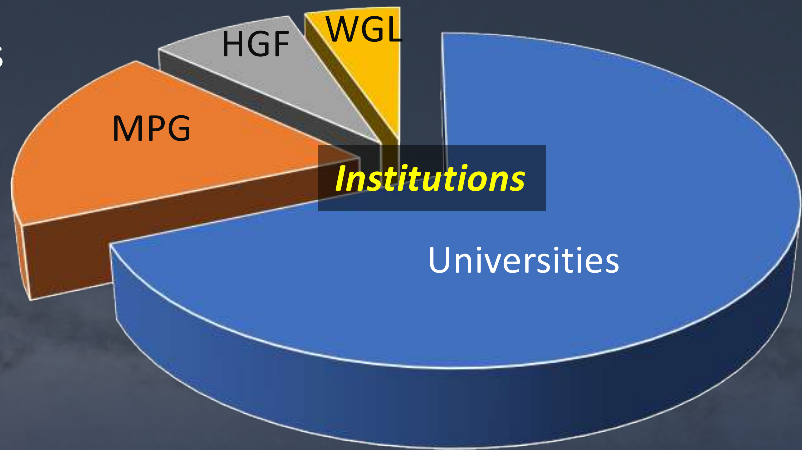

RDS@ErUM-Data
Rat der Deutschen Sternwarten (RDS)
Michael Kramer

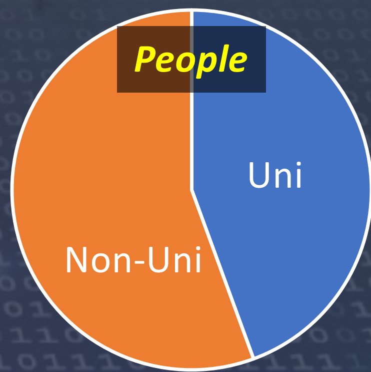
Council of German Observatories (RDS)

- Who are we?
- What are our special characteristics?
- What are our needs?
- Where do we see interfaces with NFDI & ErUM-Data?
- What are our plans?

Council of German Observatories (RDS)

- Executive committee of Astronomische Gesellschaft (AG)
- About 2700 researchers in the area of astronomy and astrophysics in Germany including about 700 PhD students*
- About 40 university and non-university institutions (Max-Planck, Leibniz, Helmholtz)
- Few big centres, many smaller institutions
- Tightly linked to international and intergovernmental organization and facilities (e.g. ESO, SKA, CTA, EST – see also EU ASTRONET Roadmap)
- For comprehensive description of status and goals see ”Denkschrift 2017”
- Representing about 20% of ERUM-Data Community**

