

# Collaborative Research Centre 940 - "Volition and Cognitive Control"

-

## Data Management, Workflow Optimization and Science Gateway

Richard Grunzke

[richard.grunzke@tu-dresden.de](mailto:richard.grunzke@tu-dresden.de)

# Overall Goal of the CRC

---

- Elucidate cognitive and neural mechanisms underlying adaptive volitional control as well as impaired control in selected mental disorders

# Project Overview

---

- 12M Euro DFG funding for this second phase
- Start at 6/2016
- 4 year duration
- 16 research and 4 support projects with 1 to 3 PIs each
- 35 full time equivalents, 1 at ZIH

# Starting Point of the INF Project

---

- In Phase 1 project execution optimized for research tasks, e.g.
  - Z2: NIC-Database, central storage for MRI data, and processing of neuroimaging data with the NICePype workflow engine
  - C3: SQL-based system RedCap for non-imaging data
- Advanced individual solutions requiring integration
- Worldwide awareness towards openness and sharing of research data
    - FAIR data principles [[datafairport.org](http://datafairport.org)]
    - Findable, Accessible, Interoperable, Re-usable

# Starting Point of the INF Project

- Overview page of the NiCePype image processing framework
- Design Concept of the NiCePype

**NicePype**  
Neuroimaging Center Python Pipelines

Overview Data Processing Jobs Administration Logout Project/Study: B4 (030)

Project Overview: B4 (030)

Found a total of 1 data processing jobs for this project in the database, whereof 0 queued, 0 running, 0 have errors, 0 aborted, and 1 finished

Project	B4 (030)
Department	asyn
Patients	148
Data basedir	/media/mf0hd/atenidata
Last data sync	Wed Feb 3 17:34:42 2016

