competitors became fellow scientists and the overlap between the research interests of individual scientists who had not met yet led to exchanges and joint initiatives.

Community-building measure: In Fig. 5 we show the envisaged structure of the Partnership for Innovative Digitization. The different pillars of the partnership are closely linked. Data storage hardware is closely related to infrastructure services, i.e. programs that provide optimal access to storage capacity for research data. This is also linked to the initiatives in the area of improved research data management. Furthermore, computer architectures are to be purchased in close coordination with the requirements of Big Data Analytics and made efficiently usable by suitable programs. The training of young scientists is to be carried out by the experts.

Strategic measures: The partnership opens up a platform to kick start the communication and cooperation within and between the communities. We envision an approach which identifies the needs, proposes concepts and overcomes challenges in the field of Big Data. Performing series of educational workshops by specialists bridges gaps of knowledge between experienced and less skilled communities on the one hand and optimally opens a new perspective on the other hand. Bringing together different communities with similar needs within the partnership can be a nucleus to carve out possible common ground and develop effectively cooperating working groups. The sharing of knowledge from existing initiatives (e.g. projects within the EU data research projects - Horizon2020) can be a common nucleus to define the working groups for future data research and synergetic participation in broader initiatives like EOSC and NFDI. The educational aspect can lead to a multiplication of competences and, at best, to the qualification of data specialists.

Relation to mathematics and computer science: Cooperations to understand research data through mathematical modeling exist in several areas of ErUM research. Beyond this, current developments in computer science and mathematics are of key importance for the further development of Big Data Analytics in ErUM research. Many concepts developed in computer science were transferred to ErUM research questions in a timely manner. Developments in the field of deep neural networks are particularly recent examples and are highly variable due to the millions of adjustable parameters for the description of very complex systems and processes. Newly developed concepts of hybrid architectures of interacting networks offer as yet unfeasible new possibilities, e.g. for simulations of experiments or for refining simulations based on research data.

All of these new research directions need mathematical foundations to prove the functionality and stability of the new methods from fundamental principles. This concerns questions of representation, learning algorithms, generalization, and explainability, which concern a wide range of disciplines from applied harmonic analysis to differential geometry all the way to learning theory and optimization. Thus, to ensure a successful implementation of modern methods of Big Data Analytics, cooperation between ErUM scientists and scientists from computer science and mathematics is of utmost importance. The umbrella organization will be an ideal breeding ground for the collaboration and continuous exchange between scientists from the different fields.

Relation to industry: The partnership also provides an ideal hub for connections to industry. In particular, it can bring together the right partners from the private sector and the ErUM communities for joint research projects in the area of Big and Smart Data. There are a few examples of such joint research projects. For example, via the CERN Openlab, scientists are given access to new developments in industry, and the

companies benefit from the close collaboration with early adopters. Thanks to the contributions of students participating in the Google Summer of Code, several innovations and improvements are available to researchers and companies as open source. While these examples show the potential that such collaborations offer, they are often limited to one community.

Furthermore, the partnership could enhance the transfer of technology via the support of spin-offs. A good example of a successful spin-off is Blue Yonder. Founded by a physicist, it applied Big Data Analytics methods developed for science to problems in industry and became the market leader for supply chain management. Industry is also an important user of synchrotron large-scale facilities in Germany. Bringing together industrial scientists and scientists with Big Data knowledge and expertise will enhance the optimum usage of facilities for industry and facilitate the exchange at this important interface between industry and science.

Dynamic funding: Especially in the field of digitization there are a multitude of ideas for the realization and implementation of the new possibilities. Here, we recommend equipping the self-organized partnership for the optional allocation of dynamic funding, which fosters promising approaches for a given development time in which prototypes can be developed and evaluated. Only then should a joint solution strategy be developed, funded and implemented by the participating scientists.

