

The PAHN-PaN Consortium

Particle, Astroparticle, Hadron & Nuclear Physics accelerate the NFDI



Thomas Schörner-Sadenius (DESY)
Karlsruhe, 30 September 2019

The NFDI

... And what happened so far, and budget



Die nationale Forschungsdateninfrastruktur (NFDI) soll die Datenbestände von Wissenschaft und Forschung systematisch erschließen, nachhaltig sichern und zugänglich machen sowie (inter-)national vernetzen. Sie wird in einem aus der Wissenschaft getriebenen Prozess als vernetzte Struktur eigeninitiativ agierender Konsortien aufgebaut werden.



March 2019	Submission of extended abstract
4 Jul 2019	Submission of letter of intent 22 binding and 23 (3) non-binding letters submitted
30 Aug 2019	DFG governance workshop
15 Oct 2019	Deadline for proposal submission
5/6 Dec 2019:	Evaluation
1 Oct 2020:	Start of funding

Budget: Max grant sum 5 MEUR, including 22% overheads, corresponding to ~52 FTE / year.

The NFDI

... and discrimination against ErUM-Data

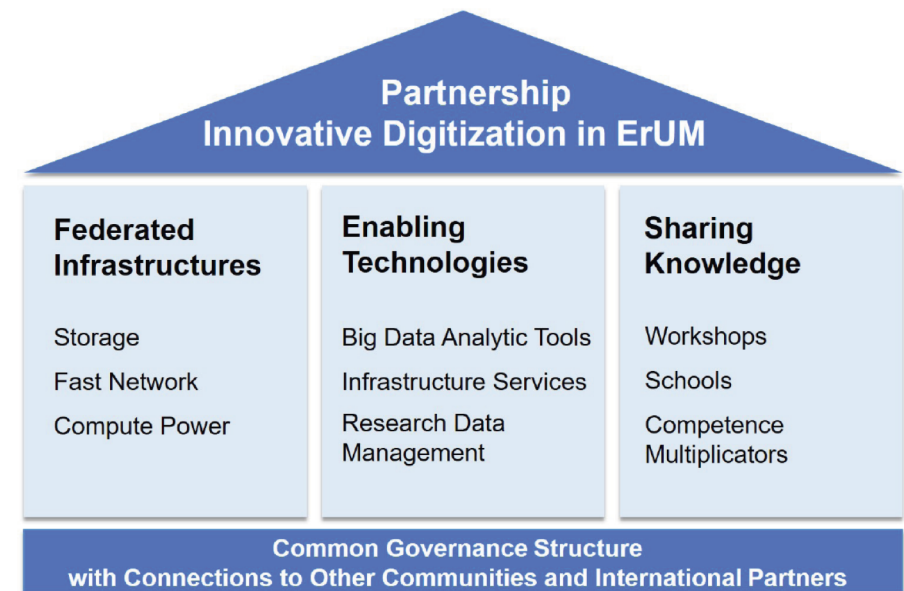


Difficult, unfortunately.

One clear distinction:

- NFDI focuses on the curation of scientific data across all disciplines.
- Emphasis in NFDI is on the development and provision of data management infrastructures and services within and beyond the consortium.

However, significant overlap especially in the „Enabling Technologies“ part of the ErUM Partnership for Innovative Digitization.



The PAHN-PaN Consortium



Applicant: Deutsches Elektronen-Synchrotron (DESY)

Co-applicants:

RWTH Aachen University, University of Bielefeld, Rheinische Friedrich-Wilhelms-Universität Bonn, TU Darmstadt, TU Dortmund, Technische Universität Dresden, Friedrich-Alexander-Universität Erlangen, Frankfurt Institute of Advanced Studies (FIAS), Forschungszentrum Jülich (FZJ), Albert-Ludwigs-Universität Freiburg, Georg-August-Universität Göttingen, GSI Helmholtzzentrum für Schwerionenforschung GmbH, Universität Hamburg, Ruprecht-Karls-Universität Heidelberg, Karlsruhe Institute of Technology (KIT), Universität zu Köln, Johannes Gutenberg-Universität Mainz, Ludwig-Maximilians-Universität München, Westfälische Wilhelms-Universität Münster, Universität Regensburg, Bergische Universität Wuppertal

Participants:

