KFS – Digitalization priorities

ErUM DATA Meeting

Jan. 15, 2021

Komitee Forschung mit Synchrotronstrahlung

Christian Gutt

KFS UNIVERSITÄT SIEGEN

Bridget Murphy

KFS C|A|U Christian-Albrechts-Universität zu Kiel

Outreach activities to and input from the community

- 2/2018 KFS user survey on needs for Data management
- 3/2018 KFS+KFN workshop #1 „Digital transformation“ at DESY
- 9/2018 KFS+KFN symposium at SNI Munich
- 6/2019 KFS+KFN workshop #2 „Digital transformation“ at DESY
- 2018-2020 Participation in the BMBF ErUM Data process
- 2019/2020 DAPHNE4NFDI proposal

Anton Barty
DESY photon science

Astrid Schneidewind
KFN
JÜLICH Forschungszentrum

Wiebke Lohstroh
KFN
Technische Universität München

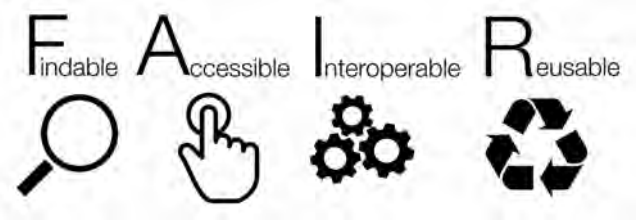
Research data management

DAPHNE4NFDI

- Universities
- Research institutions
- KFS-Committee for Synchrotron Research
- KFN-Committee for Neutron Research
- Large-scale photon and neutron research facilities
- Wider community

UNDER REVIEW

FAIR Data life cycle in DAPHNE4NFDI



Collect

Proposal

Capture
metadata
Control software
Data formats

Document
Workflow
Electronic Logbook
Measurement

Parameters
Sample
Environment
Instrument

Store

Curate
Validate
Store & Archive
Metadata &
Data

Publish Data

Access
Share Data
Access Controlled
or Open Access

Data catalogue
Federated
Searchable
Interlinked
Reuse

Publish
Papers
Patents
Presentations

Publish Results

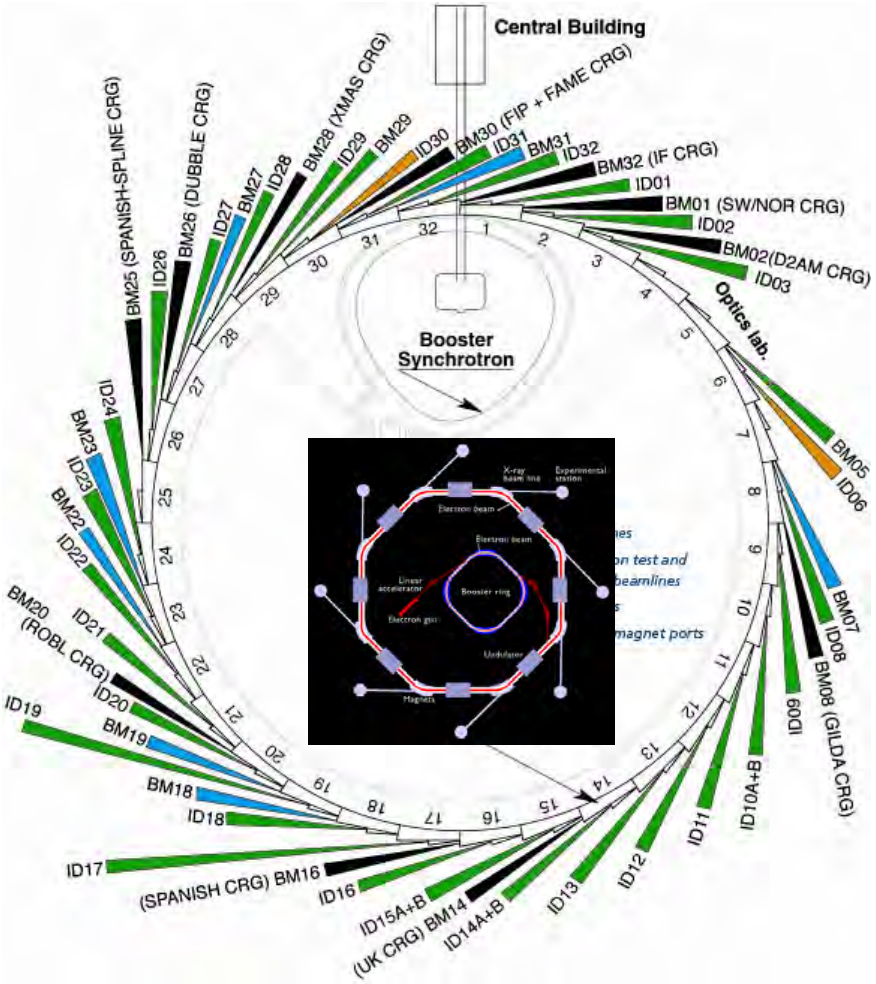
Analyse
Online
Integrate Data
Central access
Create Output

Process
Store Data
Validate Data

Visualise
Online
Live view
Reduction
Validate

Evaluate

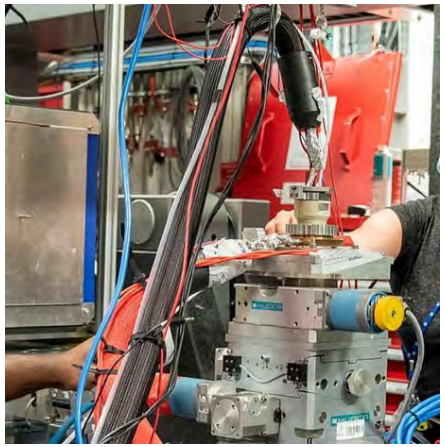
Challenge #1 - variety of disciplines, X-ray techniques and samples, set-ups and data