Cost model: We estimate the partnership's annual budget for innovative digitization as 3 M€. This includes the salaries of postdoctoral researchers in the various ErUM research fields to initiate and organize all the common activities, and of secretaries supporting them. It also includes financial support to enable workshops, schools and the dynamic funding of pioneering work. One particular example of such a workshop planned in this framework could be *Rapid Reaction* meetings, organized to bring together a group of about 15 to 20 world-leading experts in order to address and clarify a specific problem. Such a meeting could result in a recommendation of a rapid start-up funding of an urgent measure.

5.8 Cost estimates

In this section, we summarize the expected costs for the portfolio of actions in the field of ErUM research. We would like to emphasize once again that all measures build on an already functioning scientific environment provided by the many stakeholders in ErUM. The portfolio of measures to address the new challenges of the digital era will have to be implemented beyond all existing resources.

Overall, the majority of the recommended funding measures are needed to employ ErUM scientists who will drive progress in data models, compute models and Big Data Analytics, and who will broadly distribute the

Full Time Equivalents		MEuro/y /position in 2020
1. Workflows to exploit infrastructures	**100**	0.072
2. Management of research data	**100**	0.072
3. Big Data Analytics in physics research	**200**	0.072
4. Scientist's web working environment	**100**	0.072
5. Tenure track ErUM programme + 1 RA*	**100**	0.158
Total FTE	**600**	

*RA=Research Associate

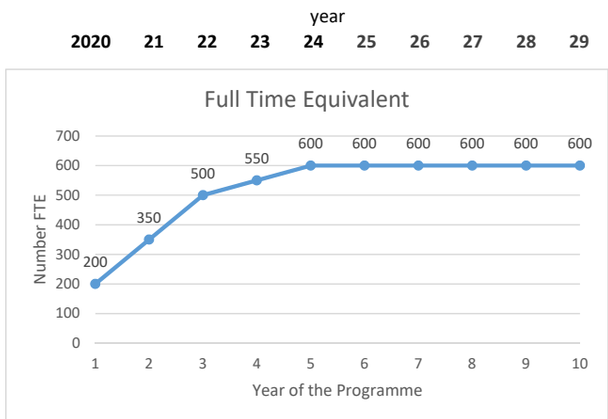

Figure 6: Recommended development of new scientist positions over 10 years as full time equivalents (FTE) for 1) Integration of workflows to exploit infrastructures (section 5.2), 2) Comprehensive management of research data (section 5.3), 3) Modern Big Data Analytics in physics research (section 5.4), 4) Scientists' integrated web working environment (section 5.5), 5) Tenure-Track programme: knowledge in digitization (section 5.6). The right column shows the conversion key from FTE to € used below.

knowledge needed to cope with the digital evolution within the German society. The table in Fig. 6 shows the recommended growth for the five task fields in which the number of scientific positions is to gradually increase over several years to the level of the estimated personnel requirements as outlined in the previous sections 5.2-5.6. The number of positions required is expressed in units of full-time equivalents.

Depending on the task, we foresee medium term (5-10 years) or permanent positions in order to achieve sustainability in the measures taken. Measured against the 8,400 scientists in ErUM's field of research, the growth presented in the table corresponds to an average annual personnel increase of 0.7% in relation to the ErUM scientists.

In the cost estimates presented in Fig. 7 we convert the FTE figures into units of million Euro (M€) using the conversion factors shown in the right column of the table in Fig. 6. For the Tenure-Track programme, for the year 2020 for the Tenure-Track position we estimate 0.0858 M€ per year plus 0.072 M€ for a research assistant assigned to the leading scientist. We also take into account an inflation compensation of 3% for all positions. For the hardware of the federated infrastructures we envisage a growth over 5 years, so that in the mid-2020s the projected additional annual needs of 30 M€ will be reached on the basis of the cost model described in section 5.1. The measures mentioned here exclusively comprise the necessary additional computing power and storage for handling research data and the operating costs to run this additional hardware.

The table in Figure 7 outlines the recommended course of funding over time. Thus, these infrastructures will be available on time for the start of the operation of experimental facilities with very high data rates and will at the same time supply the broad spectrum of the experimental programme in ErUM.