Ruhr-Universität Bochum, Technische Universität Braunschweig, CERN, Deutsche Physikalische Gesellschaft, Helmholtz-Zentrum Dresden Rossendorf (HZDR), Johann-Wolfgang-Goethe-Universität Frankfurt, Justus-Liebig-Universität Giessen, Friedrich-Schiller-Universität Jena, Helmholtz-Institut Jena, Max Planck-Institut für Physik München, Technische Universität München, Max Planck-Institut für Kernphysik Heidelberg, Universität Siegen, Leibniz Information Center for Science and Technology (TIB), Julius-Maximilians-Universität Würzburg

Task Areas

And cross-cutting topics

- Task area 1: Developing workflows and tools for data management
- Task area 2: FAIR Data Life Cycle Concepts and Open Data
- Task area 3: Data analysis procedures and services
- Task area 4: Real-time data analysis and selection
-
- Cross-cutting topic A: Synergies
- Cross-cutting topic B: Services
- Cross-cutting topic C: Training, education, outreach

TA4 deals centrally with the question of how the necessary data reduction can be achieved in real-time with a minimum loss of physically relevant information. Also using prototyping and demonstrator setups for hardware-software-code design, special intelligent algorithms and methods for data quality assurance will be developed. The focus of these developments is on exemplary problem solutions that can be applied in the entire community.

TA1 addresses developments and their technical integration for the PAHN-PaN consortium and the embedding of solutions in the international context. The required services and infrastructure will be developed towards industry standard to allow shared usage with other scientific disciplines, cost effective operations and maintainability.

The goal of **TA2** is to develop and apply a concept of research data management for the entire field of PAHN research. Concept must meet broad requirements to be applicable to other fields of research. A demonstrator should generalize and combine the different developments in the various research areas of the consortium. In addition, the complete life cycle and each individual step of the handling of research data must be equally integrated.

The projects of **TA3** maximise the scientific value of data by providing high-level services and tools for analysis. Procedures for interactive analysis and inference on large data sets are provided as well as numerical methods and hardware-specific optimisations. Based on these developments, automated machine learning services greatly simplify the training and deployment of deep learning. Finally, generic studies of systematic uncertainties, stability and visualisations of machine learning improve the usefulness of these new algorithms.

Task Areas

And cross-cutting topics

- Task area 1: Developing workflows and tools for data management
- Task area 2: FAIR Data Life Cycle Concepts and Open Data
- Task area 3: Data analysis procedures and services
- Task area 4: Real-time data analysis and selection 1g
-
- Cross-cutting topic A: Synergies
- Cross-cutting topic B: Services
- Cross-cutting topic C: Training, education, outreach

In the task area **"Cross-cutting topic C: Training, education, outreach"**, a focus is placed on teaching young students and early-career researches in PAHN-PaN-relevant topics, mentor and advise them, and disseminate concepts, tools and results to the general public. Teaching material is provided for university education, and courses and hands-on masterclass are provided for high-schools in order to inspire them and to get them involved.

One aim of the NFDI is to create synergies between different fields of science, and with task area **"Cross-cutting topic A: Synergies"** we will contribute to this effort by following up concrete and existing joint projects or ideas with other consortia, and by explicitly looking for new ideas to be exploited in the future.

Already now there are some services operated for and within the PAHN-PaN consortium. Sizable effort that is spent in this task areas is targeted at identifying, improving and delivering existing services, or developments towards new ones. The **cross-cutting topic activity „Services“** identifies the needs for services in the consortium and beyond in order to foster the development of generic services that address broad spectrum of needs, ideally from multiple communities.