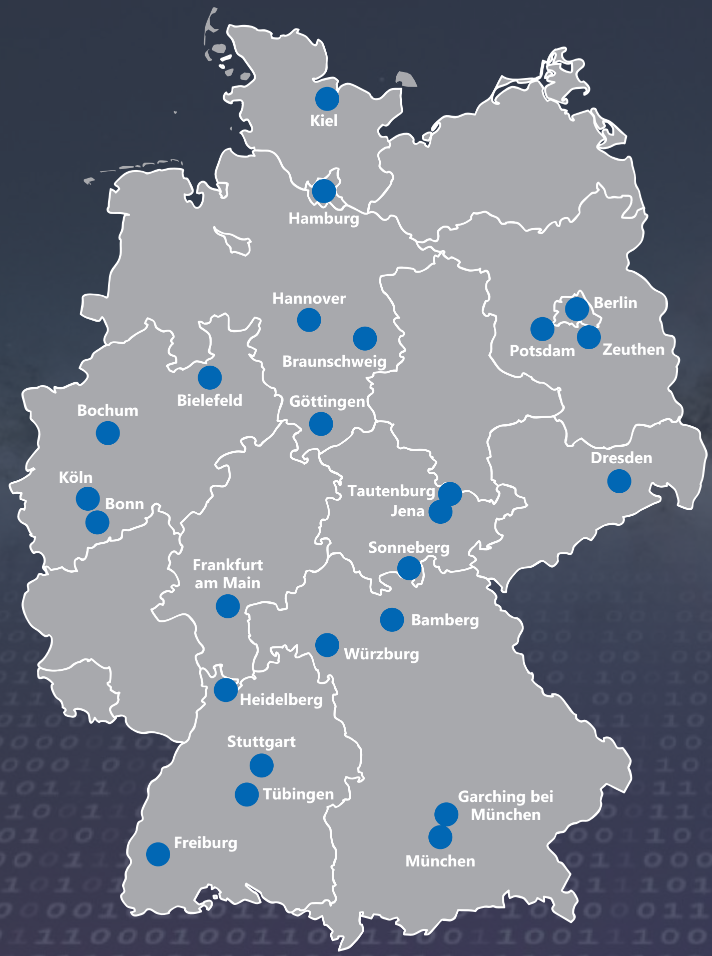




*as in Denkschrift (2017)

**as in Recommedations of the ErUM Committees (2019)

Heterogeneous, distributed, diverse… in many aspects

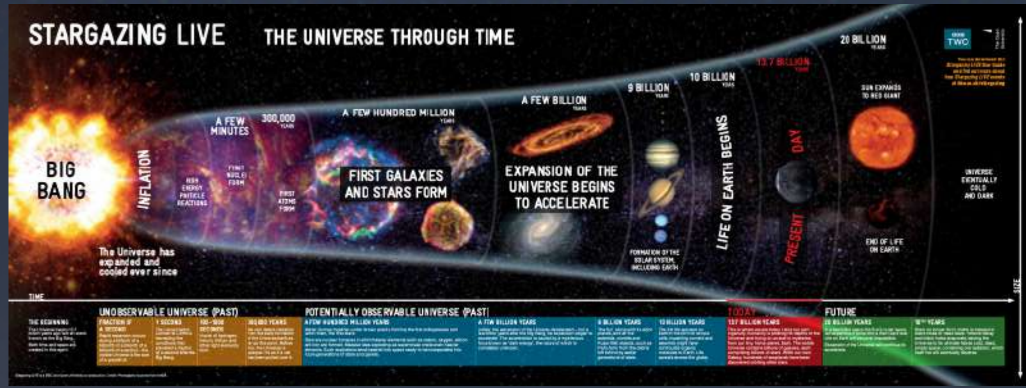


Denkschrift (2017)

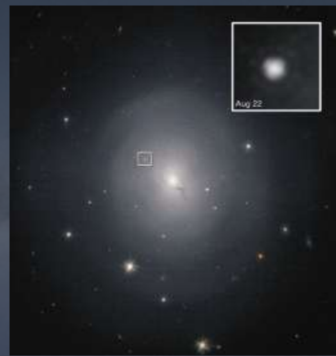
Denkschrift (2017)

- From ground to air & space
- Single telescopes & networks
- Multi-messenger science with and without triggers
- Exploiting EM spectrum and beyond
- Probing timescales from nanoseconds to centuries
- Involving professionals, amateurs, citizen scientists

Our Data …



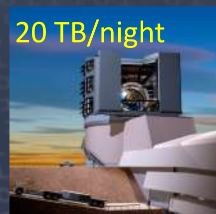
- From cosmic laboratories and experiments that cannot be controlled <u>or repeated</u>!
  Optimal data exploitation is essential
- Diverse in nature and multi-use
  - Electromagnetic (radio, submm, Optical/IR, UV, X-ray, Gamma-ray)
  - Gravitational Waves, particles and particle showers
  - Time Domain, Imaging, Spectroscopy, Polarization etc.
  - Computer Simulations
  - Laboratory Astrophysics (Astrochemistry, Astrobiology)



- Large data volumes (hitting technological boundaries)
  - Large scale surveys and dedicated experiments
  - Distributed & dynamic, non-conservative compression
  - Deep learning methods for archive coherence and fast & flexible interoperation
  - Problem of information loss and data irreversibility
- Particular features of our data
  - Mostly open data (except proprietary periods of ~1 year)
  - Lots (but far from all) accessible via virtual observatories
  - Huge amounts with increasingly grueling data rates (e.g. 150TB/s)