The table in Figure 7 also shows the costs of the community-building measures for the Partnership for Innovative Digitization for which we recommend 3 M€ per year, which includes personnel for the organization, funding for workshops and schools, as well as dynamic funding for pioneering research (section 5.7).

In total, the programme will grow from 26 M€ to 86 M€ in the first five years. We would like to again emphasize that the broad positioning of our recommended measures lies in the successful tradition of Germany:

Excellent educational opportunities for our society will ensure that both science and the economy will continue to develop in a healthy and successful way.

Cost estimate of recommended measures /MEuro	year 2020	21	22	23	24	25	26	27	28	29	Meuro /topic over 10y
Full Time Equivalents	17.8	32.1	47.3	52.6	58.3	60.0	61.8	63.7	65.6	67.6	526.9
Large-scale federated infrastructures	5.0	10.0	15.0	20.0	25.0	30.0	30.0	30.0	30.0	30.0	225.0
Partnership for innovative digitization	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	30.0
Total Cost Estimate	**25.8**	**45.1**	**65.3**	**75.6**	**86.3**	**93.0**	**94.8**	**96.7**	**98.6**	**100.6**	**781.9**

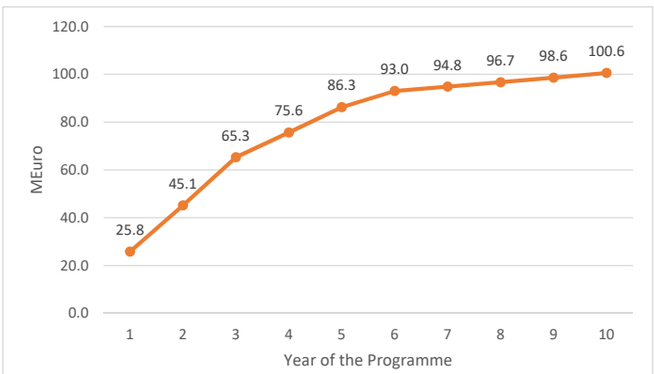

Figure 7: Cost estimate of the recommended measures for the ErUM Data Programme over 10 years. The table shows separately the costs for scientific positions according to Fig. 6, for hardware investments for the federated infrastructures (section 5.1) and for the Partnership for Innovative Digitization (section 5.7).

Appendix A Committees of research within the ErUM Programme

Committee for		Scientists with doctoral degree
Forschung mit Synchrotronstrahlung	KFS	2,300
Rat Deutscher Sternwarten	RDS	1,500
Hadronen- und Kernphysik	KHuK	1,500
Elementarteilchenphysik	KET	1,300
Forschung mit Neutronen	KFN	1,000
Astroteilchenphysik	KAT	500
Beschleunigerphysik	KfB	200
Forschung mit nuklearen Sonden und Ionenstrahlen	KFSI	100
Total number of scientists with doctoral degree		**8,400**

Figure 8: The total number of scientists with a doctoral degree working on research within the Universe and Matter programme ErUM is approx. 8,400.

Description of the Committees

Komitee für Astroteilchenphysik (KAT)

Astroparticle physics combines our knowledge of the largest structures in the universe with that of the smallest building blocks of matter and the forces between them. It is a fascinating field of science that lives at the interfaces of astronomy, astrophysics, cosmology, elementary particle physics and nuclear physics. Activities in this young research field have increased dramatically in the last two decades. Astroparticle physics has become independent and has developed its own profile in many countries worldwide. The Committee for Astroparticle Physics (KAT) represents German physicists working in the field of astroparticle physics at German universities, Helmholtz Centers and Max-Planck Institutes. The aim of KAT is to bring together the different research directions, discuss current developments and represent the interests of the scientific community. KAT seeks close contact with the community of German astroparticle physicists with the aim of achieving the greatest possible consensus while at the same time including strategic aspects. KAT represents the common goals and interests of the global community.

