# Towards Efficient Open Source File Formats for the Atom Probe Tomography Experimentalist and the Full-Field Mesoscopic Scale Microstructure Evolution Modeling Communities

MAX-PLANCK-INSTITUT FÜR EISENFORSCHUNG

Markus Kühbach

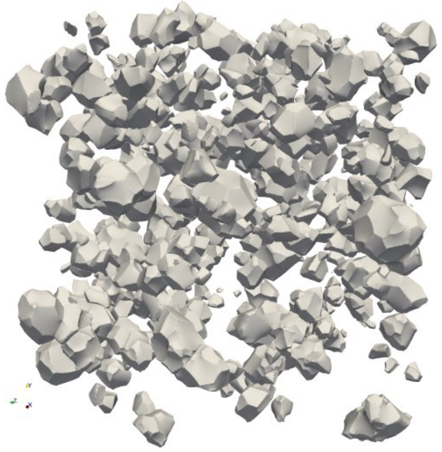
m.kuehbach@mpie.de

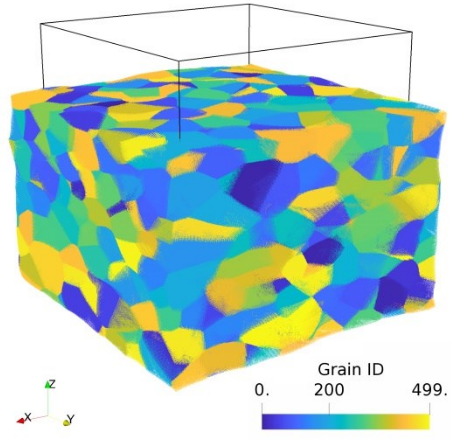
Friday, July 12, 2019

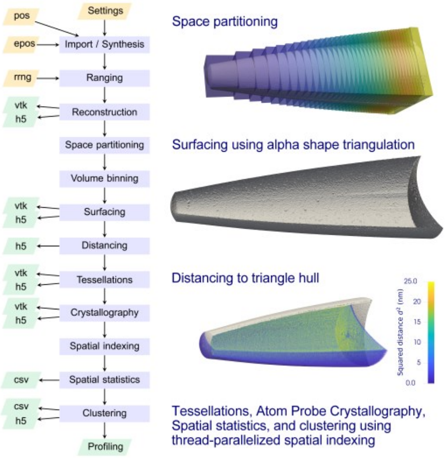
# Personal Background and Motivation

## Myself in a nutshell:

- Dr-Ing. in Metallurgical and Materials Engineering
- Development of parallelized software tools for full-field microstructure evolution modeling and atom probe tomography
- BiGmax software engineering consultant






## Our motivation:

*data must be connected to established metadata and to workflows of their production. all sub-communities need to be brought together*

*A not fully solved challenge is the definition of the sample materials. Obviously, closely coupled to the definition of metadata is the description of workflows in the sample preparation and running of the experiment"* [Landing page statement of this workshop]

## This talk:

☞Metadata and file format situation for atom probe tomography experiments

☞Motivate that the microstructure evolution modeling community is another well-suited candidate to have on the FAIRmat radar in the future

# The BiGmax Network

https://www.bigmax.mpg.de/projects

CompMat | NPJCompMat | IMMI | MetTransA | MatChar | MatDesign | MRSBull | PhilMagA | Elsevier Editorial Syste...

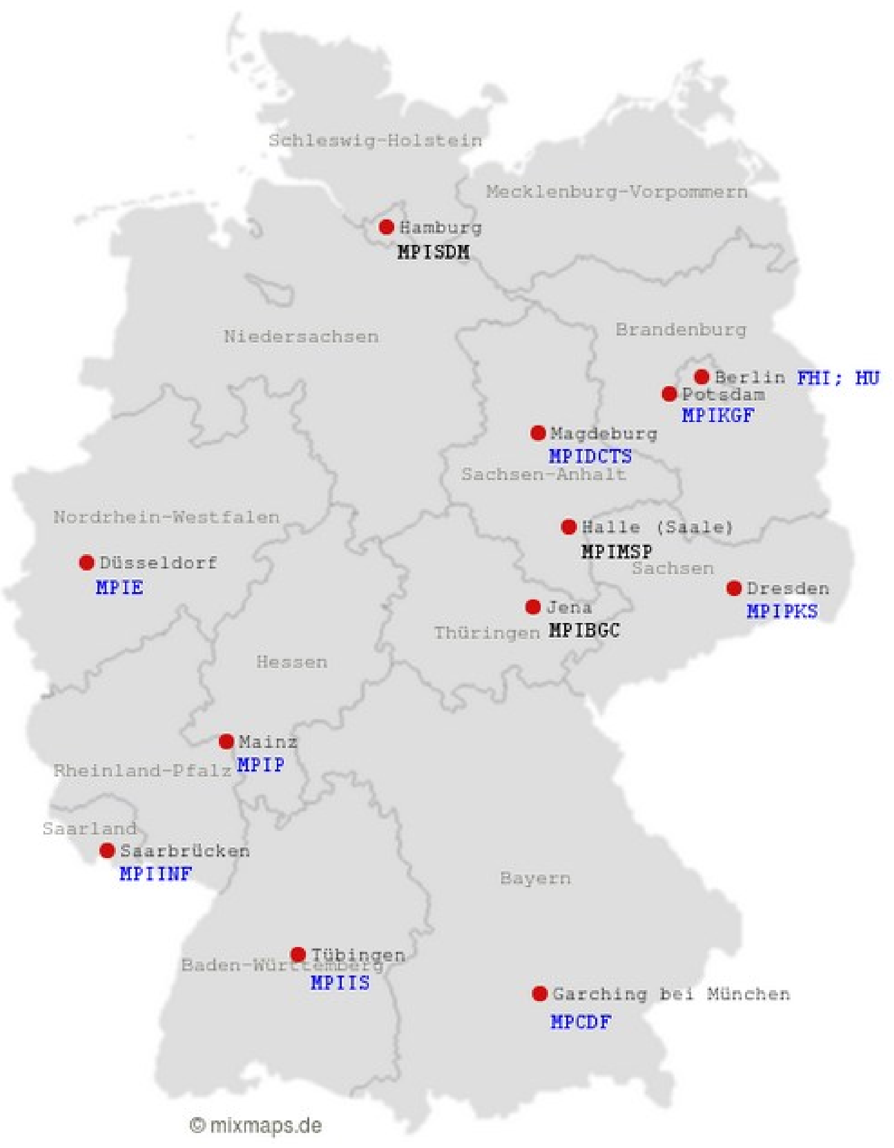
**MAX PLANCK RESEARCH NETWORK**
**on big-data-driven materials science**

ABOUT US | RESEARCH | EVENTS | NEWSROOM | TEAM | CAREER | PUBLICATIONS

Home > project areas

BiGmax is concerned in particular with the following 5 topics:

1. **Structure and plasticity of materials**
2. **Data diagnostics in imaging**
3. **Discovering interpretable patterns, correlations, and causality**
4. **Learning thermodynamic properties of materials**
5. **Materials Encyclopedia (incl. metadata for exp. samples and methods)**




www.bigmax.mpg.de
fairdi.eu

SCIENTIFIC DATA

Comment | OPEN | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier [...] Barend Mons


| | |
|---|---|
| Computational materials science - NOMAD | Experimental materials science |
| Pillar A | Pillar B |

## What is FAIR ?

Research data management paradigm

- Findable
- Accessible
- Interoperable
- Reproducible / Repurposable / Recyclable

## What is FAIR-DI e.V. ?

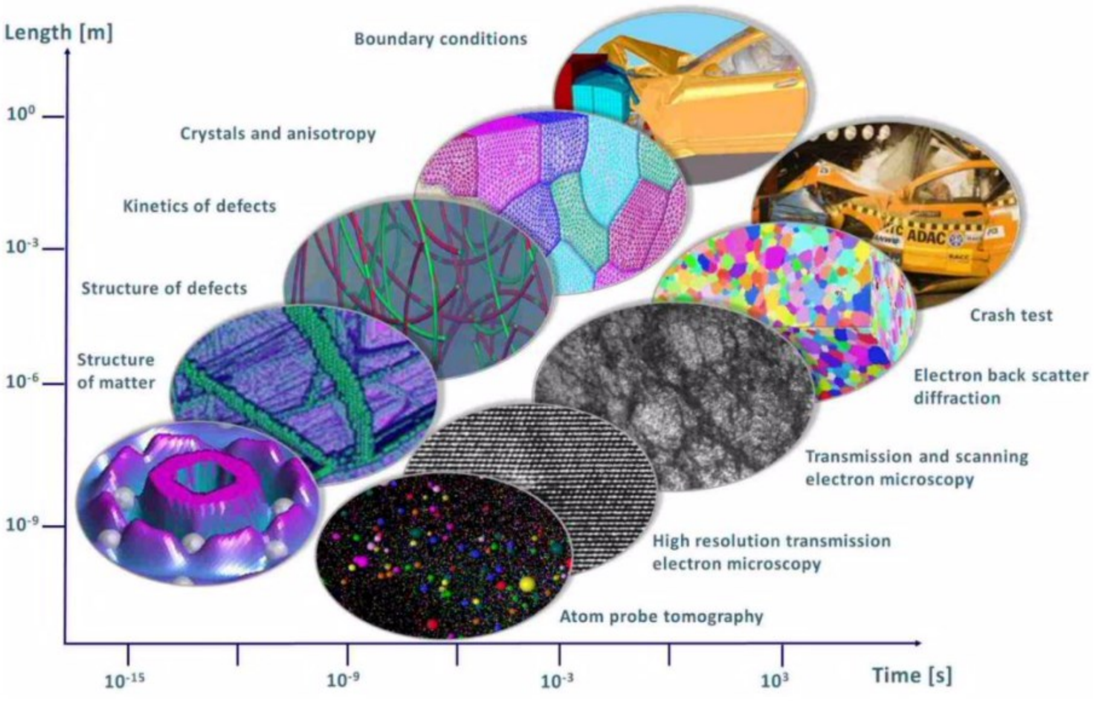
An association to promote FAIR materials science

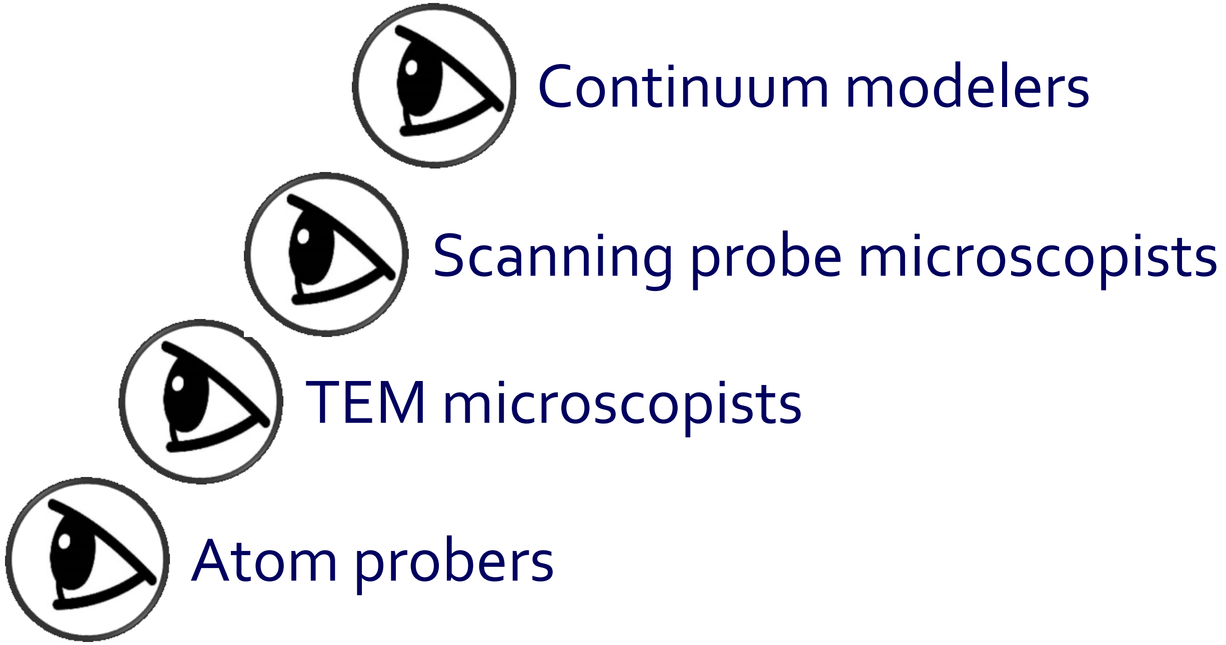
## What is FAIRmat ?