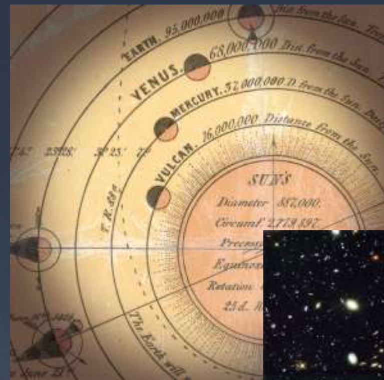
GERMAN ASTROPHYSICAL GAVO VIRTUAL OBSERVATORY

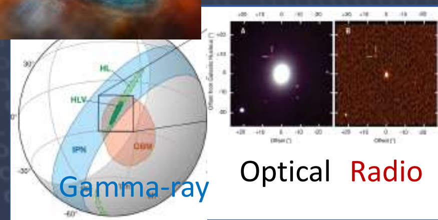
IVOA



Astrophysics is a Data and Discovery Science

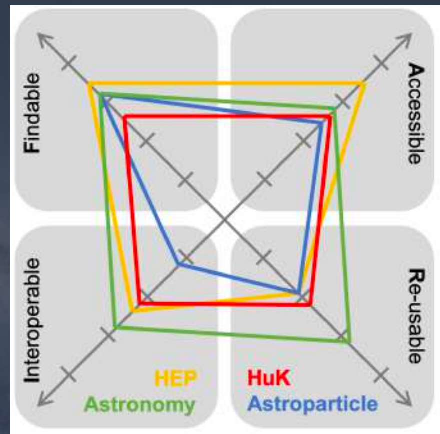
- FAIR: Findable, Accessible, Interoperable and Reusable is key!
- Historically:
  - Copernican revolution (long term observations)
  - Halley’s comet (inference from different records)
  - Neptune’s discovery (combining theoretical predictions & excellent maps)
  - Mercury perihelion advance leading to first test of general relativity
- Some examples:
  - Hubble Deep Fields (from exoplanets to cosmology)
  - 50% of publications utilizing Hubble Telescope data based on archival work
  - Gravitational Waves by merger of two neutron stars (multi-messenger approach)
- Future:
  - “Data Avalanche” (archiving Exabytes/yr) by large surveys and next generation of telescopes & simulations. But this has already started! (e.g. EHT, MeerKAT, LSST)
  - Data mining and deep learning – not only on isolated data sets but connected across instruments and archives





Needs to be addressed within PUNCH4NFDI

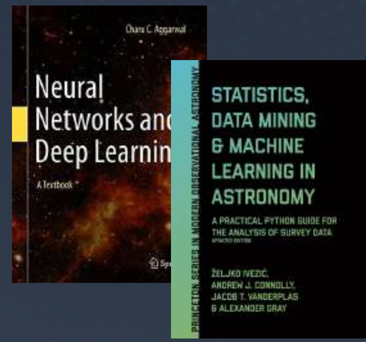
- Building sustainable competence centers for (astronomical) data
  - Data management (curation, provenance, publication) & Data publication software
  - Solutions for ‚last dirty mile' (small data collections)
  - Well defined, internationally compatible interfaces to big projects



- Extension of FAIR data policies in our discipline
  - Interoperable, interdisciplinary standards und metadata, DOI
- Code to the Data
  - Deep learning
  - Distributed data processing and Data Mining
- Scientific Software Curation
  - supporting generic Open Source Software (e.g. astropy, gammapy, psrchive)
  - managing data and connected software as units

- Dynamical Archives
  - Resolving lossy data problems and managing avalanche of online-data
  - Data irreversibility and information loss
  - Especially in the presence of unpredictable background noise

Needs to be addressed by ErUM-Data



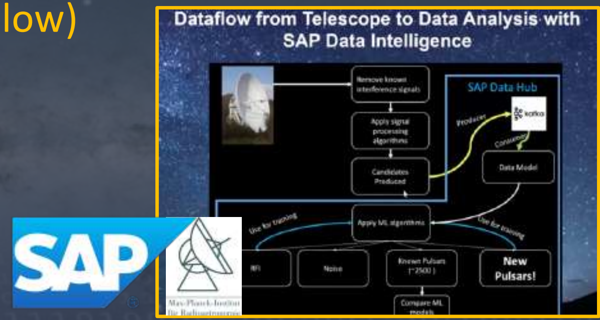
Depending on NFDI funding and implemented measures, we will have to (re)visit:

- Research data management
- Digital infrastructure (see below):

  - large amounts of compute power, sustainable storage, data lakes, dynamical archives & cloud computing
- Novel software & autonomous algorithms empowered by deep learning (see below)
- Knowledge transfer & cross-fertilization:

  - within community, with other communities and especially with industry

  - nationally and also internationally



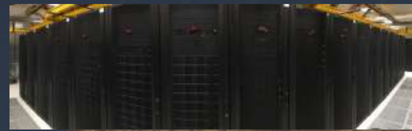
- Education & Training:

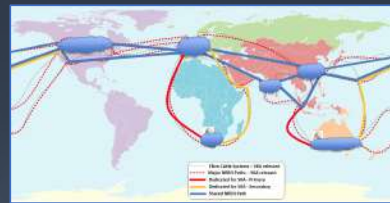
  - workshops and schools, especially for the analysis of Big Data, cross-cutting technology & code, etc.
- Publicizing efforts & results:

  - utilising large existing network of planetariums and public observatories

  - building on experience in citizen science and public outreach

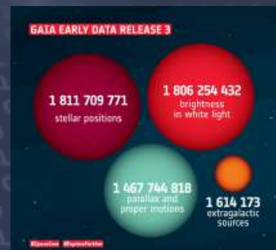
Important need for digital infrastructure





- Powerful compute infrastructure required also near instruments during data taking
- Sustained wide-bandwidth data transport capabilities (nationally and globally)
- Evolution of data lakes in combination with archives in suitable Tier-structure
  - dynamical, redundant and flexible
  - required by our heterogeneous and distributed data

- Challenge to guarantee sustainability of archives in landscape of international experiments
  - Global (or at least EU-) distribution of the data with access to cross-multi-messenger archives
  - Different approaches by different partners (e.g. ESO, ESA)
  - Even within organisations like ESO heterogeneous set of approaches (e.g. ALMA, APEX, VLT)
  - Financing & sustainability, incl. that of software, workflows & pipelines (e.g. many astronomers already use tools provided/financed internationally, e.g. AWS): national solution?



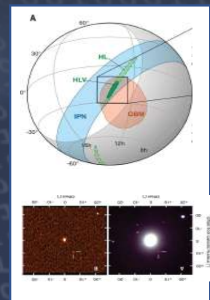
- General and important problem: certain urgency vs timeline of corresponding ErUM-Data call

First call of ErUM-Data: Software & Algorithms with focus on machine learning

Astronomy is a discovery science (e.g. CMB, pulsars, GBRs, FRBs etc.) in special requirements for AIs. Many applications needed during, after and even before measurements, e.g.:

- AI working on different hardware with real-time and off-line analysis

- Ability to spot and identify outlier, unusual and unexpected data items and patterns in huge data
- Ability to mine data autonomously and attach errors and uncertainties to AI results
- Ability to understand and exploit data with different properties simultaneously, including information from different domains (time, frequency, imaging etc.)
- Ability to utilise multi-messenger information
- Ability to utilise AI methods for simulations and synthetic observations
- Ability to access, combine and utilise archive data in combination or even planning of new data



- Ability to plan and organise multi-messenger observing campaigns (decide on telescopes to use, which information to obtain and which set-up to use, watch also weather, satellites etc.)
- Ability to provide dynamical, flexible survey strategy orchestrated between different telescopes and in response to triggers

Conclusions

- Large range of challenges that astro community expects to address with ErUM-Data given the increasing resource gap, especially for university groups
- Lots of opportunities to share knowledge & technology with other communities
- With ErUM-Data fewer discoveries will be missed, especially thanks to Deep Learning
- Important additional way to exploit investment in our infrastructures

- The next logical step in mankind's oldest science