Astroparticle physics is characterized by its globally distributed and diverse infrastructures. Information is obtained in particular by linking the data of different experiments (Multi-Messenger Astroparticle Physics), in which the digitalization of the research field plays an important role. The infrastructures currently funded by ErUM-pro are the Gamma-ray Astronomy Observatory *CTA*, the Cosmic Ray experiment *Pierre Auger Observatory*, the Neutrino Observatory *IceCube*, the Dark Matter Experiments *CRESST*, *XENON*, the neutrino mass experiment *KATRIN*, and the Double Beta Decay Experiment *GERDA*.

Komitee für Elementarteilchenphysik (KET)

The Committee for Elementary Particle Physics (KET) is the elected body representing 1300 particle physicists with a PhD degree from more than 20 universities, at the European research center CERN, at the Helmholtz Centre DESY, at two Max Planck institutes and at non-European particle research centers. In close collaboration with the particle physics community, KET formulates common goals and interests. Topics covered include support for young scientists and outreach to the public. KET moderates discussions about the future strategy of particle physics with German funding agencies and represents the German community on the international level. Particle physics research is characterized by its large international research facilities like CERN in Geneva, Switzerland, or KEK in Japan. The actual flagship projects are the experiments ATLAS, CMS and LHCb at the Large Hadron Collider (LHHC) at CERN and Belle II at KEKB accelerator in Japan. The experiments address fundamental questions

concerning the smallest constituents of matter and their interactions under conditions similar to those at the beginning of the Universe. Answers to such questions provide important input to cosmology in its search for a consistent description of the origin and development of the Universe.

Experiments in particle physics are among the most data-intense endeavors of mankind. The particle physics community operates the largest distributed science grid (the Worldwide LHC Computing Grid, WLCG) with contributions from 170 computer centers in 42 countries, providing in total one million computer cores and one thousand petabytes of storage space. The well-defined roadmap for future developments of accelerator performance and novel, more precise detector components will lead to an annual increase of data rates by one order of magnitude in the next decade.

Komitee für Beschleunigerphysik (KfB)

Particle accelerators play a prominent role in the investigation of elementary particles, hadrons and nuclei, in the investigation of condensed matter with photons, neutrons and charged particles as well as in other areas of science, technology and medicine. In addition to the construction and operation of particle accelerators, accelerator physicists are dedicated to the further development of accelerator systems and the development of new concepts and basic technologies.

The KfB represents accelerator physicists at German universities, at Helmholtz centers, at other German research institutes or at foreign institutes with German participation. It fosters contact within the community through the organization of regular meetings or workshops and promotes the training and education of young scientists. The interconnected R&D activities of the involved institutes and accelerator laboratories require a common strategy process which the KfB drives forward in coordination with the user communities, e.g. elementary particle physics or photon science.

The design, operation and optimization of modern accelerators, based on either RF or novel technologies, relies increasingly on the availability of high-throughput and high-performance computing. Examples are feedback loops or full start-to-end simulations to increase machine performance. High-end beam diagnostic devices with high data-rate data acquisition or large-scale plasma simulations trigger the need for advanced analysis tools and common standards across facilities.

Komitee für Forschung mit Neutronen (KFN)

The Committee Research with Neutrons (KFN) represents the interests of the German neutron user community to politics and to neutron facilities. The members of the KFN are elected by around 1700 registered regular neutron users (1000 PhD) that are active at Universities (about two thirds), at public research institutes (Helmholtz, Leibniz, Max Planck, Fraunhofer and other societies) or at industrial societies. Representatives of the research facilities and the project management are guests in the KFN to facilitate the direct dialogue. The scientific interests of the neutron community are wide-spread covering amongst others physics, chemistry, biology, geoscience, medicine, electro-mechanical engineering and cultural heritage science. Consequently, the instrumentation required to meet this variety of scientific problems is also widespread. Neutron experiments are irreplaceable in all these scientific fields but construction and operation of neutron sources require large societal efforts that scientists must compensate with substantial results. It is the central role of the KFN to shape this necessary dialogue between the neutron user community and the society.

