HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

HELMHOLTZ INCUBATOR
INFORMATION & DATA SCIENCE

Further developing an innovative, association-wide approach

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

CONTENTS

1. INTRODUCTION

The digital transformation is the biggest challenge facing science, business, and society in the early 21st century. It offers **enormous opportunities** in nearly all aspects of life: innovative modes of working and community living, completely new platforms for trade and science, unimaginable possibilities in the field of medicine, revolutionary approaches to networks supplying water and energy, forward-thinking practices for environmental protection and migration, and much more.

At the heart of the digital transformation is the knowledge value chain of „**data – knowledge – innovation.**" This chain also represents a crucial challenge faced by every scientific discipline aiming to shape society's future progress.

Ambitious, courageous, and innovative steps are needed to ensure Germany will participate in and shape this future progress. Germany is known as a place of innovation and research. Combining computer science, mathematics, statistics, sensor technology, simulation, and data-intensive computing with challenging applications in natural sciences, engineering, medicine, the humanities, and social sciences will be a dynamic driver in research going forward. These innovative links will generate completely new scientific insights and offer significant benefits to business and society at large.

Germany boasts significant transformative high-tech potential and is currently in a solid starting position globally. In order to maintain and further develop this position, **entirely new approaches and larger-scale efforts are needed within the scientific system**.

2. INFORMATION & DATA SCIENCE AT THE HELMHOLTZ ASSOCIATION

The **Helmholtz Association of German Research Centers** contributes to solving large, pressing problems facing society, science, and business by excelling in six fields of scientific research: Energy; Earth and Environment; Health; Key Technologies/Information; Matter; and Aeronautics, Space, and Transport. With over 38,000 employees at 18 research centers, it is the largest research organization in Germany.

As a research organization aiming to address major current societal issues, the Helmholtz Association has accumulated an impressive portfolio of expertise in the field of **Information & Data Science** over the past decades. This includes areas such as information processing, Big Data, data analytics, simulation, modeling, bioinformatics, imaging techniques, research data management, high performance computing, robotics, technical and biological information systems, and more. The Helmholtz Association has access to a trove of Big Data that is growing exponentially and incomparable to that of almost any other research organization.

The Helmholtz Centers and research fields are also in an **outstanding position** to facilitate the successful pooling of expertise in order to create new synergies. They operate large research infrastructures (including satellite missions and major facilities for nuclear and particle physics), use the latest generation of supercomputing technologies, and are experts in complex simulation (e.g., for extensive Earth and climate models, virtual material design, and systems biology). In this regard, the Association has long been conversant with every aspect of the entire data life cycle, from research planning and data collection, handling and maintenance, to the analysis, evaluation, and practical application of very large and complex data sets. All of the Helmholtz sites utilize leading approaches – some of which are global benchmarks – and demonstrate outstanding methodological competence in their relevant fields.

Building on this, the Helmholtz Association is and will be working to develop its own competency in the highly topical field of Information & Data Science, and to create synergies in the research landscape, adopt new findings from national, European, and international contexts, and implement groundbreaking approaches. In addition to strengthening individual research areas with state-of-the-art methods in Information & Data Science, the stated goal is to develop the field across the entire organization at a multidisciplinary level.

The Helmholtz Association is **addressing the complex area of Information & Data Science at all levels:**

- It is a top priority on the President's agenda.

- Centers and research fields address the immense potential of this topic in their respective strategies.

- In the context of restructuring the research fields, what was formerly the Key Technologies field will now be further developed as a new field dedicated to Information research.

- The Association is building several new institutes for simulation and data technologies, as well as cyber security, and is currently planning to establish a new Helmholtz Center for Information Security in Saarbrücken.

- A large number of young scientists will be trained in this area using innovative methods, as part of a new network of regional Helmholtz Information & Data Science Schools.

- The Helmholtz Information & Data Science Incubator, initiated by the President of the Helmholtz Association, functions as a new type of association-wide think tank and driver of innovation by bolstering individual initiatives.

3. THE HELMHOLTZ INFORMATION & DATA SCIENCE INCUBATOR

In June 2016, the Helmholtz Association launched the **Information & Data Science Incubator** with the intention of intelligently pooling the Association's diverse and decentralized expertise. To this end, each Helmholtz Center sent two top scientists to take part in the Incubator. These 36 professionals cover a wide range of subject areas, each brings a great deal of innovative know-how to the table. Experts from renowned companies involved in research offer additional support to the Helmholtz scientists.

The Incubator integrates existing, forward-thinking initiatives across the Helmholtz Association using a **bottom-up process**. By aggregating and consolidating expertise from the various Centers on a regular basis, visionary developments in the area of Information & Data Science are being made possible. The Helmholtz Incubator is charting a new course to create a strong interdisciplinary field of the future.