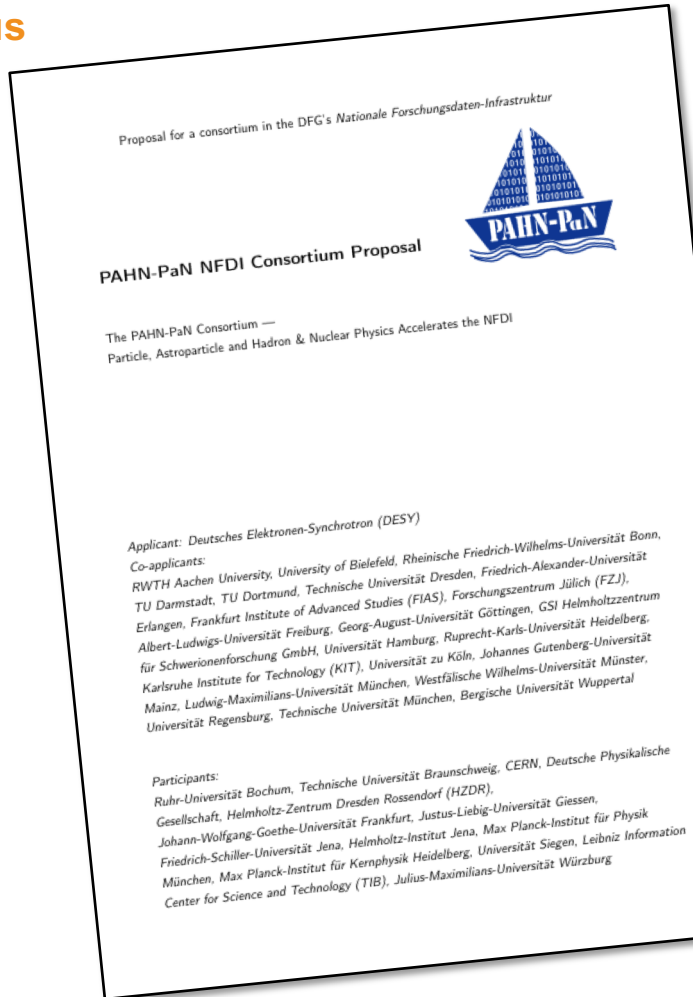
Task Areas


And cross-cutting topics



Task area 1 "Development of workflows and tools for data management" (A. Heiss, C. Wissing)		
1.1 Harnessing heterogeneous and opportunistic resources	NN	
1.2 Enabling access to data and metadata	NN	
1.3 Automation and optimisation of big data management workflows	NN	
Task area 2 "FAIR data lifecycle concepts and open data" (A. Haungs, C. Schwarz)		
2.1 Concepts and standardisation of metadata and data curation	NN	
2.2 Publication and preservation of analysis workflows	NN	
2.3 Open data and data mining	NN	
Task area 3 "Data analysis procedures and services" (G. Kasieczka, T. Kuhr)		
3.1 Analysis services and tools for large data sets	NN	
3.2 Inference on large data sets	NN	
3.3 Automated ML services	NN	
3.4 Development of numerical methods and hardware-specific optimisation		
3.5 Systematics, stability and visualisation		
Task area 4 "Real-time data analysis and selection" (A. Redelbach, W. Rhode)		
4.1 Algorithm optimisation for specialised hardware	NN	
4.2 Real-time feature extraction and pattern recognition	NN	
4.3 Hierarchical data selection systems	NN	
4.4 Prototyping and demonstrator setup	NN	
4.5 ML methods for data quality assurance	NN	
Cross-cutting topic A "Synergies" (G. Duckeck, T. Schörner-Sadenius)		
A.1 Future Synergies	NN	all
A.2 Astrophysics and astronomy / Astro-NFDI	NN	Astro-NFDI
A.3 Photon and ion physics / DAPHNE	NN	DAPHNE
A.4 Mathematics / MaRDI	NN	MaRDI
A.5 Medicine	NN	NN
A.6 High-performance computing	NN	NFDI4CS4NFDI
Cross-cutting topic B "Services" (K. Schwarz, C. Wissing)		
B.1 Authentication and authorisation	NN	
B.2 Cloud interfaces	NN	
B.3 Catalogues and database services	NN	
B.4 Storage and data management tools	NN	
B.5 Future technologies	NN	
Cross-cutting topic C "Professional training, education, outreach" (K. Kröninger, A. Quast)		
C.1 Professional training	NN	
C.2 Education	NN	
C.3 Outreach	NN	

