Data Processing on diverse Resources

Change of resource landscape for PAHN expected over the next years

- Today: Mainly Grid sites and local batch farms (often dedicated to PAHN)
- Future: More and more shared resources including
  - Allocations on High Performances Computers (HPCs)
  - Cloud resources (either commercially or owned research communities)
  - Container infrastructures like Kubernetes

DESY plans

- Providing of Cloud resources (OpenStack), Container service (Kubernetes), Jupyter Notebooks
  - Relates to Analysis Services in Task Area 3
- Transparent integration and adaption of existing workflows of the PAHN community to the “new” infrastructure
- Ensure coherence with other efforts
  - International: EOSC, ESCAPE
  - National: HIFIS by Helmholtz

Towards the “Data Lake”

Optimization of data management middleware stack

- Quality of Service (QoS)
- Caching
- Combine high level tools to federate data storage and analysis systems on the national and international level

DESY plans

- Authentication and Authorization Infrastructure (AAI)
  - Implementation of X509 alternatives, like OpenID connect
- Continue development and integration of relevant storage and data management middleware
- Integration of data lake concepts into scientific workflows and user applications
  - (Performance) Validation and adaption of existing workflow
- Extension of ILDG (International Lattice Data Grid)

Relates closely to Task Area 2