The KFN is active in the European User Association ENSA and in the KEKM (Commission Condensed Matter Re-search at Large Scale Facilities) and it is regularly consulted in any questions concerning research with neutrons. The KFN publishes strategic recommendations as brochures, see documents. Contact persons are in charge of different aspects, and users find information on the activities of the committee on the website and by email newsletters.

Neutron sources (national, European): FRMII@Heinz Maier-Leibnitz Zentrum (Garching, Germany), BerII@ HZB until 2020 (Berlin, Germany), Institute Laue-Langevin (Grenoble, France); European Spallation Source (ESS) from 2023, Lund (Sweden).

Komitee für Forschung mit Synchrotronstrahlung (KFS)

The Research with Synchrotron Radiation committee (KFS) is an elected body representing the interests of synchrotron radiation users (including X-FELs) in Germany in politics and research centers. It represents approximately 4000 users of synchrotron radiation from universities, non-university organization and research centers. Representatives of the research centers and the project sponsor (DESY PT) are represented as guests in the KFS. Information is gathered at periodic meetings in which strategic discussions are held and the activities of the KFS are planned. The KFS is represented in the European User Organization (ESUO) and is also regularly consulted in various contexts of research with synchrotron radiation. The KFS publishes strategic recommendations in brochures and documents. Users can direct their questions at contact persons in charge of various ministries, and they will find up-to-date information on the activities of the KFS on the website, e.g. in circulars and protocols.

Synchrotron and X-ray free electron laser sources: BESSYII - HZB (Berlin), PETRA III - DESY (Hamburg), European Synchrotron Radiation Facility (Grenoble), FLASH - DESY (Hamburg), European XFEL (Schenefeld)

Komitee Forschung mit nuklearen Sonden und Ionenstrahlen (KFSI)

The Research Committee on Nuclear Probes and Ion Beams (KFSI) is an elected body representing the interests of charged particle, i.e. high energy ions and positron beam users in the field of solid state physics, material sciences and interdisciplinary research in Germany vis-a-vis politics and the centers. The direct dialogue is facilitated by the presence of representatives of the research centers and the promoter as guests in the KFSI. The aim of the meetings is to gather information, engage in strategic discussions and plan the work of the KFSI. It supports research in particular at the high-energy ion beam research centers of Helmholtz organization GSI (Darmstadt) and Dresden Rossendorf (HZDR), the radioactive beam facility ISOLDE and the slow positron beams at the Munich neutron induced positron source NEPOMUC at FRM II.

Komitee für Hadronen- und Kernphysik (KHuK)

The national committee on the physics of hadrons and nuclei (‘Komitee Hadronen und Kerne’, KHuK) was founded by the Ministry for Education and Research (BMBF) as an advisory board for the ministry and a lobby for the German hadron and nuclear physics community. Nine elected KHuK-members represent different research directions including two members for theory. They are supported by four ex-officio members who represent other important national and international program committees, e.g. for BMBF funding, DFG funding, DPG, NuPECC and ESF. KHuK discusses and reports the status quo and future perspectives of hadron and nuclear physics on the national and international level. Based on this, the committee formulates recommendations for the development and strengthening of the field and the community. These recommendations serve as input to European strategy processes, for example, the last NuPECC Long Range Plan (LRP) from 2017 and the European Strategy Process in Particle Physics (EPPS) in 2018. The committee represents more than 1500 scientists with a PhD degree from German universities and research institutions.