FAIR-DI e.V. lead initiative to write a proposal to DFG's Nationale Forschungsdateninfrastruktur (NFDI) call


M. D. Wilkinson et al., Scientific Data, 3, 2016, 1
C. Draxl and M. Scheffler, MRS Bulletin, 43, 2018, 676
http://www.go-fair.org/fair-principles/
http://fairdi.eu
http://www.dfg.de/foerderung/programme/nfdi/index.html

# Big Picture: Getting Engineering Communities Exp/Sim FAIR




- Scale-bridging experiments motivate cross-community research (exp + sim)
- Mental barriers „more reward for research results rather than tool development"
- Strong permeation of proprietary/closed source software tools many of which lock data in

☞ Perception persists that documenting experiments FAIR is very challenging

☞ Zoo of data and file formats plus strong differences in mindset, technical knowledge, and sophistication how acquisition and analyses pipelines to experiments are documented

http://www.dierk-raabe.com/multiscale-modeling

# Motivating the Examples from Engineering Communities

I want to give two examples with the following motivation:

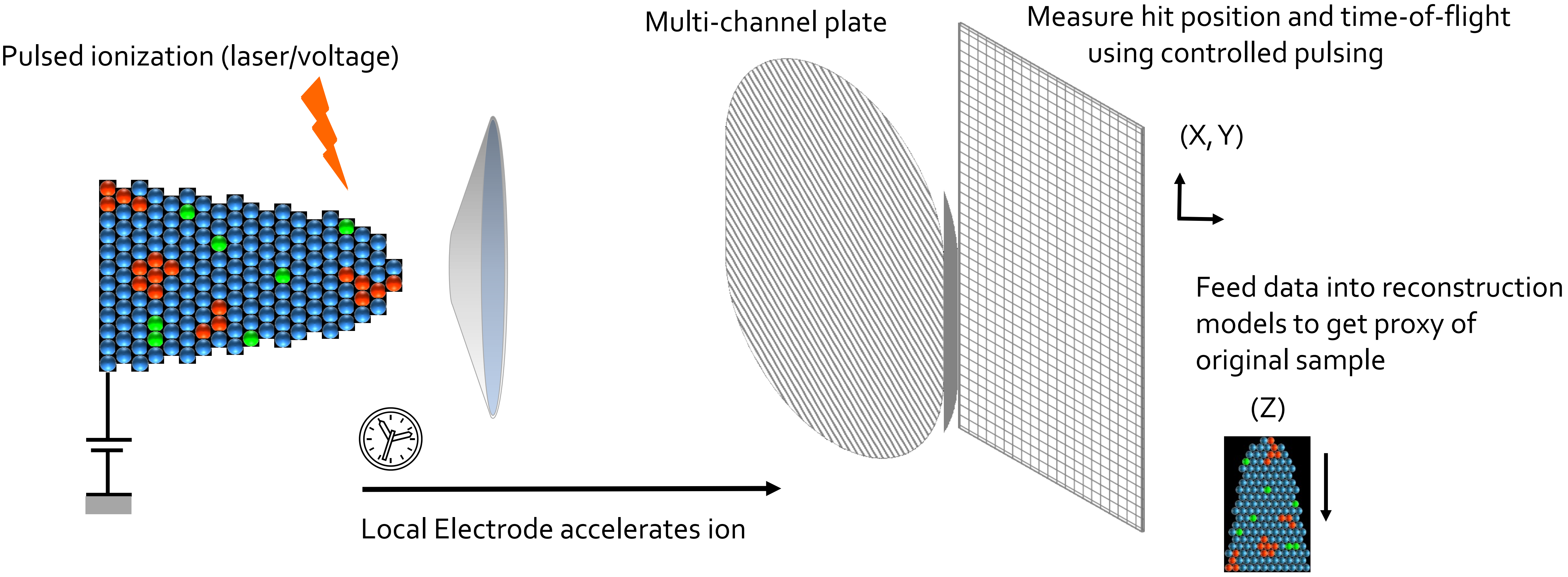
- Identify status quo of metadata and file format activities in APT/FIM and microstructure modeling communities which pinpoint that

- Planned FAIR-DI e.V. activities in pillar A and B fall on extremely fertile ground if FAIR open source tools were to be developed

- Make aware of specific activities in the continuum microstructure characterization and modeling community which are difficult to tell apart from what FAIR-DI e.V. / FAIRmat aims

# Case 1: Atom Probe Tomography Experiments

- Nano-meter-sized site-specific material specimen
- Ionization of sharp needle-shaped specimen, ≈ 50 nm radius
- Ultra high vacuum, cold (20-100K)
- Destructive



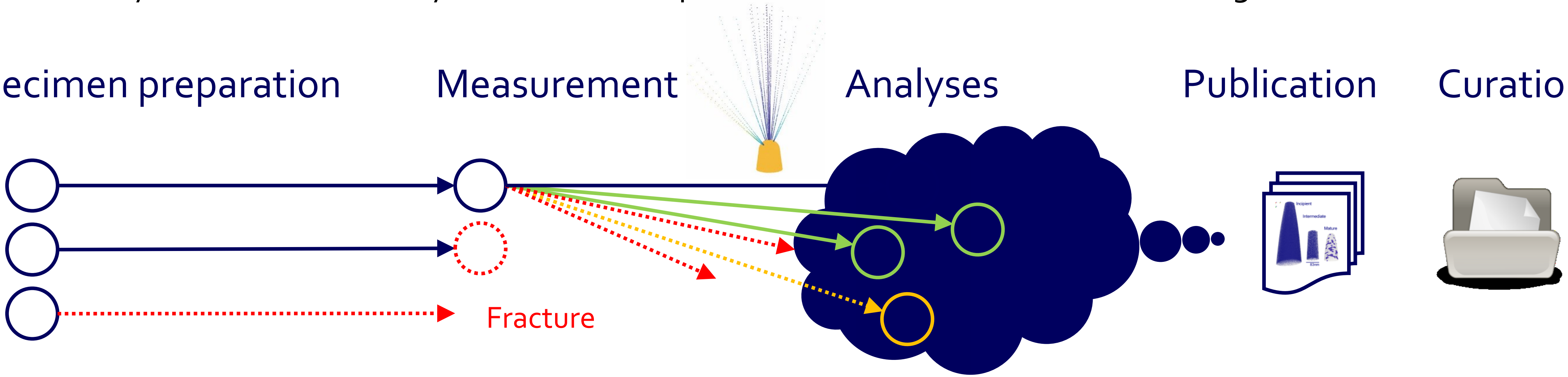

Credit to Manoj Naikade

# APT/FIM Experiments: Their Right to Exist and Workflow

## Why to do APT/FIM experiments?

- Characterize and correlate chemical composition with local crystal defects and microstructure
- Possibility to measure as many as a billion ions per measurement delivers statistical significance

Specimen preparation — Measurement — Analyses — Publication — Curation



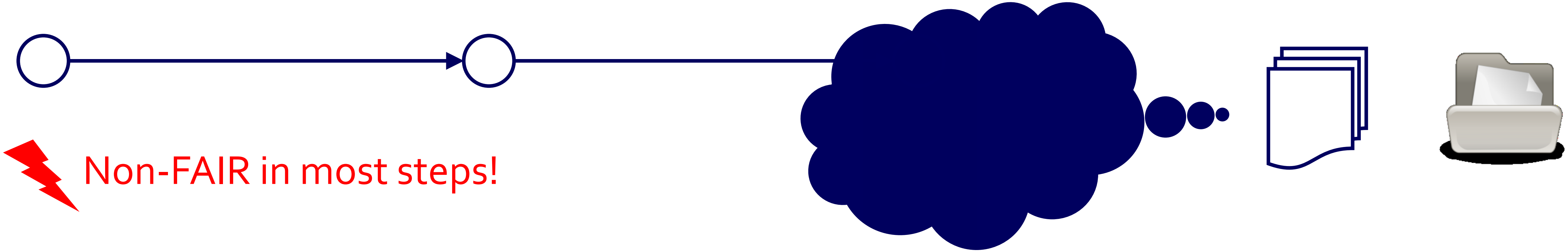

- Thermo-chemo-mechanical history
- Metallography steps
- Electrochemoical / ion milling steps
- Environment of devices between which specimen is transferred