Degree of automatization and standardization in experiments and data varies across the techniques

Protein crystallography – experiment and data highly standardized

Most others: data, metadata highly user specific

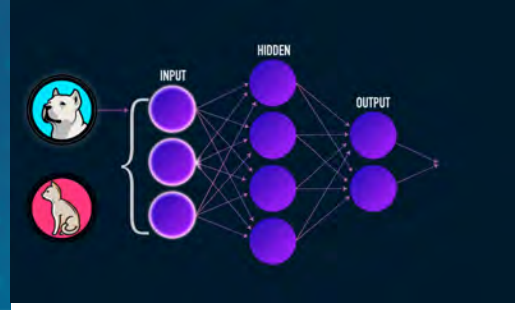


Challenge #2: increasing data rate

500 fps, 12 MPix
12MB/image
21 TB/hr (6 GB/s)

2PB free space
'full in a few days'

Per year (2019)	BESSY II	PETRA III	FLASH	EuXFEL	ELBE	MLZ	Total	ESRF EBS
Beamlines (BL) Measuring stations (MS)	27 BL 37 MS	22 BL 42 MS	2 BL 7 MS	3 BL 6 MS	1 BL 7 MS	26 BL 26 MS (+7 BL)	81 BL 125 MS	44
Experiments/ year	800	1400	38	25	70	620	approx. 2950	2000
Individual Users	1500	3000	350	499	100	600	approx. 6050	9000
User visits	3000	5000	500			>1000	>9500	
Publications	500	500	30	94	40	345	approx. 1500	2000
Data generated	1-2 PB	5 PB	1 PB	20 PB	0.5 PB	6TB	28.5 PB	> 20 PB
Expected 2025	8 PB	20 PB	4 PB	100 PB	2 PB	0.5 PB	134.5 PB	60 PB



New (ML/AI) software and algorithms for coping with data avalanche

Five possible objectives
- more welcome

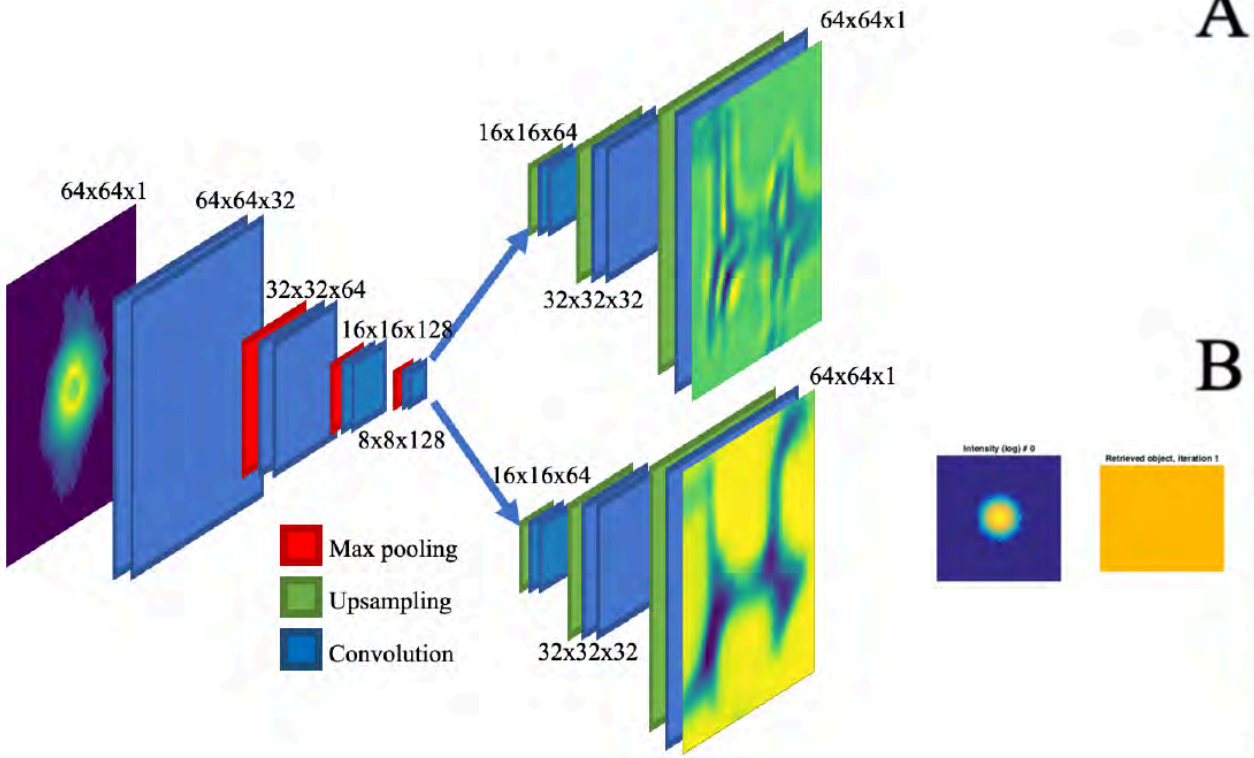
1. Faster and better classification / analysis of X-ray data and feedback loop to simulation/models/theory.