The Helmholtz Incubator is currently pursuing the **following objectives:**

- Regular interaction of creative minds from across the Association

- Creating the foundations for innovative, interdisciplinary networks and approaches

- Identifying forward-thinking fields and disruptive pilot projects

- Planning long-term, systemic, and structural changes

The President has allocated 30 million euros from the Helmholtz Association's Initiative and Networking Fund to support innovative activities and promising pilot projects. Together with additional funding from the individual Helmholtz Centers, an amount of **more than 60 million euros is currently available to develop this area.** Renowned international experts decide how these funds are to be allocated by way of a strict peer-review process.

4. THE HELMHOLTZ INCUBATOR AS
A VISIONARY CATALYST

By systematically implementing these objectives across the Association, the Helmholtz Incubator is providing **new impetus** for initiatives in the area of Information & Data Science that 1) continually contribute to the development of new topics, 2) will be further developed as essential structural initiatives, and 3) build foundations for addressing large-scale research topics in the long term. These initiatives serve as links to expand interactions and structures far beyond the Helmholtz Association and realize enormous collective potential.

PILOT PROJECTS

The Helmholtz Incubator met for several days of talks in October 2016, December 2016, and May 2017. In December 2016, the Incubator experts identified a series of **new multidisciplinary research focal points**. Drawing on these points, groups of researchers then prepared applications for disruptive, forward-thinking pilot projects intended to provide fresh impetus in the field of Information & Data Science, enabling new interdisciplinary collaborations. **Five initial pilot projects** were selected in June 2017 and allocated 18 million euros in funding. The pilot projects are already being implemented. They are developing state-of-the-art techniques for data analysis and finding new ways to generate knowledge from data sets. Additional innovative pilot projects are soon to follow.

Selected Pilot Projects

The *"Helmholtz Analytics Framework"* *aims to develop cutting-edge, basic analytics techniques using a co-design approach involving domain scientists and information experts. This approach also facilitates rapid dissemination of the methods.*

The *"Sparse2Big"* *pilot project provides a methodology for working with sparse and partially flawed massive data sets. Using imputation and robust modeling, high-quality Big Data can be achieved – especially for health research.*

The *"Reduced Complexity Models"* *project aims to develop intelligently reduced models of complex computer simulations. These allow vast quantities of data from complex, mathematical computer models to be handled.*

The *"Automated Scientific Discovery"* *project uses innovative and reliable artificial intelligence methods to automatically explore correlations in large volumes of scientific data.*

> The **_"Imaging at the Limit"_** pilot project develops interdisciplinary methods for image reconstruction that transform measurements into images. This increases the overall information extraction, as imaging techniques represent an essential source of information in virtually every field of research.

COMPREHENSIVE APPROACHES TO STRUCTURAL CHALLENGES

During the initial Incubator workshops in October 2016, experts identified **key structural challenges** facing the Helmholtz Association and the research landscape as a whole. In spring 2017, focused working groups specified these challenges in more concrete terms and developed effective concepts for the future. In May 2017, these results from the working groups were discussed and further refined at an Incubator workshop involving all of the experts.

In September 2017, the **scientific and commercial Directors of the Helmholtz Association convened for an extraordinary meeting** to look at the five identified structural challenges in more depth. They recommended these be elaborated upon in greater detail in order to prepare for potential long-term, association-wide implementation. In October and November 2017, the Incubator working groups further developed these ideas and are now working to produce **comprehensive concepts for long-term, structural changes** at the Helmholtz Association by May 2018 in the following five areas:

Next Generation of Scientists

Information & Data Science is a key field for global leaders in research and development. However, when comparing the volume of emerging opportunities and challenges to the number of qualified personnel in science, research, and development, there is a clear shortage. There is a particular lack of talent with a profound command of modern information technologies and a dedicated scientific domain as well. Thus, our aim is to invest massively in a wide-reaching network of postgraduate training opportunities in this area, thereby developing a new environment with a critical mass of researchers in both Information & Data Science and a scientific domain. This network will be comprised of half a dozen dedicated, regional Helmholtz Information & Data Science Schools, which are currently being designed in close cooperation with partner universities. As the Helmholtz Information & Data Science Academy (HIDA), they will work to bring together the most relevant training activities in the field.

Machine Learning

The development of artificial intelligence – and specific, cutting-edge machine learning and deep learning processes - is already having a disruptive effect across the entire research landscape. These areas of development are enabling scientists to address completely new questions and are significantly improving research, both quantitatively and qualitatively. In order to make this enormous potential even more usable in its own research and for society at large, the Helmholtz Association is developing groundbreaking technology platforms to combine its tremendous expertise in methodology development and scientific

disciplines. This allows the Association's own experts to join forces with other partners from the wider scientific system.