Rat Deutscher Sternwarten (RDS)

The Council of German Observatories (Rat deutscher Sternwarten, or RDS for short) represents the common interests of all German astronomical research institutions and their scientists vis-a-vis funding bodies, governments, international organizations and other relevant boards and committees. The RDS provides professional consultation to state institutions and other decision makers. The RDS represents the German astronomical community within the International Astronomical Union (IAU). It also reviews applications from German astronomers and astronomy enthusiasts applying for individual IAU membership. The RDS nominates representatives for national and international committees in the area of astronomy and astrophysics. The RDS issues research strategy papers and participates in the development of European astronomy within the ASTRONET initiative. The RDS elects an Executive Committee from the representatives of its member institutions. The Executive Committee advises and supports the chair on questions relating to research policy and research strategy, and can endorse formal statements at short notice. The RDS currently has 40 institutional members.

Appendix B Participants of the Workshop

Last name	First Name	Institution
Appel	Sabrina	GSI Helmholtzzentrum für Schwerionenforschung GmbH
Bayan	Oya	Rheinisch-Westfälische Technische Hochschule Aachen (RWTH)
Blank	Kuno	Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.
Bock	Hans Georg	Ruprecht-Karls-Universität Heidelberg
Boine-Frankenheim	Oliver	Technische Universität Darmstadt
Brockhauser	Sandor	European X-Ray Free-Electron Laser Facility GmbH (XFEL)
Bründermann	Erik	Karlsruher Institut für Technologie (KIT)
Brünger-Weilandt	Sabine	FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur
Büscher	Volker	Johannes Gutenberg-Universität Mainz
Bussmann	Michael	Helmholtz-Zentrum Dresden-Rossendorf e. V.
Conrad	Tim	Freie Universität Berlin
de la Mar	Jurry	T-Systems International GmbH
Desch	Klaus	Rheinische Friedrich-Wilhelms-Universität Bonn
Dettmar	Ralf-Jürgen	Ruhr-Universität Bochum
Dietz	Volkmar	Bundesministerium für Bildung und Forschung (BMBF)
Dimper	Rudolf	European Synchrotron Radiation Facility (ESRF)
Ebert	Barbara	Technische Universität Dresden
Ehrenfeld	Wolfgang	Projektträger DESY
Elsen	Eckhard	European Organization for Nuclear Research CERN
Elsing	Markus	European Organization for Nuclear Research CERN
Engel	Andreas	SAP SE
Erdmann	Martin	Rheinisch-Westfälische Technische Hochschule Aachen
Fischer	Andrea	Bundesministerium für Bildung und Forschung (BMBF)
Föhlisch	Alexander	Helmholtz-Zentrum Berlin für Materialien und Energie GmbH
Frahm	Ronald	Bergische Universität Wuppertal
Gast	Mikael	Bundesministerium für Bildung und Forschung (BMBF)
Geddes	Neil	STFC Rutherford Appleton Laboratory
Giubellino	Paolo	GSI Helmholtzzentrum für Schwerionenforschung GmbH
Gülzow	Volker	Deutsches Elektronen-Synchrotron DESY
Gutt	Christian	Universität Siegen
Halm	Christine	Projektträger DESY
Hamaekers	Jan	Fraunhofer-Institut für Algorithmen und Wissenschaftliches Rechnen (SCAI)
Hammer	Barbara	Universität Bielefeld
Hasler	Tim	Zuse Institute Berlin (ZIB)
Haungs	Andreas	Karlsruher Institut für Technologie (KIT)
Helbing	Klaus	Bergische Universität Wuppertal
Hinton	James Anthony	Max-Planck-Institut für Kernphysik
Hoffmann	Jens-Uwe	Helmholtz-Zentrum Berlin für Materialien und Energie GmbH
Hoppe	Diana	Projektträger DESY
Hradil	Klaudia	Technische Universität Wien
Karalopoulos	Athanasios	European Commission
Karsch	Frithjof	Universität Bielefeld
Kisel	Ivan	Frankfurt Institute for Advanced Studies (FIAS)
Knüpfer	Andreas	Technische Universität Dresden
Kramer	Michael	Max-Planck-Institut für Radioastronomie Bonn
Kramer	Tobias	Zuse Institute Berlin (ZIB)
Kranzlmüller	Dieter	Leibniz-Rechenzentrum der Bayrischen Akademie der Wissenschaften (LRZ)
Kroseberg	Jürgen	Bundesministerium für Bildung und Forschung (BMBF)
Kuhr	Thomas	Ludwig-Maximilians-Universität München