- Tomograph environment
- Tomograph conditions
- Experiment settings
- Measure ion hit positions
- Measure time-of-flight

- Case specific/operator analysis tasks and workflows
- Proprietary software + user scripts
- Ion type identification based on time-of-flight
- Reconstruction models to get atomic positions
- Descriptive spatial statistics and computational geometry on atom positions to extract features
- Distill results in light of research questions

# Status Quo on Metadata in APT/FIM Experiments

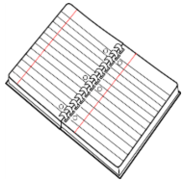
Specimen preparation | Measurement | Analyses | Publication | Curation



Non-FAIR in most steps!

**Specimen preparation**

- Thermo-chemo-mechanical history
  Third-party infos + manual lab book
- Metallography steps
  Manual lab book
- Electrochemical or ion milling
  Manual lab book
- Transfer environment
  $T, p, p_{O_2/H_2}, t$



**Measurement**

- Tomograph environment
  Electronic database
- Tomograph conditions
  E-database + closed vendor
- Experiment settings
  E-database + manual lab book
- Measure $x_{det}, y_{det}, t_{flight}$
  Closed non-public manufacturer file
  Neither file layout nor fields completely documented, transcoding required

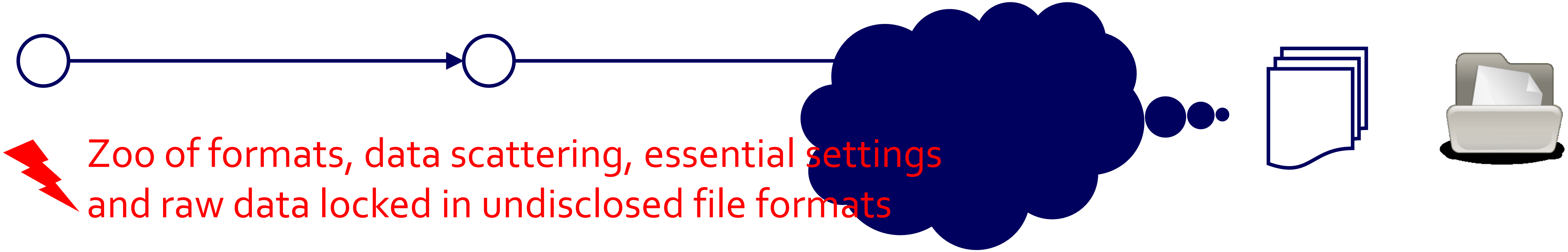
**Analyses**

- Case specific analysis tasks and pipeline
  Explorative, electronic + manual lab book
- Proprietary software + user scripts
  Electronic + manual lab book
  Ad hoc shell scripting

**Publication / Curation**

Paper + researchers directory archive
Ad hoc selective uploading to online repositories is on the rise

Specimen preparation | Measurement | Analyses | Publication | Curation



Zoo of formats, data scattering, essential settings and raw data locked in undisclosed file formats

- Thermo-chemo-mechanical history

Vendor specific ASCII machine protocols and SQL databases

- Metallography steps
- Electrochemical or ion milling
- Transfer environment

- Tomograph environment
- Tomograph conditions

  Closed binary RHIT/HITS
- Experiment settings

  Closed binary RHIT/HITS
- Measure $x_{det}, y_{det}, t_{flight}$

  Transcoding selected content from RHIT/HITS to open binary POS/EPOS

  No hashing mechanisms commonly place

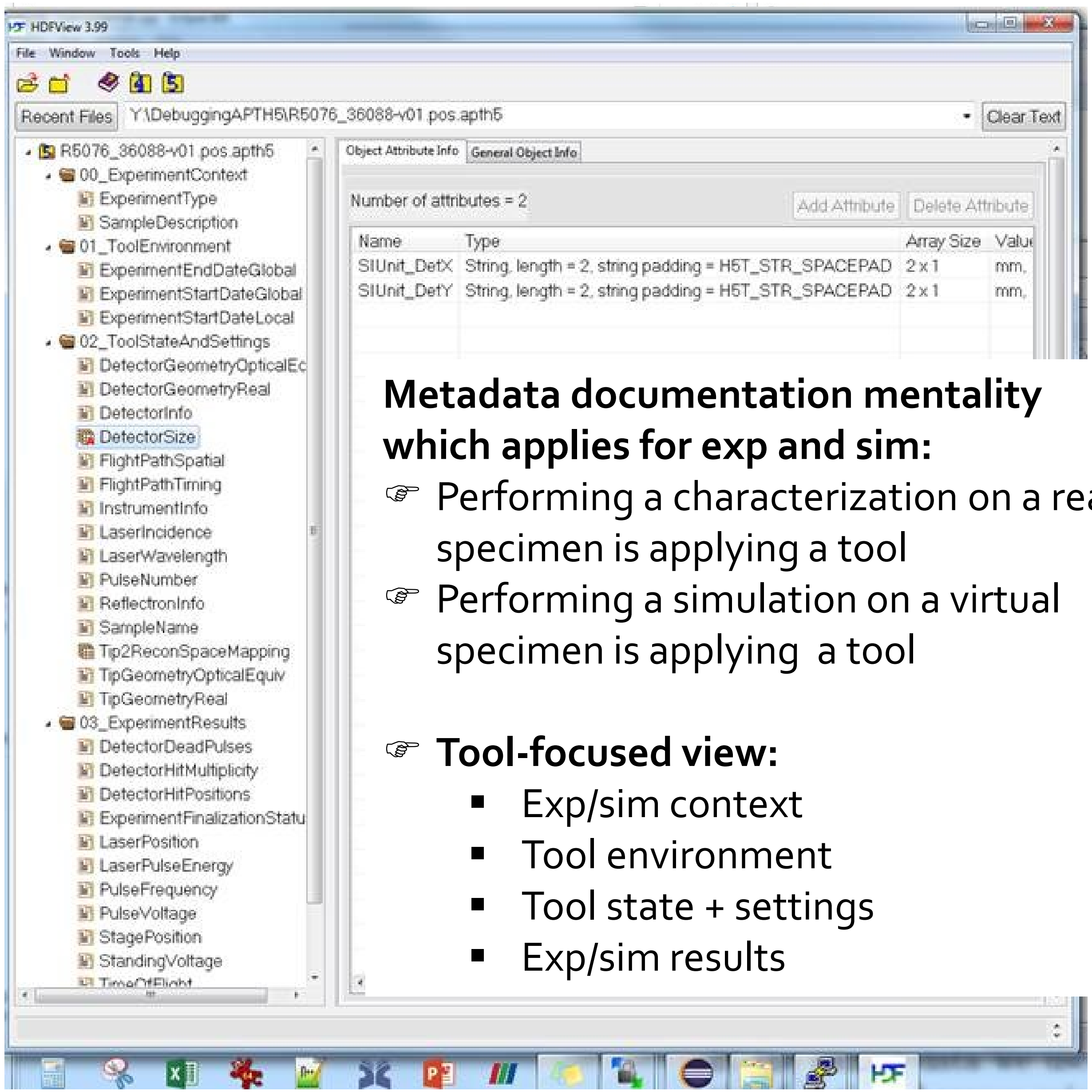
- Case specific analysis tasks and pipeline