Imaging Techniques

Image recognition and imaging techniques are essential for all areas of modern research and development, especially now that an ever-increasing amount of information is collected and analyzed in image format. The Helmholtz Association plans to systematically develop and improve cutting-edge approaches and make these more accessible in order to facilitate new discoveries in many fields of research.

Adding Value to Research Data with Metadata

Efficient and sustainable research data management is essential as researchers are collecting ever more diverse and complex data sets. In order to generate knowledge from data, reproduce results, and reuse data (FAIR principles), both the context and process of collection must be well documented. Building on initial initiatives in this area, the Helmholtz Association is strengthening its expertise in the collection, management, and analysis of metadata. This allows the Association to make a significant contribution to constructing effective national, European, and international research data infrastructures (NDFI, EOSC, etc.).

Basic Technologies and Services

Research into pressing societal challenges is increasingly being carried out across decentralized, dynamic research networks, with enormous amounts of data being collected, evaluated, and stored at various locations. An efficient network of supportive services (e.g., clouds, remote computing, authentication networks, etc.) is necessary to best support advanced collaborative research. The Helmholtz Association has been developing competences and capacities in these areas for decades and will work to systematically pool these core technologies for the entire Association and its partners. In doing so, the Association will also draw on the preliminary work of the existing Helmholtz Data Federation (HDF) and continue to develop its initial findings in a tailored manner.

5. INFORMATION & DATA SCIENCE VISION

NEXT STEPS

The Incubator's first concrete results and projects demonstrate the **enormous potential** of this approach. The pilot projects are providing pioneering impetus and realizing completely new opportunities for interdisciplinary collaboration. The Helmholtz Incubator allows us to face fundamental challenges in the field of Information & Data Science with partners in a comprehensive, science-driven, and collaborative manner. Together, internal and external experts analyze scientific needs and engage in an ambitious process intended to develop sustainable, long-term potential courses of action. Members of top management at the Helmholtz Association then find ways to strategically integrate these options across the entire Association. Continuous involvement at all levels allows for detailed, comprehensive planning as the basis for a concerted, needs-based implementation of findings. As an influential player in the research landscape, the Helmholtz Association continues to provide **fresh impetus** to the field and develop sustainable, **long-term solutions**.

The following **concrete steps** are planned for the near future:

- By the middle of 2018, five concepts addressing the structural changes outlined above will be developed. This will involve efforts by the entire Helmholtz Incubator, its working groups, and additional researchers as well as external experts in a dynamic, concerted process. In September 2018, these concepts will be submitted to the Association management and recommended for implementation.

- These concepts supported by the entire Association will then be rapidly implemented and, following completion of an initial development phase, presented to partners in the wider German research system as an offer for collaboration and participation.

- The Helmholtz Incubator will continue to gather leading experts from the Helmholtz Association to develop new approaches for disruptive research projects. These new pilot projects will complement the established projects, further expanding research on Information & Data Science with new strategies.

- Additional Incubator workshops are intended to promote the networking of activities in Information & Data Science, continue to develop the new environment, and create new offerings for science and society at large.

VISION

The Helmholtz Association supports visionary association-wide efforts to advance the field of Information & Data Science as necessary in its entirety. The Helmholtz Incubator functions as a **think tank** that will also support the Helmholtz Association as a powerful driver of innovation in the future. In order to manage a diverse set of requirements, the Incubator will continue to evolve going forward, adapt to current trends, and consider new perspectives.

The Helmholtz Incubator works to create a **community with a critical mass of researchers across the Association** in the areas of Information & Data Science and other scientific domains. This ensures that new approaches are continuously being fed back into the individual disciplines and better promotes the spread of new methodologies and knowledge. This novel approach puts the Helmholtz Association in a leading research position both nationally and internationally.

The activities of the Helmholtz Incubator offer manifold links to outside partners as well as numerous opportunities for symbiotic collaboration with other competent parties. They form a solid foundation for **large-scale research projects** in both national and European contexts. The Helmholtz Incubator is designed to work as a platform to encourage interaction far beyond the Helmholtz Association and realize the enormous potential of these topics of the future together. The Helmholtz Association and its partners are dedicated to making important contributions to building a national research data infrastructure.

Pioneering research requires **collaboration between renowned experts from all over the world**. The same can be said for attracting and retaining excellent researchers and young scientists. The Helmholtz Incubator and its programs help to build lasting networks between research sites in Germany and attractive international partners in the field of Information & Data Science.

Drawing on the Association's manifold strengths, the Helmholtz Incubator is **giving new momentum to the development of advanced interdisciplinary research fields**. It will continue to develop powerful concepts and courses of action and make sound proposals for the allocation of additional funding.

Finally, the Helmholtz Incubator positions the **Helmholtz Association as a formative player** in the field of Information & Data Science. It provides decisive impetus for the Helmholtz Association's involvement in this topical field going forward, in both the German and international research landscapes.

www.helmholtz.de