Status

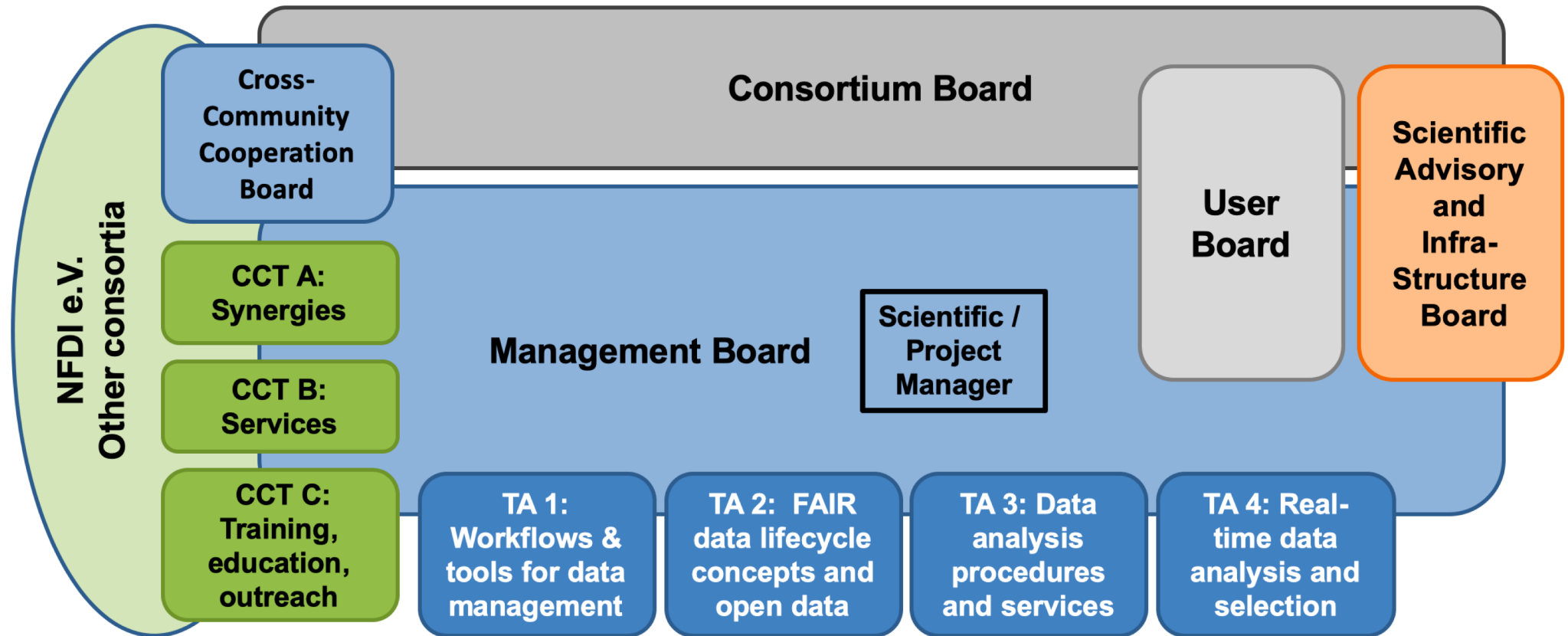


	
<hr/>	
<h1>Contents</h1>	
1	General Information 3
2	Consortium 8
2.1	Research domains or research methods addressed by the consortium, objectives 8
2.2	Composition of the consortium and its embedding in the community of interest . . . 11
2.3	The consortium within the NFDI 11
2.4	International networking 11
2.5	Organisational structure and viability 12
2.6	Operating model 15
3	Research Data Management Strategy 17
3.1	Metadata standards 19
3.2	Implementation of FAIR principles and data quality assurance 19
3.3	Services provided by the consortium 20
4	Work Programme 24
4.1	Overview of task areas 24
4.2	Task area 1: Developing workflows and tools for data management 27
4.3	Task area 2: FAIR data lifecycle concepts and open data 35
4.4	Task area 3: Data analysis procedures and services 48
4.5	Task area 4: Real-time data analysis and selection 62
4.6	Cross-cutting topic A: Synergies 75
4.7	Cross-cutting topic B: Services 81
4.8	Cross-cutting topic C: Training, education, outreach 86
4.9	Administrative tasks 90
5	Overall Funding Request 92
6	General Compliance 94
Appendix 1. Bibliography and list of references 95	
Appendix 2. Curricula vitae and lists of publications 96	
Appendix 3. Letters of commitment by the participants 97	
Appendix 4. Letters of support 97	
Appendix 5: Initial board composition in PAHN-PAN 98	
Appendix 6: The "Berlin Declaration" of NFDI consortia 98	
<hr/>	
2	PAHN-PaN NFDI Consortium Proposal

- > 100 pages and filling quickly
- Especially introductory text parts still missing
- Numbers being consolidated
- 2,5 weeks to go ...

