

PAHN-PaN:

Task 2: FAIR data lifecycle concepts and open data

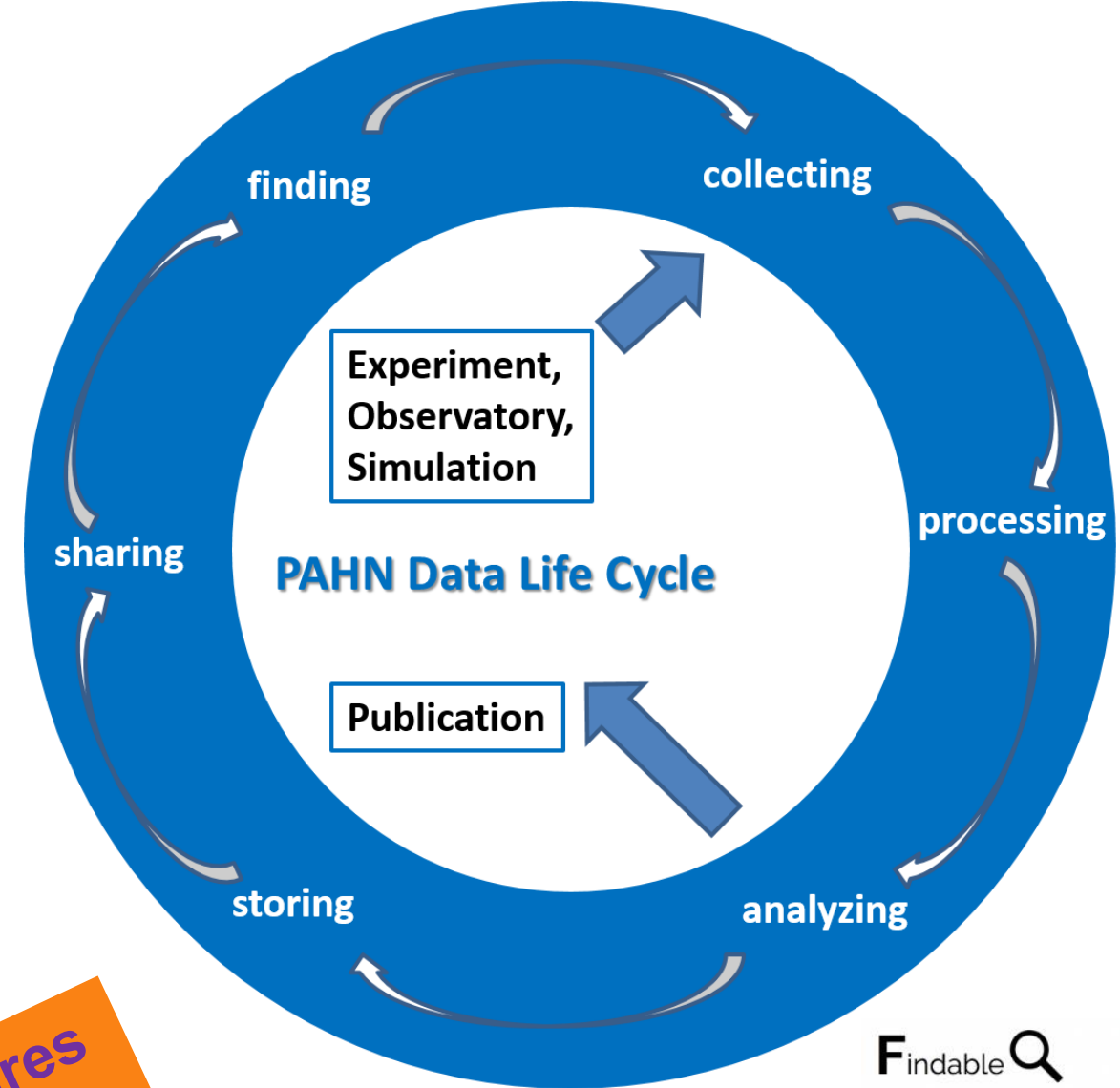
Workshop | PAHN-PaN
26 September 2019, Kassel