  RHIT/HITS/POS/EPOS/ASCII
- Proprietary software + user scripts

Word/TEX

RHIT/HITS/POS/EPOS/ASCII/TAR/ZIP

# Towards FAIR Metadata in APT/FIM Experiments




**Metadata documentation mentality which applies for exp and sim:**

- Performing a characterization on a real specimen is applying a tool
- Performing a simulation on a virtual specimen is applying a tool

- **Tool-focused view:**
  - Exp/sim context
  - Tool environment
  - Tool state + settings
  - Exp/sim results

- 201x - ongoing
  Efforts by individuals (especially B. Gault) to get RHIT content opened up through manufacturer
- 2017-2018
  APT/FIM community initiative to form IFES TC
  „Initiate community-wide efforts what constitutes APT metadata and how to open up closed formats“
- Q1/2019
  IFES TC drafted metadata specification for acquisition
  Agreement to move to HDF5 as file format
- Me implementing first HDF5 proof-of-concept for acquisition and post-processing side, now in contact with manufacturer about opening up file format
- Q3-Q4/2019
  Manufacturer about to roll out complementary open binary file format, unfortunately custom brew but company signalized that HDF5 „is a good choice“

# APT/FIM Experiment Curation through FAIRmat

## Proof-of-concept trials to use Zenodo + NOMAD to archive APT/FIM experiments

Files (7.3 GB)

**Name**

PARAPROBE.SimID.20231.STDERR.txt

md5:a590d8e02a0d204b86c5ace82ae61516

PARAPROBE.SimID.20231.STDOUT.txt

md5:e3e307bfd6d596908407fc19f55f3459

PARAPROBE.Transcode.EPOS2APTH5.R76_20

md5:fb0de7efd45565ec3a3cd2e13429069a

preview.png

md5:3e4910b43e4263202aa0caaad4cc415f

R76-20231-v01.aptfim

md5:55106627b93b2159416c07039dc41de2

R76_20231-v01.epos.apth5

md5:e86dd64c1789e6e65bf797f376a40f67

R76_20231-v01.epos.tar.gz

md5:da005298e322cb07425f2c76cbed5b7e

C:\Users\m.kuehbach\AppData\Local\Temp\R76-20231-v01.aptfim - Notepad++

File Edit Search View Encoding Language Settings Macro Run Plugins Window ?

CONFIG_Reconstruction.h | CONFIG_Reconstruction.cpp | new 1 | R76-20231-v01.aptfim


```
{
    "data_repository_name": "zenodo.org",
    "data_repository_url": "http://zenodo.org/record/2669484",
    "data_preview_url": "http://zenodo.org/record/2669484/preview.png",

    "experiment_method": "atom probe tomography or field ion microscopy",
    "experiment_location": "Düsseldorf, Germany",
    "experiment_facility_institution": "Max-Planck-Institut für Eisenforschung GmbH",
    "instrument_info": "LEAP R76-20231-v01",

    "experiment_date_global_start": "00.00.0000 00:00:00",
    "experiment_date_global_end": "00.00.0000 00:00:00",
    "experiment_data_local_start": "00.00.0000 00:00:00",
    "experiment_operation_method": "apt",
    "experiment_imaging_method": "laser",

    "specimen_description": "WC nano grains with Co, Cr, Fe, V segregated at the grain boundaries bulk sample polycrystalline",
    "specimen_chemistry": ["W", "C", "Co", "V", "Cr", "Fe"],
    "specimen_microstructure": "poly crystal",
    "specimen_constitution": "single phase",

    "measured_nions_evaporated": 127459578,
    "measured_detector_hit_pos": "yes",
    "measured_detector_hit_mult": "yes",
    "measured_detector_dead_pulses": "yes",
    "measured_time_of_flight": "yes",
    "measured_standing_voltage": "yes",
    "measured_pulse_voltage": "yes"
}
```