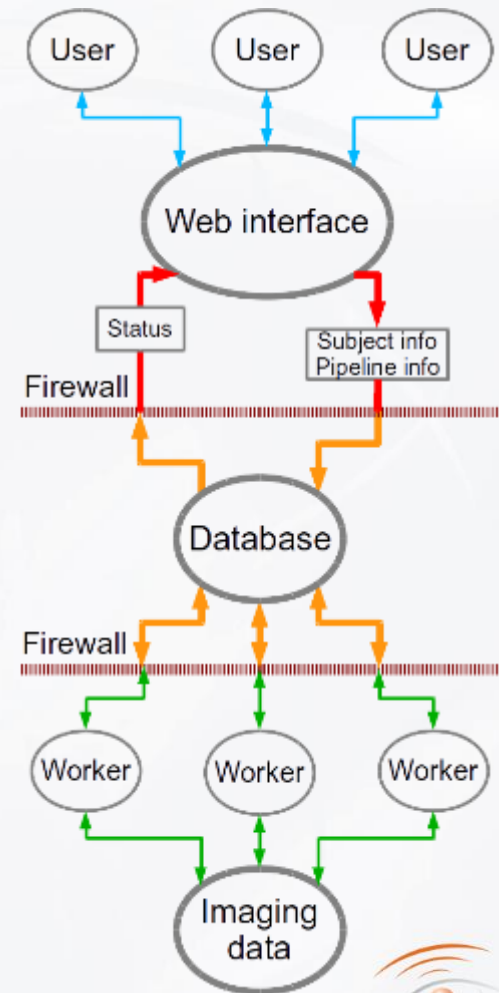
Request project updates

Status of pipeline information

Last update: Wed Feb 3 17:21:57 2016

Request pipeline information updates

Neuro Imaging Center Dresden



# Motivation

---

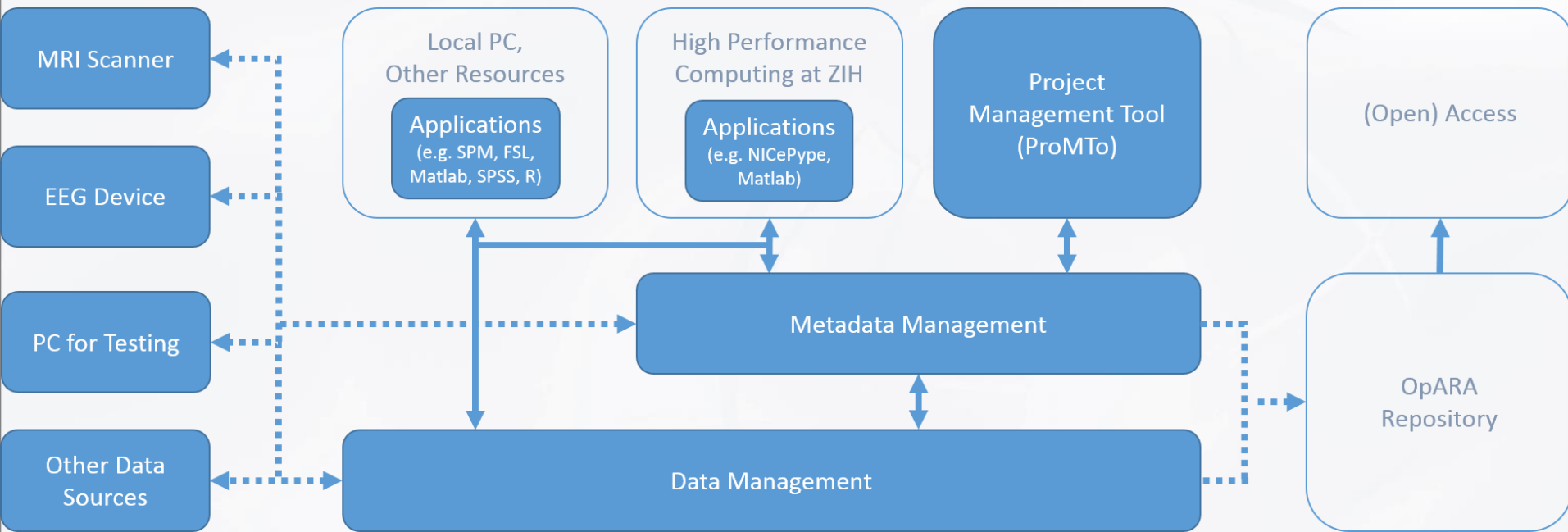
- Enable systematic and long-term use of data according FAIR principles
- Efficient use of existing subject data between projects (e.g. B3, B7, B8)
- Continuous gathering of complex and large (300 subjects) and varied (behavioral, clinical, fMRI, experience sampling) datasets (C1, Z2)
- Sharing and analysis of large datasets, e.g. 210 GB in 500.000 files for 200 subjects and, additionally, 4 TB after processing with NICEPype (C1, Z2)
- Multiple processing options and 10x faster measurements multiply demands

# INF Project Aims

---

- Design and implementation of a comprehensive data, metadata and project management system.
- Seamless and transparent access to data in HPC and user environments
- Data privacy and data security as integral part
- Facilitate Open Access, data sustainability, and re-use

# Overall Architecture





# Work Plan

---

- Analyses of data and metadata requirements
  - in cooperation with all projects
  - with consideration of phase 1 work and
  - creation of a data management plan
- Concept for data privacy and data security, Integral parts e. g.
  - possibility for anonymization before download
  - cooperation with official information security unit of TU Dresden

# Work Plan

---

- System Design and Implementation based on
  - requirement analysis
  - phase 1 work
- NICEPype Integration with
  - HPC resources
  - metadata framework, and
  - hardening of web-interface

# Work Plan

---

- Continuous Evaluation
- Training and documentation
- Data migration of first funding period data into new system

# INF Role within CRC

---

- INF builds upon
  - Phase 1 work from various projects (e.g. B3, B9, C1, C3, Z2)
  - Comprehensive ZIH experiences and existing technologies
- INF provides
  - Working environment for CRC researchers to significantly enhance their scientific capabilities
  - Consulting for all projects with focus on data management and analysis
  - Tutorials and training for optimal research efficiency

# Conclusion

---

- Integrated system enabling scientists to more quickly generate better scientific results based on their data

# Thank you

---



**TECHNISCHE  
UNIVERSITÄT  
DRESDEN**

Richard Grunzke

14



Center for Information Services &  
High Performance Computing