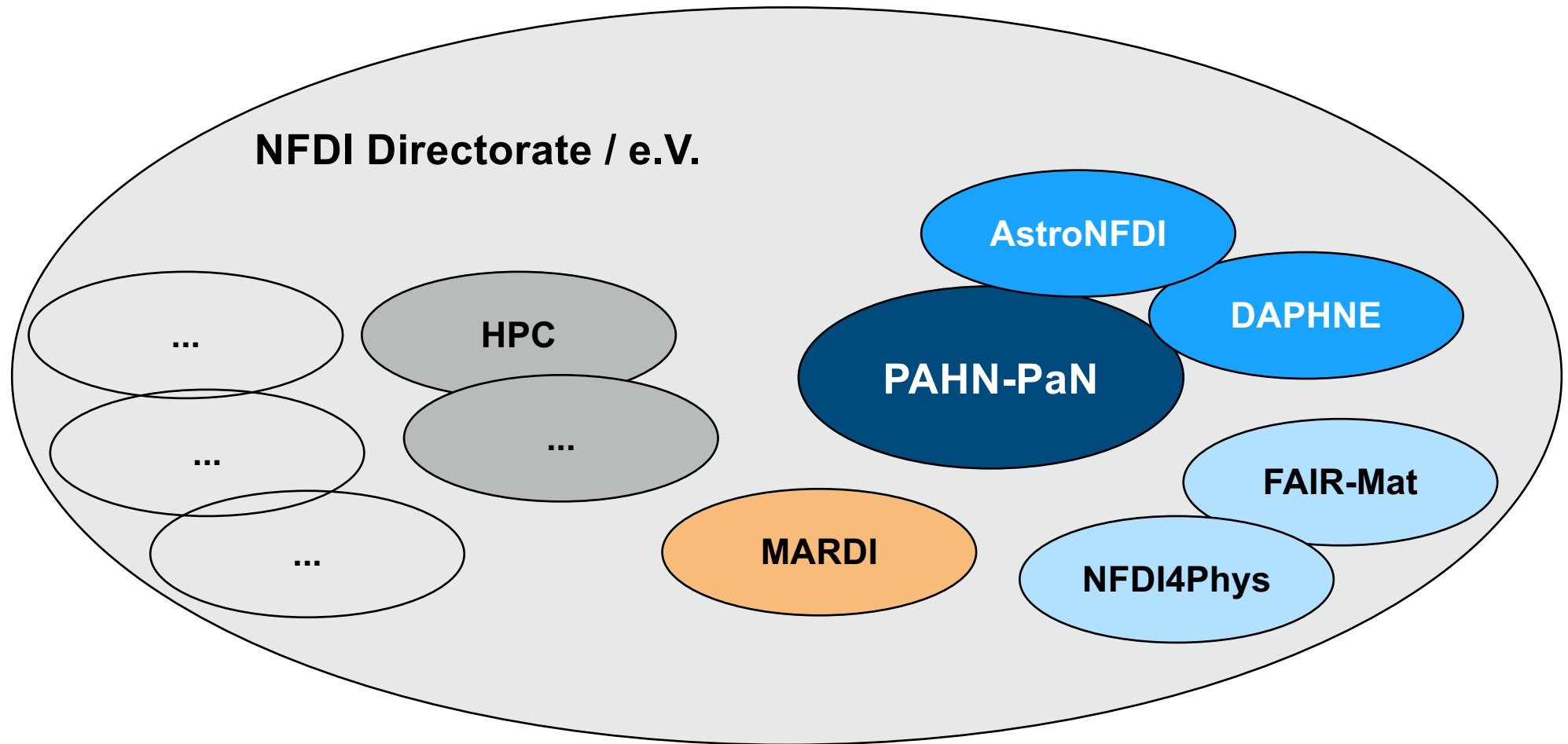
Governance and Organisation

Of the PAHN-PaN consortium



A Word on Synergies

With other NFDI consortia



Thank you



Contact

DESY. Deutsches
Elektronen-Synchrotron

www.desy.de

Thomas Schörner-Sadenius

DESY FM / CMS

E-Mail: thomas.schoerner@desy.de

Phone: +49 40 8998 3429

Backup