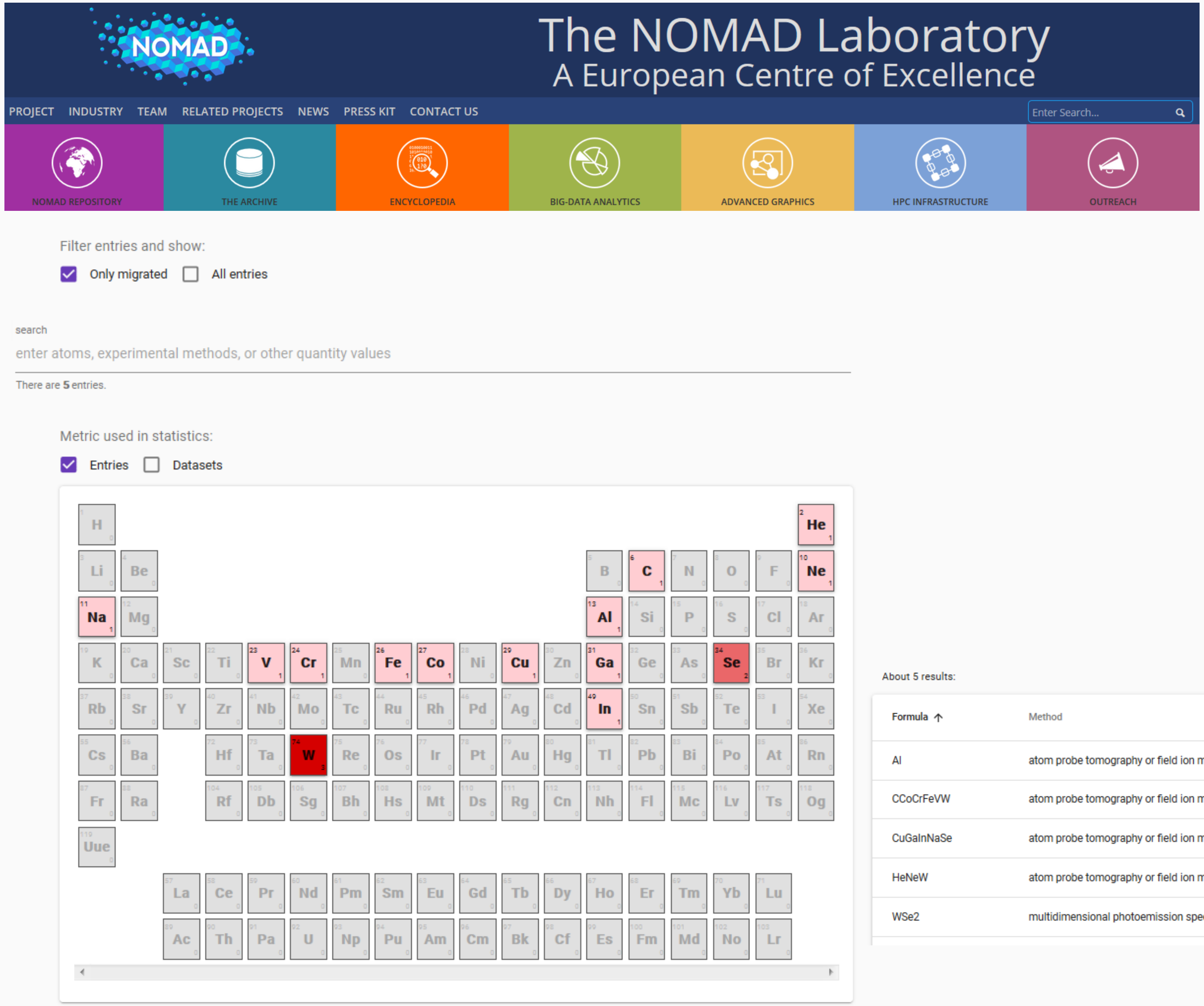

zenodo Search Upload Communities

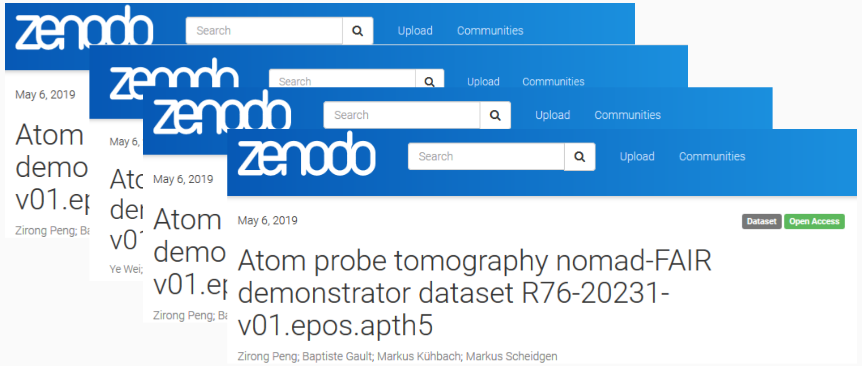
May 6, 2019 Dataset Open Access

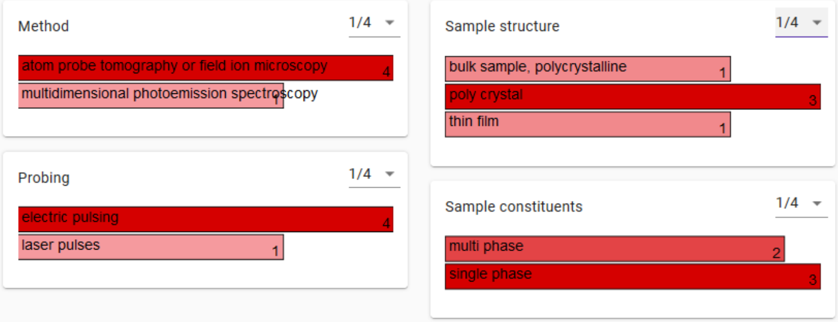
Atom probe tomography nomad-FAIR demonstrator dataset R76-20231-v01.epos.apth5

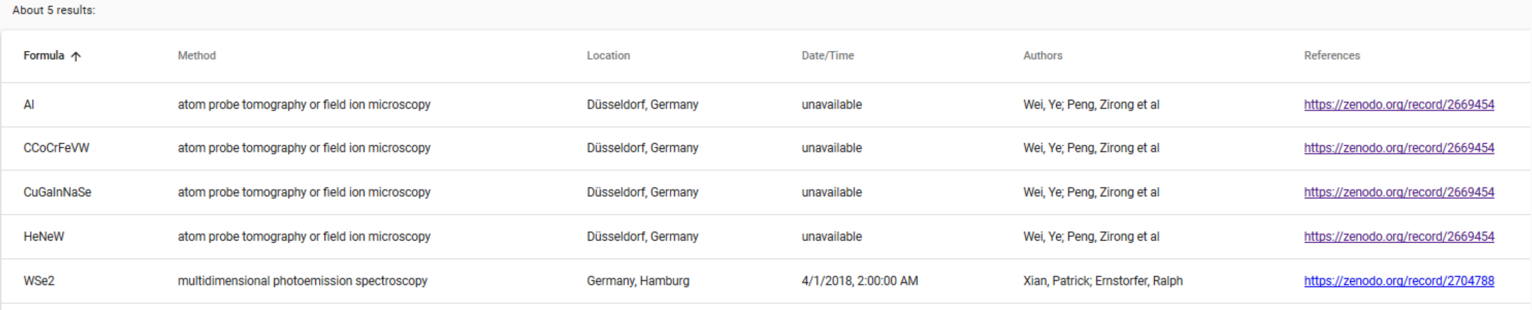
Zirong Peng; Baptiste Gault; Markus Kühbach; Markus Scheidgen

## Proof-of-concept trials to use Zenodo + NOMAD to archive APT/FIM experiments




Credit to Markus Scheidgen here
https://labdev-nomad.esc.rzg.mpg.de/fairdi/nomad/experiments/gui/search

# Take Home 1: APT/FIM Experiments and FAIR

☞ APT and FIM are state-of-the-art microscopy techniques for probing quantitatively the 3d architecture of many structural materials at the atomic scale.

☞ Metadata to APT/FIM experiments + analysis pipeline have so far neither been consistently defined nor used homogeneously in the community, also given the monopoly of the manufacturers analysis software and ubiquitous use of manual documentation procedures. Files use a zoo of formats with essential experimental settings locked in undisclosed layouts.

☞ The APT/FIM community has understood the necessity to improve on above FAIR barriers. Key individuals have been convinced to implement FAIR compliant tools: more complete metadata collecting on the acquisition/measurement side, use of more efficient, open, and collating file formats. However, metadata for the post-processing side remain a challenge:

☞ Virtually all APT/FIM analysis tools are either undisclosed source or scattered across labs, hence calling for FAIRmat activities: data consolidation, doing a lot of persuading, and professionalized software development.