Example: Ptychography, ML training

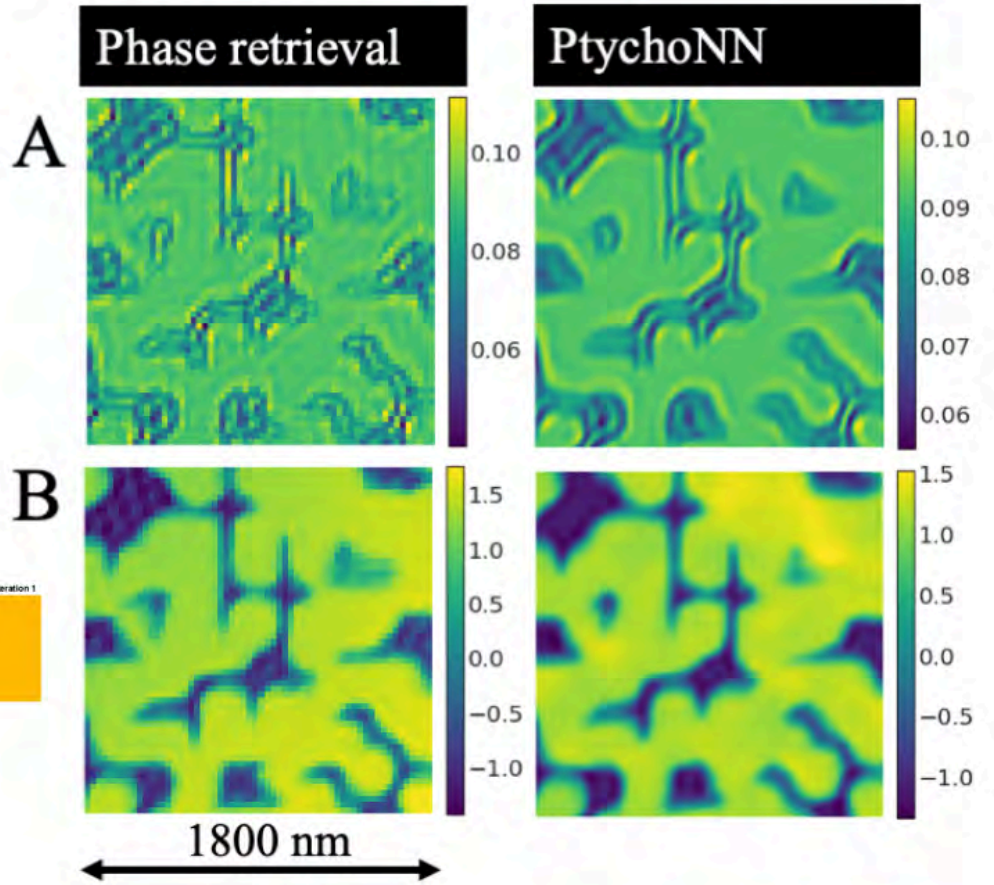
- on a small subset of the data/sample
- on 180k artifical samples

Allows to improve speed and relax on sampling conditions

Cherukara, M.J et al. *Sci Rep* **8,** 16520 (2018)
Cherukara, M.J et al. arXiv 2004.08247 (2019)



Neural network reconstruction 300 x faster than Phase retrieval- outlook for real time imaging



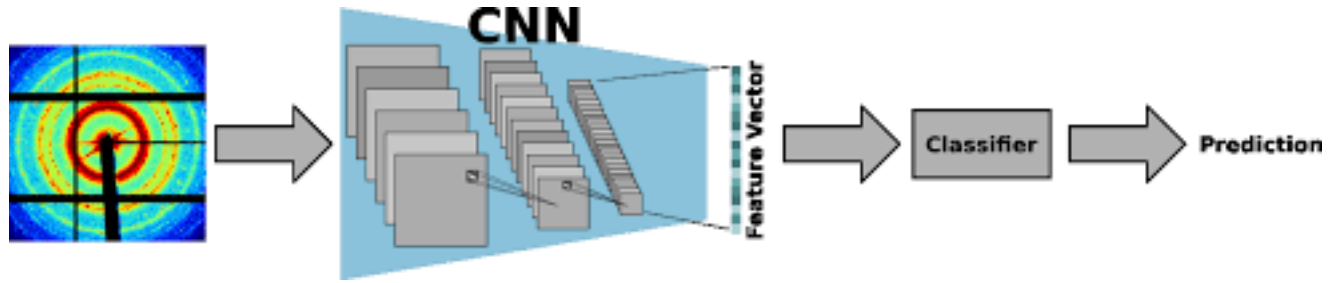
1. Faster and better classification / analysis of X-ray data and feedback loop to simulation/models/theory.