Andreas Haungs (KIT, KAT) Kilian Schwarz (GSI, KHuK)



TA2 - Intro

- **GOAL** = foster/build/provide concepts and demonstrators (infrastructure) for a general PAHN Data Management Plan
- Where possible, common standards should be established to foster interoperability (FAIR)
- Importance of “data stewards” to manage the data life cycle and to act as a curator for metadata
- Allowing Data (i) availability (ii) method development (iii) analysis (iv) education (v) open access (vi) archive
- Further points to be considered:
 1. Connection to other TAs
 2. Connection to PAHN physics
 3. Connection to other Consortia
 4. Relevance for NFDI



→ Requires dedicated manpower

Findable 🔍
Accessible 🖱️
Interoperable ⚙️
Reusable ♻️

TA2: Work Packages

1. Concepts and standardisation of Metadata and Data Curation

- Data format
- Data quality
- Simulation as Data
- Metadata (DOI)
- Data storage
- Generalize the different ideas and levels available in PAHN physics

2. Publication and Preservation of Analysis Workflows

- Consider the workflows (to obtain the published data) as 'FAIR Data' themselves.
- Support middleware
- Improve and Generalize workflows as tools
- Include small and midsize experiments

3. Open Data (and Data Mining)...

- Data preservation
- Data publication
- Provide access to PAHN data
- User management
- Legal aspects

TA2: Deliverables

1. Demonstrator for PAHN platform of data life cycle

- ...
- ...
- ...

2. Data Management Plan Template for PAHN physics

- ...
- ...
- ...

3. Publication? Handbook?

- ...
- ...
- ...

TA2: Services

1. Catalogues (with real data)

- What needs a catalogue
- How to use a catalogue
- How to manage a catalogue
- ...

2. Data Management Plan

- Standardization of meta data
- How to treat data and meta data and workflows FAIR
- ...

3. Platform for a data life cycle of PAHN data (concepts)

- How to use the hardware
- Data mining
- User management

TA2: Risks and Mitigation

1. Requirements justified

- FTE ?
- Eigenanteil ?
- ...

2. Entire community under one hat

- ...
- ...

3. Shall we do a SWAP analysis?

- ...

TA2: Budget Plan

TA2	11,5	11,5	11	10,5	10,5	55
2.1	3,75	3,75	3,75	3,75	3,75	18,75
Bonn	0,25	0,25	0,25	0,25	0,25	1,25
Darmstadt	0,5	0,5	0,5	0,5	0,5	2,5
GSI	1	1	1	1	1	5
KIT	0,5	0,5	0,5	0,5	0,5	2,5
Köln	0,5	0,5	0,5	0,5	0,5	2,5
Jülich	0	0	0	0	0	0
Regensburg	0,5	0,5	0,5	0,5	0,5	2,5
Bielefeld	0,5	0,5	0,5	0,5	0,5	2,5
2.2	3,75	3,75	3,25	3,25	3,25	17,25
Bonn	0,75	0,75	0,75	0,75	0,75	3,75
Darmstadt	0,5	0,5	0,5	0,5	0,5	2,5
Dortmund	1	1	0,5	0,5	0,5	3,5
Köln	0,5	0,5	0,5	0,5	0,5	2,5
Jülich	0,5	0,5	0,5	0,5	0,5	2,5
Regensburg	0	0	0	0	0	0
Bielefeld	0,5	0,5	0,5	0,5	0,5	2,5
2.3	4	4	4	3,5	3,5	19
DESY	0,5	0,5	0,5	0,5	0,5	2,5
Erlangen	1	1	1	1	1	5
GSI	0,5	0,5	0,5	0,5	0,5	2,5
KIT	1	1	1	1	1	5
Köln	0,5	0,5	0,5	0	0	1,5
Mainz	0,5	0,5	0,5	0,5	0,5	2,5