# Case 2: Characterizing Physics of Evolving Microstructures

**Deformation** **Annealing** **Terminology**

Recovery Recrystallization Grain growth

Dislocation and point defect storage **Mechanisms**

Alteration of microchemistry

Dislocation annihilation

Modification of dislocation walls + sub-grain boundaries

Nucleation

Grain boundary migration

- Frequently overlapping static and dynamic phenomena

**Driving forces**

Free enthalpy of dislocations

Free enthalpy of grain boundaries

Solute atoms in thermodynamic interaction with crystal defects modify rates and mechanisms

A. Rollett, G. S. Rohrer and J. Humphreys, Recrystallization and Related Annealing Phenomena, 3rd, 2017

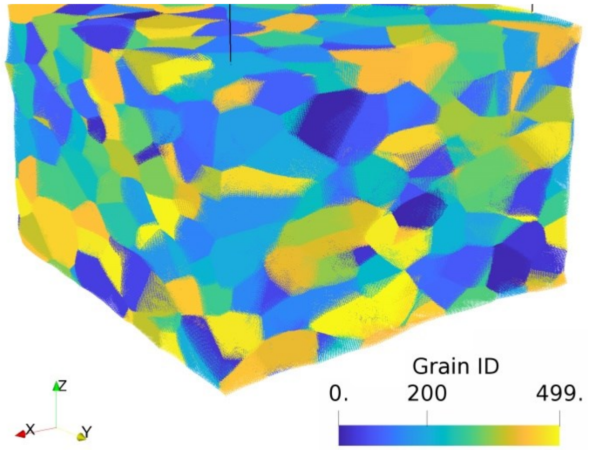
## Key motivation:

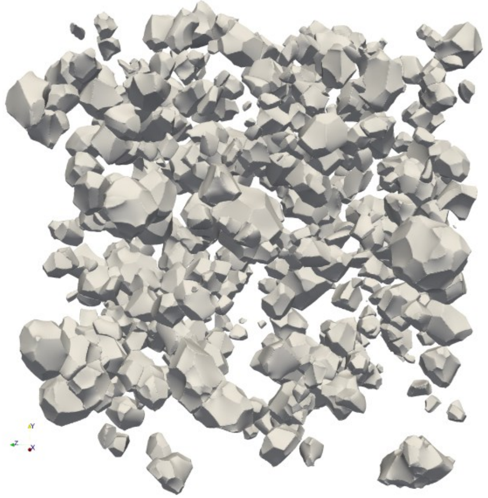
- Quantify state variable values (dislocation density, stresses/strains) through exp/sim to feed into mean-field/continuum models to predict material properties
- Cost-efficient alloy development, understand ensemble dynamics

## Key assumptions:

- Parameterized (ab-initio-guided) crystal defect properties, statistical representative volume element (RVE)
- Continuum-scale models: 3d geometry of crystal defects + chemistry:
- Solute concentration fields, dislocations, grain and phase boundaries, multi-phases
- Few physical mechanisms dominate kinetics, thermodynamically-driven evolution

$$\dot{\gamma} = \varrho b v_0 \exp\left(-\frac{Q}{kT}\left(1-\frac{|\tau_{eff}|^p}{\hat{\tau}}\right)^q\right) sgn(\tau_{eff})$$



$$v_{GB} = m\nabla G \coloneqq mp$$

$$v_{GB} \coloneqq m\left[{}^1/_2\, Gb^2\Delta\rho - \gamma\kappa - \alpha\gamma\, {}^f/_r\right]$$

$$m = m_0 exp(-\frac{H}{kT})$$

D. Turnbull, Trans AIME, 1951, 191, 661-665
G. Gottstein et. al., CRC Press, 2. ed, 2010
U. F. Kocks et al., Pergamon Press, 1975
F. Roters et al., Comp. Mat. Sc. 158 (2019), 420-478

# Efficient File Formats + Metadata: Deformation Example

Evidence of typical friction in the microstructure modeling community:

- Move to HDF5 when letting established proof-of-concept tools loose on larger RVEs / more numerous parameter runs does not scale

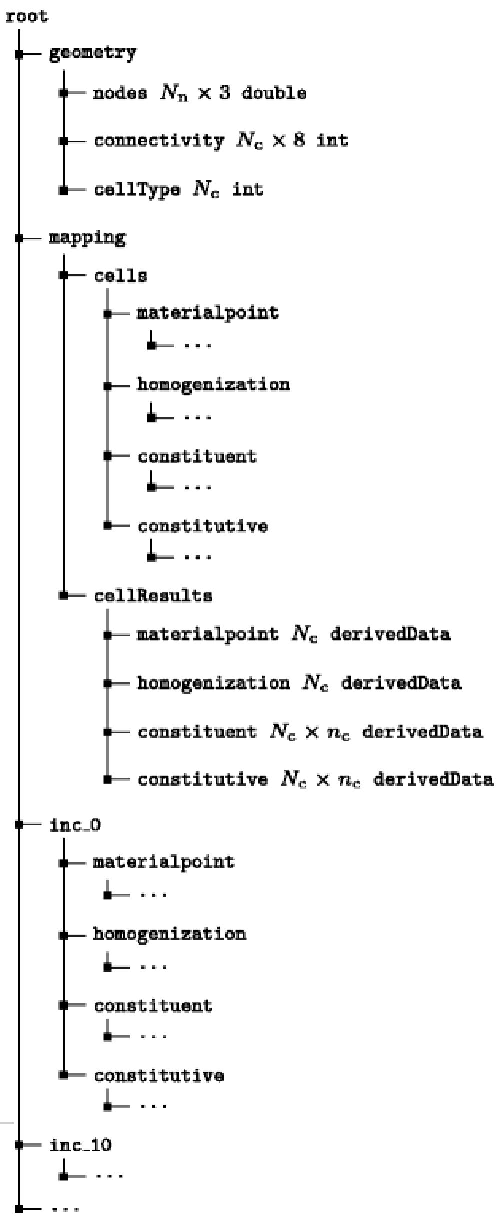
- Often then motivating deeper thougIhs on metadata

Integrating Materials and Manufacturing Innovation
March 2017, Volume 6, Issue 1, pp 83–91 | Cite as