X-ray techniques

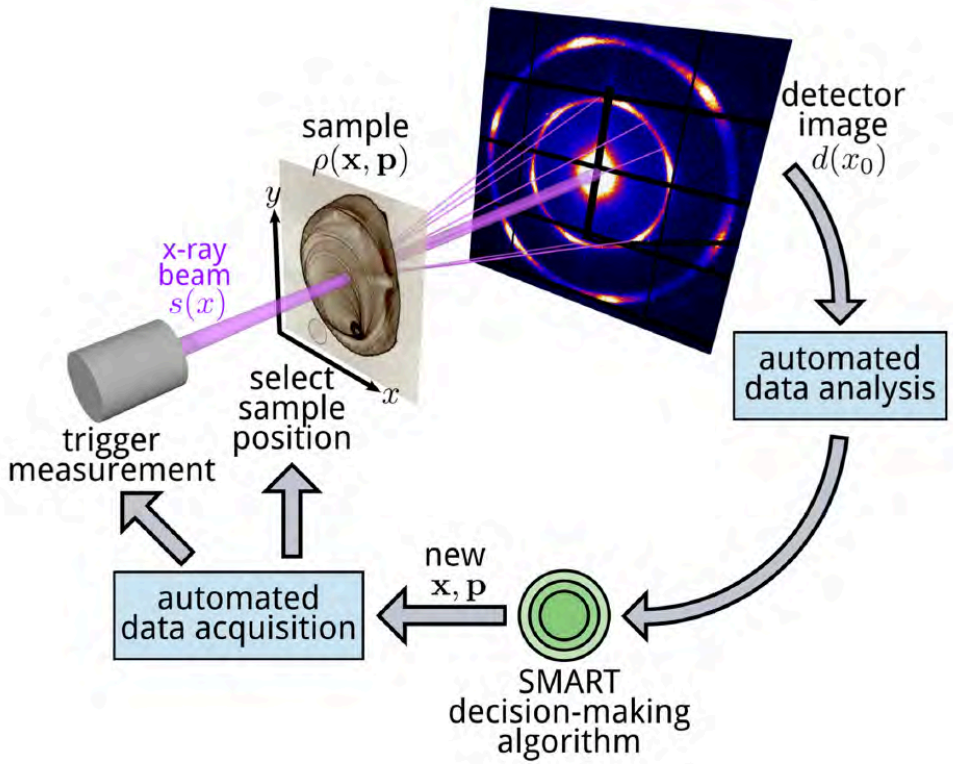
- X-ray reflectivity
- GISAXS
- SAXS
- WAXS
- Ptychography
- Tomography
- XANES, EXAFS
- X-ray spectroscopy
- Resonant magnetic scattering
- XPCS
- Serial crystallography
- FEL diffraction pattern classification
- …

ML techniques

- Pattern/image recognition and classification
- Machine learning for inversion problems
- Automated segmentation in tomography
- Superresolution
- …