TA2: Budget Plan

Summe Institute						
	Jahr	1	2	3	4	5
5	Bonn	1	1	1	1	1
5	Darmstadt	1	1	1	1	1
2,5	DESY	0,5	0,5	0,5	0,5	0,5
3,5	Dortmund	1	1	0,5	0,5	0,5
5	Erlangen	1	1	1	1	1
10	Gitter	2	2	2	2	2
7,5	GSI	1,5	1,5	1,5	1,5	1,5
7,5	KIT	1,5	1,5	1,5	1,5	1,5
6,5	Köln	1,5	1,5	1,5	1	1
2,5	Mainz	0,5	0,5	0,5	0,5	0,5
		11,5	11,5	11,0	10,5	10,5
	Gitter:					
2,5	Regensburg	0,5	0,5	0,5	0,5	0,5
5	Bielefeld	1	1	1	1	1
2,5	Jülich	0,5	0,5	0,5	0,5	0,5
55	Ziel:	55				

TA2: Budget Plan

WP2.1	Concepts and standardisation of Metadata and Data Curation						
Jahr		1	2	3	4	5	
Gitter							
	Regensburg	0,5	0,5	0,5	0,5	0,5	
	Bielefeld	0,5	0,5	0,5	0,5	0,5	
	Jülich	0,5	0,5	0,5	0,5	0	
Bonn		0,5	0,5	0	0,5	0,5	
Darmstadt		0,5	0,5	0,5	0,5	0,5	
GSI		1	1	1	1	1	
KIT		0,5	0,5	0,5	0,5	0,5	
Köln		0,5	0,5	0,5	0,5	0,5	
TiB		0	0	0	0	0	
WP2.2	Publication and Preservation of Analysis Workflows						
Gitter							
	Regensburg	0,5	0,5	0,5	0,5	0	
	Bielefeld	0,5	0,5	0,5	0,5	0,5	
	Jülich	0,5	0,5	0,5	0,5	0	
Bonn							
Darmstadt		0,5	0,5	0,5	0,5	0,5	
DESY		0	0	0	0	0	
Köln		0,5	0,5	0,5	0,5	0,5	
Mainz		0	0	0	0	0	
TiB		0	0	0	0	0	
Dortmund		1	1	0,5	0,5	0,5	
WP2.3	Open Data and Data Mining						
Erlangen		1	1	1	1	1	
DESY		0,5	0,5	0,5	0,5	0,5	
KIT		1	1	1	1	1	
Mainz		0,5	0,5	0,5	0,5	0,5	
GSI		0,5	0,5	0,5	0,5	0,5	
Köln ?							

TA2: How important is to incorporate the Lol?

- **Exploit added-value for full NFDI**
 - in applying workflows in data management and data processing
 - in applying workflows in handling open data
- **Build-up and operate a data management infrastructure**
 - to serve for (large-scale) experiments
 - to handle large-scale simulations (for theory and experiment)
 - pursue the concept of data lakes
 - ensure interoperability for different functionalities: data archive – data access – public data – simulations – method developments – big data analytics
 - includes middleware for storage systems and data mining
- **Define Standards**
 - for data formats
 - for metadata
 - for data curation
 - for data preservation
 - for open data

→ Vision/Mission?
→ Towards a virtual
German Science Cloud ?

TA2: Next Steps

- **Further Text parts?**
 - Gitter Gruppe Text (06.09) is in overleaf and will be incorporated
 - Update Olaf Kaczmarek (18.09) will be considered
 - Frankfurt (Uni) (lattice QCD?)
 - Text von Matthias Schott, Mainz (05.09) will be included (wp 3)
 - Dortmund: new Text will be incorporated mail of 17.09

Will be done next few days

**→ TA 2 phone meeting next week: 2/10/2019 11:00
or 4/10/2019 11:00 oder 14:00**

TA2: Next Steps

- **How to write coherent proposal?**
 - Include further text (work) parts
 - Examples for Backbone Use
 - ...
- **What we are providing/delivering to NFDI, ErUM-Data, EOSC,...?**
 - ...
 - ...
- **Who is doing what?**
 -
 -
- **And still:**
 1. Connection to other TAs
 2. Connection to PAHN physics
 3. Connection to other Consortia
 4. Relevance for NFDI