Kunze	Marcel	Ruprecht-Karls-Universität Heidelberg
Lindenstruth	Volker	Johann Wolfgang Goethe-Universität Frankfurt am Main
Maas	Frank	Johannes Gutenberg-Universität Mainz
Mallmann	Daniel	Forschungszentrum Jülich GmbH
Mannheim	Karl	Julius-Maximilians-Universität Würzburg
Markl	Volker	Technische Universität Berlin
Maerten	Lena	Bundesministerium für Bildung und Forschung (BMBF)
Masciocchi	Silvia	Ruprecht-Karls-Universität Heidelberg
Müller	Anke-Susanne	Karlsruher Institut für Technologie (KIT)
Müller	Matthias	Rheinisch-Westfälische Technische Hochschule Aachen
Murphy	Bridget	Christian-Albrechts-Universität zu Kiel
Müssner	Rainer	Bundesministerium für Bildung und Forschung (BMBF)
Mutti	Paolo	Institut Max von Laue - Paul Langevin (ILL)
Neuroth	Heike	Fachhochschule Potsdam
Osterhoff	Jens	Deutsches Elektronen-Synchrotron DESY
Pippow	Andreas	Fraunhofer-Institut für Angewandte Informationstechnik FIT
Polsterer	Kai	HITS gGmbH - Heidelberger Institut für Theoretische Studien
Quast	Günter	Karlsruher Institut für Technologie (KIT)
Reichart	Patrick	Universität der Bundeswehr München
Richter	Tobias	European Spallation Source ESS ERIC
Ritz	Raphael	Max Planck Computing and Data Facility
Schaeper	Jannis	Georg-August-Universität Göttingen
Schinnerer	Eva	Max-Planck-Institut für Astronomie
Schmidt	Alexander	Rheinisch-Westfälische Technische Hochschule Aachen
Schneidewind	Astrid	Forschungszentrum Jülich GmbH
Schroer	Christian	Deutsches Elektronen-Synchrotron DESY
Schumacher	Markus	Albert-Ludwigs-Universität Freiburg
Schwarz	Kilian	GSI Helmholtzzentrum für Schwerionenforschung GmbH
Seeger	Bernhard	Philipps-Universität Marburg
Steinmetz	Matthias	Leibniz-Institut für Astrophysik Potsdam (AIP)
Sträter	Rainer	1&1 Internet
Streit	Achim	Karlsruhe Institute of Technology (KIT)
Surzhykov	Andrey	Technische Universität Carolo-Wilhelmina zu Braunschweig
Swiebodzinski	Jacek	Projektträger DESY
Taylor	Jonathan	European Spallation Source ESS ERIC
Torge	Sunna	Technische Universität Dresden
Tschentscher	Thomas	European X-Ray Free-Electron Laser Facility GmbH (XFEL)
Wambsganß	Joachim	Ruprecht-Karls-Universität Heidelberg
Weber	Marc	Karlsruher Institut für Technologie (KIT)
Wender	Jan	Atos Information Technology GmbH
Weinheimer	Christian	Westfälische Wilhelms-Universität Münster
Wenzel-Constabel	Peter	Bundesministerium für Bildung und Forschung (BMBF)
Wiesenfeldt	Sören	Helmholtz-Gemeinschaft Deutscher Forschungszentren
Wilms	Jörn	Friedrich-Alexander-Universität Erlangen-Nürnberg
Winkler-Nees	Stefan	Deutsche Forschungsgemeinschaft DFG
Wolfangel	Eva	Moderation
Würges	Jochen	Projektträger DESY
Wuttke	Joachim	Forschungszentrum Jülich GmbH
Zach	Karin	Deutsche Forschungsgemeinschaft DFG