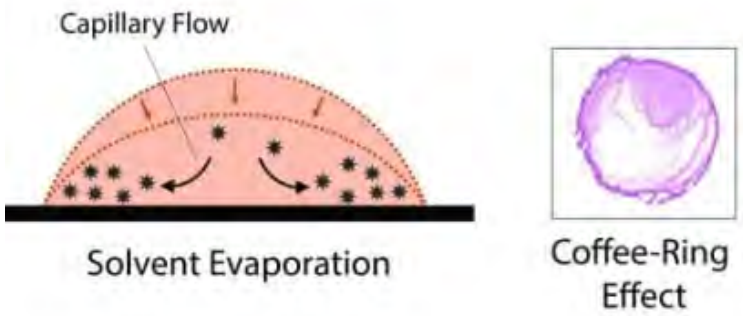


2. Towards autonomous experiments

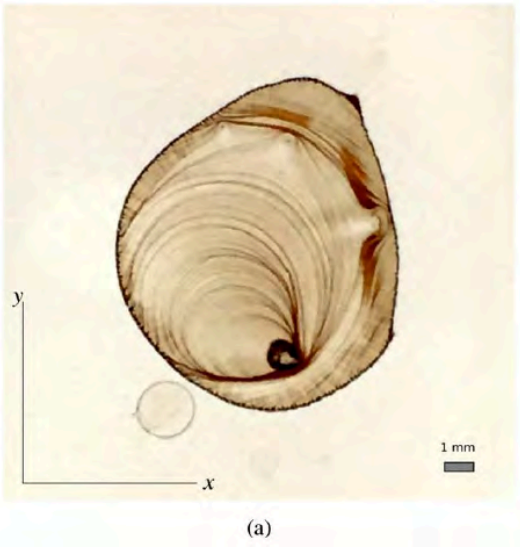


Noack et al. Scientific reports (2019) 9:11809

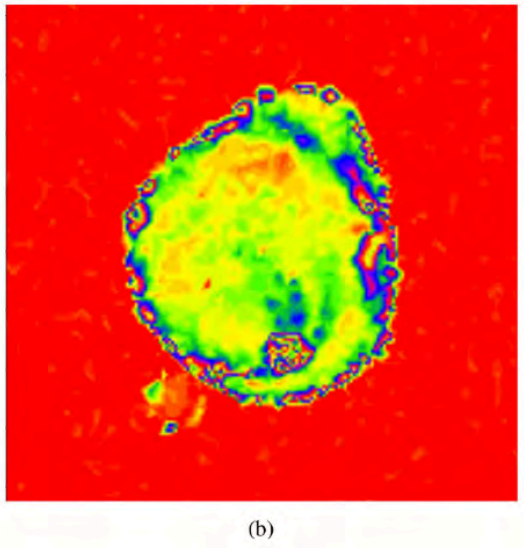
Coffee ring with nanoparticles



Sample



Grid scannning



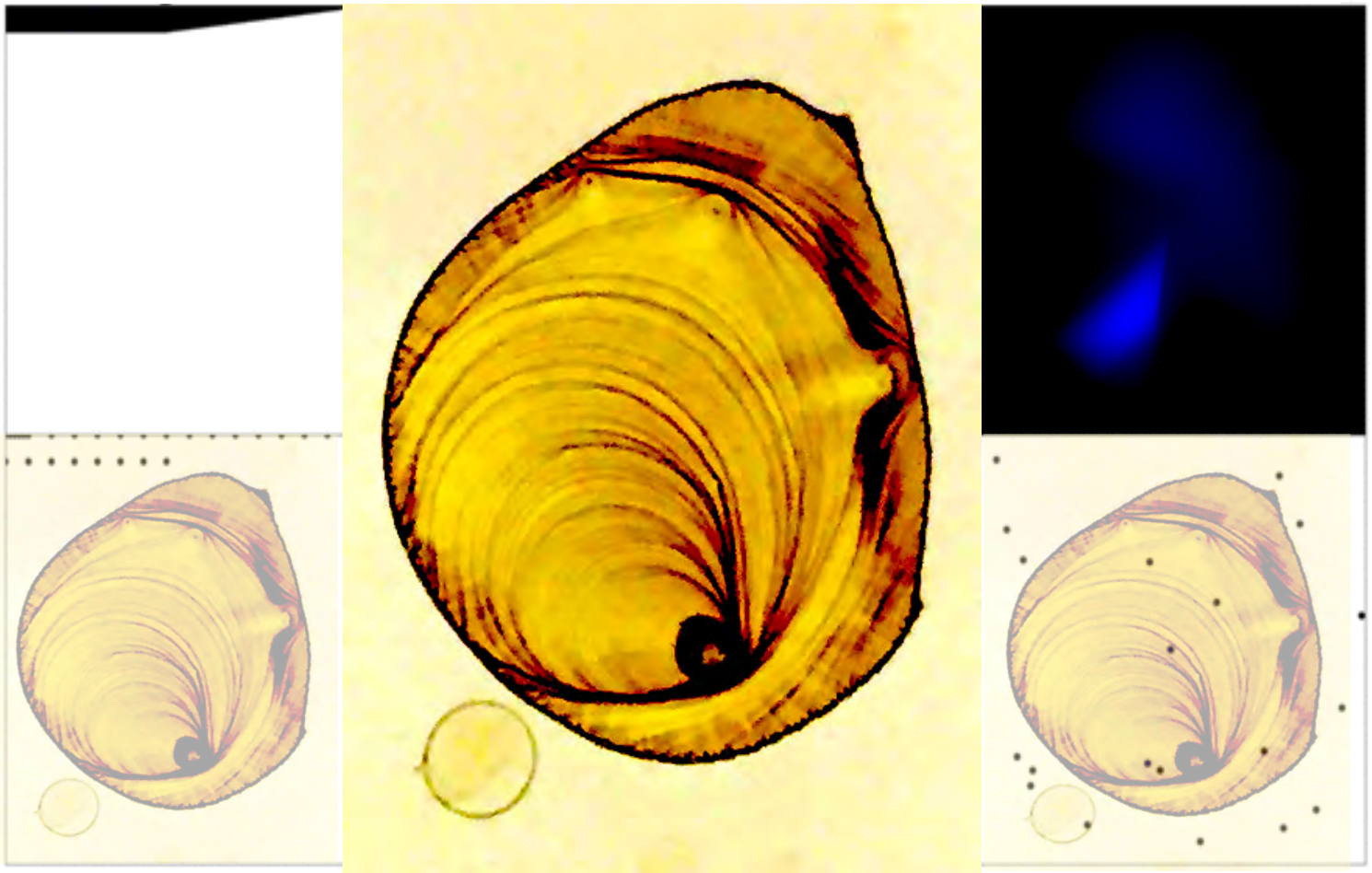
2. Towards autonomous experiments

Add:
beamline control
accelerator

Noack et al. Scientific reports (2019) 9:11809
Credit animation: Talk S. Jethian (Berkely + CAMERA), ESRF-ILL workshop 2019

