ASTRO - NFDI
Astronomy, Astrophysics, Astroparticle Physics
Astrophysics in the NFDI
Matthias Hoeft
Thüringer Landessternwarte Tautenburg
on behalf of ASTRO - NFDI
Community

- Rat Deutscher Sternwarten
- More than half of German astronomical institutes are actively participating in Astro-NFDI, others support the consortium
- LRZ, KIT, FZJ as HPC (and data storage) providers are involved
- participating in ESCAPE, SOLARNET and other European roadmap projects, as well as in EOSC
- involved in many international collaborations (instrument and facility building, surveys and satellite projects)
- International Virtual Observatory Alliance via
Our data is highly diverse

- Highly varying observational data types from
  - Various frequency ranges of EM spectrum
  - Gravitational data
  - Particle (decay) data

- Simulation data
  - Cosmological simulations
  - Starformation processes
  - Magneto Hydro Dynamical processes

- Experimental data
  + Biological, geophysical, meteorological data

- Almost no commercial value
- No privacy protection problems
What can we contribute to the NFDI?

• Connecting and working with multitudes of data structures and distributed archives
  − Collaborative use of distributed data archives
  − International reuse of data collections
  − Collaborative Research Environments

• Standardised data exchange protocols and metadata schemata
  − partially used and proved
  − International Virtual Observatory Alliance (+GAVO) travails (almost 20 years)
    • standardized access to data collections, interoperability, (partially FAIR)
    • registries for services and data archives

• Open Source Software
  − sustainable scientific software development

• Interdisciplinary contributions
  − Research Data Management Organiser (RDMO)
What do we need/expect building and collaborating in NFDI?

- Extension of FAIR data policies in our discipline
  - Interoperable, interdisciplinary standards und metadata, DOI
  - Support and enabling of collaborative research in national and international collaborations

- Building sustainable competence/data centers for (astronomical) data
  - Data management (curation, provenance, publication)
  - Data publication software
  - Solutions for 'last dirty mile' (small data collections)

- Code to the Data:
  - Collaborative Research Environments (CRE)
  - Machine learning
  - Distributed data processing

- Scientific Software Support
  - supporting generic Open Source Software (e.g. astropy)
  - managing data and connected software as units

- Resolving lossy data problems and managing 'live data'

- Connecting data and publications
Eco-System of scientific data

Data sources: Observations
Expected data production: several Exabyte/year

Data sources: Simulations
Expected data production: several hundred Petabyte/year

Metadata
Raw data
Quality Control Processing Curating

Metadata
Science-ready data
Standards, API

Metadata
Science-ready data
Standards, API

Metadata
Raw data
Quality Control Processing Curating

Scientific article
Astro-NFDI: Consortium Workpackages

Current Workpackages

1. Governance, Consortium Management
2. Distributed Services and Structures
3. Data Workflows
4. Software for Data
5. Data Irreversibility Challenges
6. Synergies / Interaction with other consortia
7. Training, Summerschools etc., (Education)
Structure of consortium
* organisation
  - starting structures
  - intermediate target structures
  - longterm target structures

* NFDI interaction
* European and International interaction

* policy development and implementation
Astro-NFDI: WP Distributed Services and Structures

Available resources and facilities:
assessment + development
Virtual Observatory
data archives
available services
interactions of facilities and users
applied metadata systems

Digital Data Identifiers
Metadata packages,
FAIR data requirements
Provenance

changing requirements:
’real time‘ astronomy vs. ‘conventional‘ astronomy
Astro-NFDI: WP Default Data Workflows

- from 'fresh' data to science ready data
- from science ready data to science data

* required metadata, components
* pipelining, verification
* curation, preservation, publication

* enabling small(er) data collectors

metadata packages,
FAIR data requirements
Astro-NFDI: WP Software in data access and processing

* (collaborative) access to remote data collections

* (collaborative) access to distributed data collections

* lossy data taking / processing vs. data integrity

* software for data management and maintenance
Astro-NFDI: WP Challenges in Data Irrevisiblity

Cooperation
- Real-time algorithms
- Federated infrastructures

PAHNPaN

Dyn. Archiving
Regional Computing
Archives

Distributed Sensors
Scaling Pipelines
Dyn. Filtering
Central Computing
Data Products

M. Kramer
Astro-NFDI: WP Synergies

Intersection with other consortia:
• common approaches / data structures
• form a working group between consortia (in physics at least)

- data management workflows
- metadata systems
- methods
- management of data access
- software development (simulations)
- lossy data vs. reproducibility