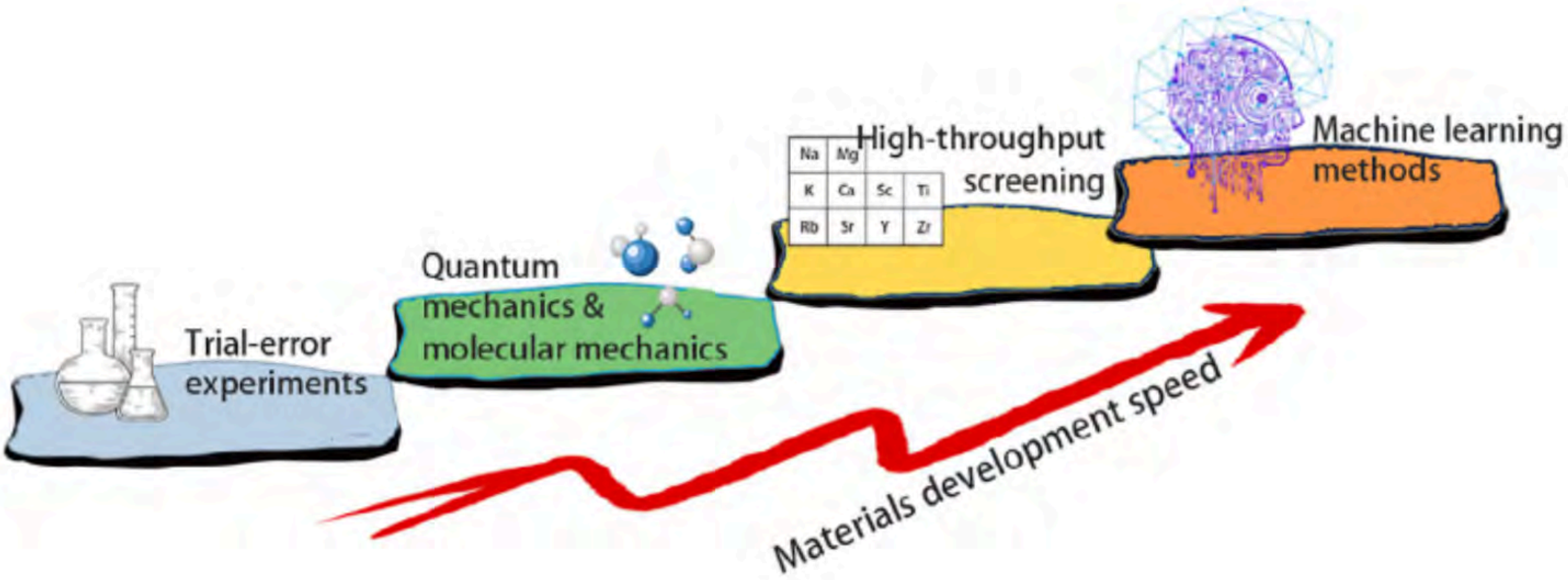
Grid

SMART



3. X-ray data mining
ML based cross analysis using data archives etc.

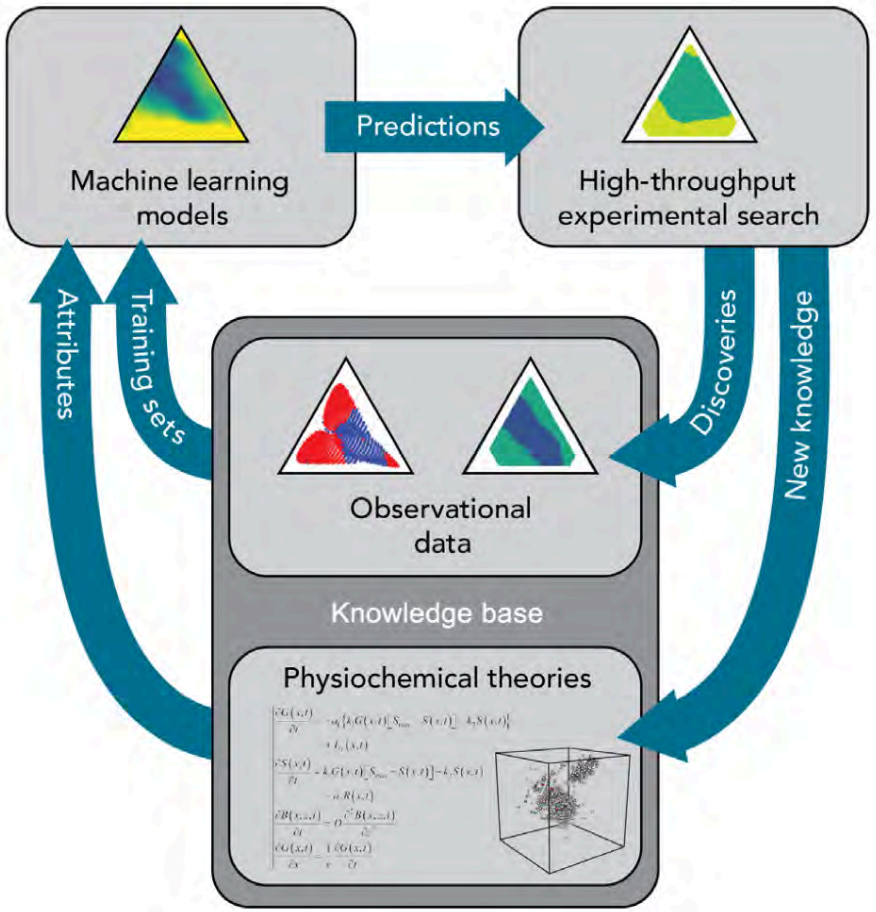
A. Chen et al. InfoMat. 2020;2:553–576.



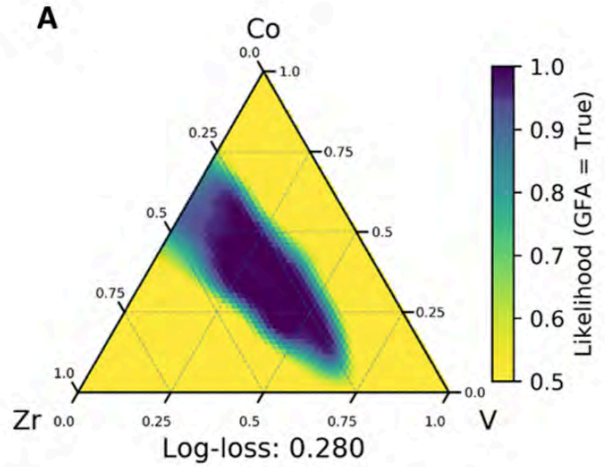
3. X-ray data mining
ML based cross analysis using repositories etc.

Ren et al., Sci. Adv. 2018;4:eaaq1566

Example: Predicting material properties



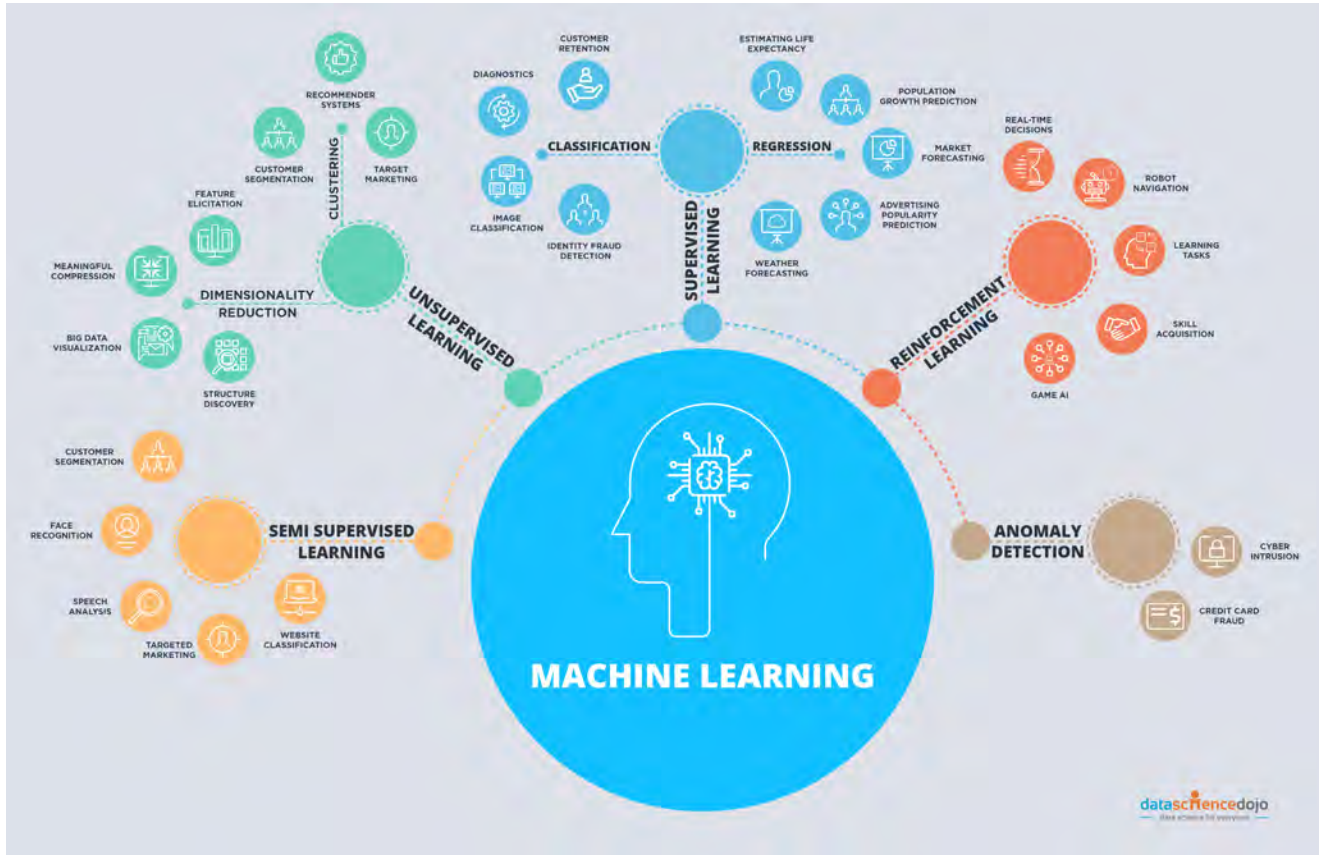
ML predicted glass forming ability (GFA) of tenary metallic glasses based on X-ray data

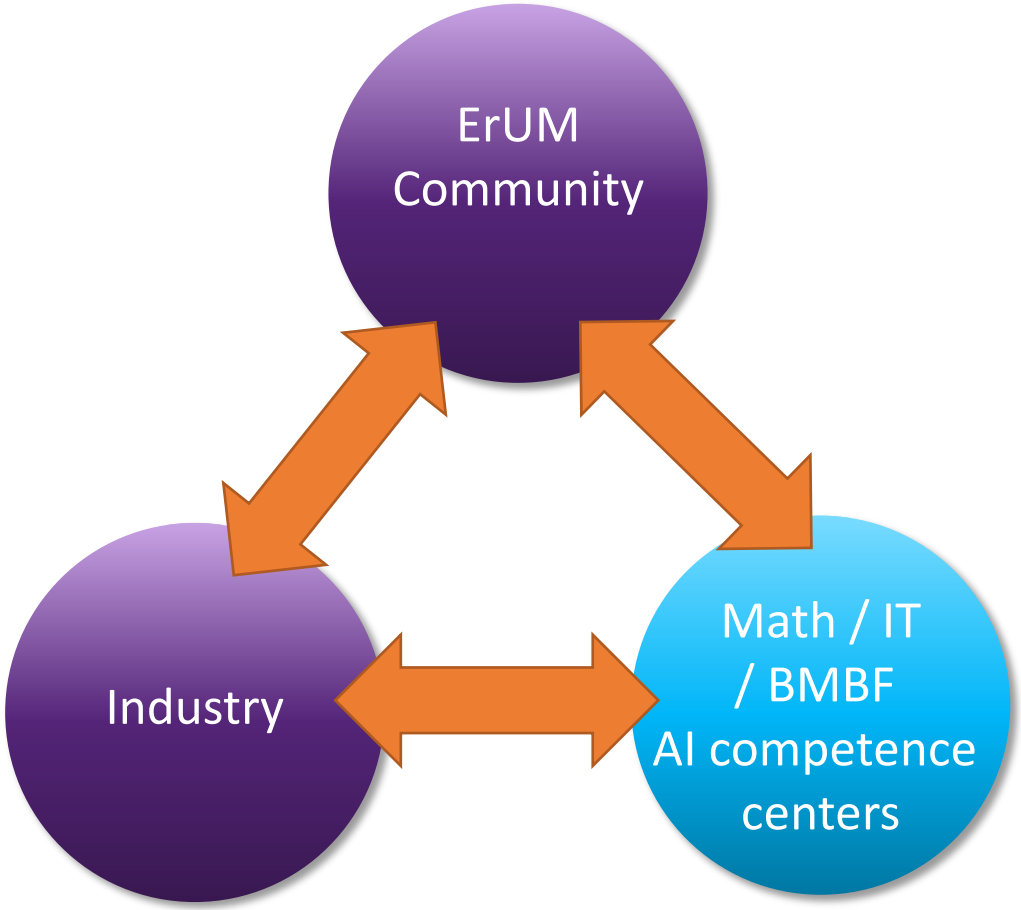


Requires data research management

- complete metadata
- electronic logbooks
- curated and documented analysis software
- proper repositories
- ontologies
- descriptors for material properties. (PRL 114, 105503 (2015)

4. Benchmarking, development, use and understanding of ML algorithms

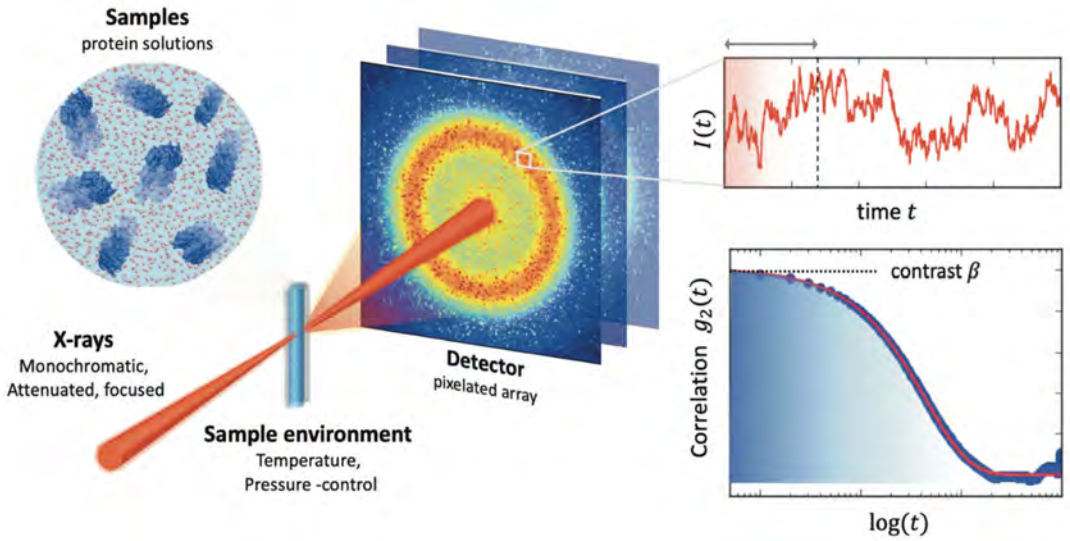




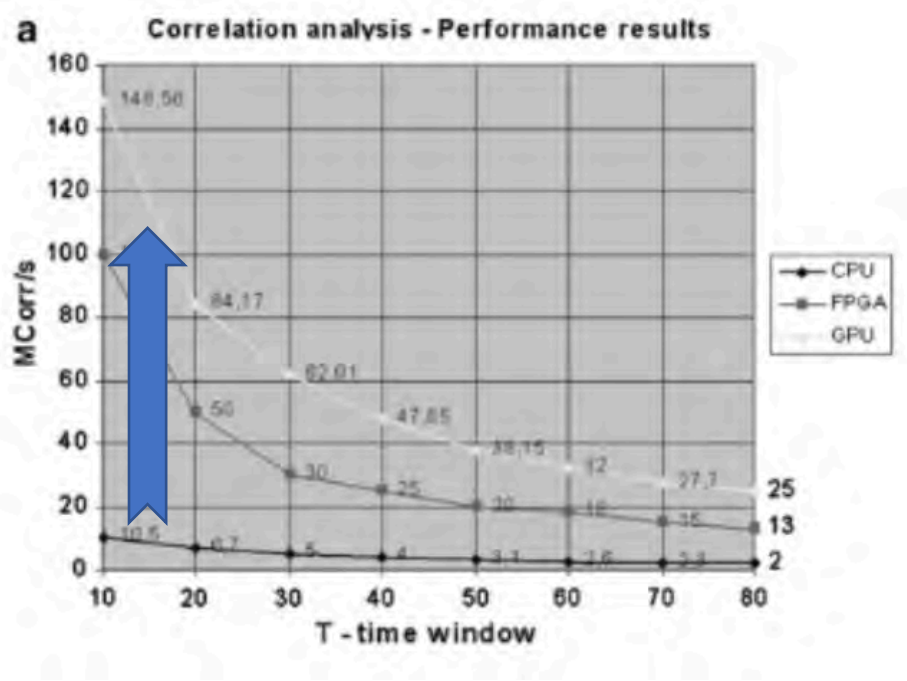
5. Parallelization, GPU + FPGA computing

Example: X-ray photon correlation spectroscopy.

Pixel by pixel correlation $g_2(\vec{Q},t) = \frac{\langle I(\vec{Q},t')I(\vec{Q},t'+t)\rangle}{\langle I(\vec{Q},t')^2\rangle}$



Benefits from parallel computing



Priorities for KFS

- We are striving for **science- and user-driven** software solutions for our community.
- Close collaboration with the national facilities including European XFEL and ESRF
- Evaluate cross community efforts e.g. together with KFN, KFB, KFSI
- Include ML/AI stakeholders and other ErUM communities
- Understanding ErUM Data as a long-term project for the KFS community aiming for establishing new algorithms, new data management, federated infrastructure and data science education.
- **Today and tomorrow: KFS encourages you to discuss and plan possible collaborations for ErUM Data.**