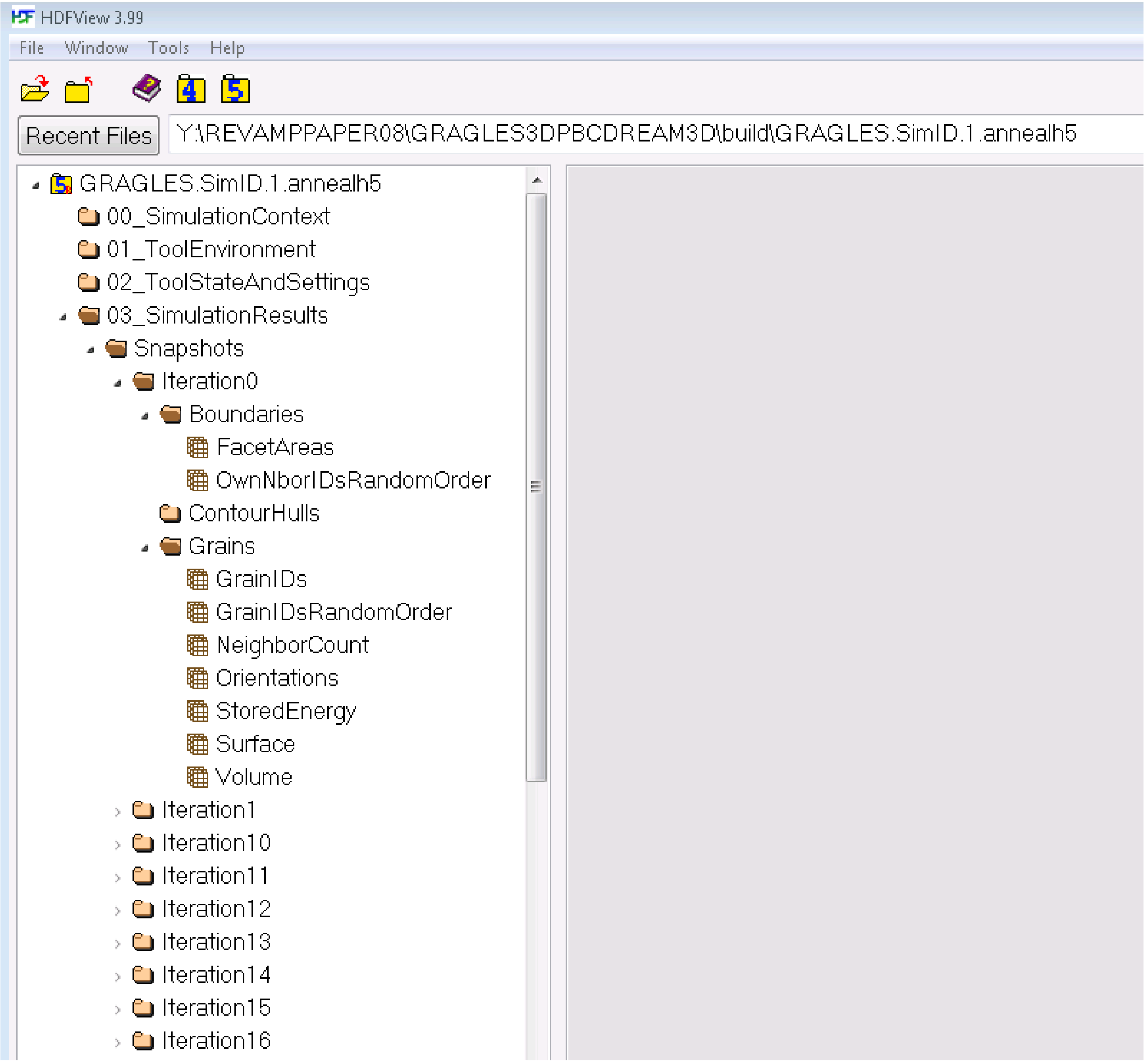
## A Flexible and Efficient Output File Format for Grain-Scale Multiphysics Simulations

Authors | Authors and affiliations

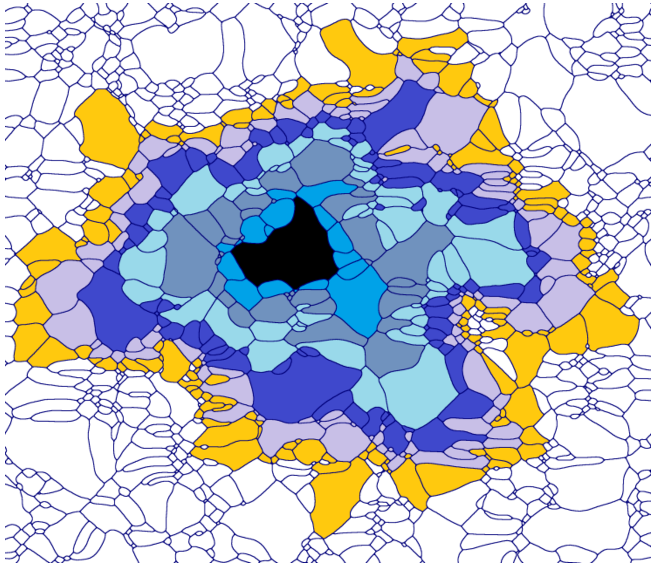
Martin Diehl, Philip Eisenlohr, Chen Zhang, Jennifer Nastola, Pratheek Shanthraj, Franz Roters



# Efficient File Formats + Metadata: Annealing Example



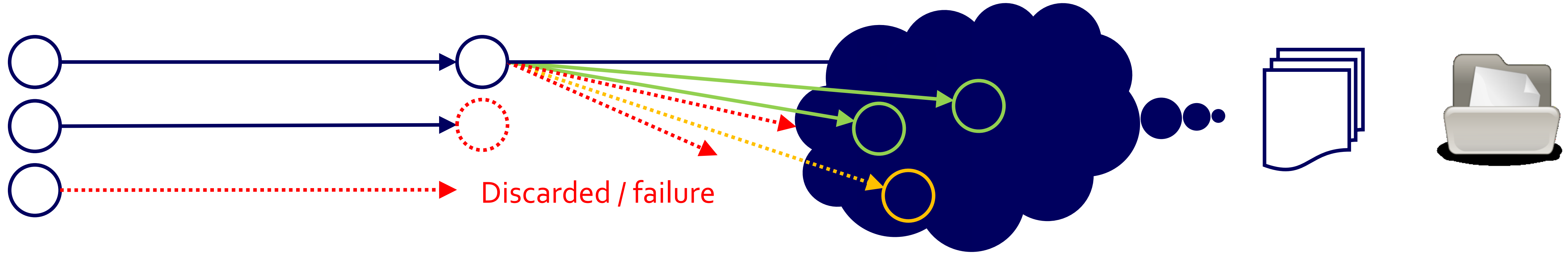

2012-2017
- GraGLeS anisotropic grain growth solver
- Level-set method
- $10^4 - 10^6$ *grains* grains per RVE



2016-2019
- Understand effect of long-range neighbors on abnormal grain growth through monitoring grain boundary network geometry evolution
- Only each 1000 grain survives the coarsening!
- ≈ 5000 triangles x 50000grains x 2000 time steps / 2
- 1-10 TB per simulation + replica studies

# Status Quo on Metadata in Experiment/Simulation

Specimen preparation | Measurement | Analyses | Publication | Curation




**Specimen preparation**
- Thermo-chemo-mechanical history
- Metallography steps
- Electrochemical or ion milling
- Transfer environment

**Measurement**
- Microscope environment
- Microscope conditions
- Experimental settings
- Measure multi-spectral image

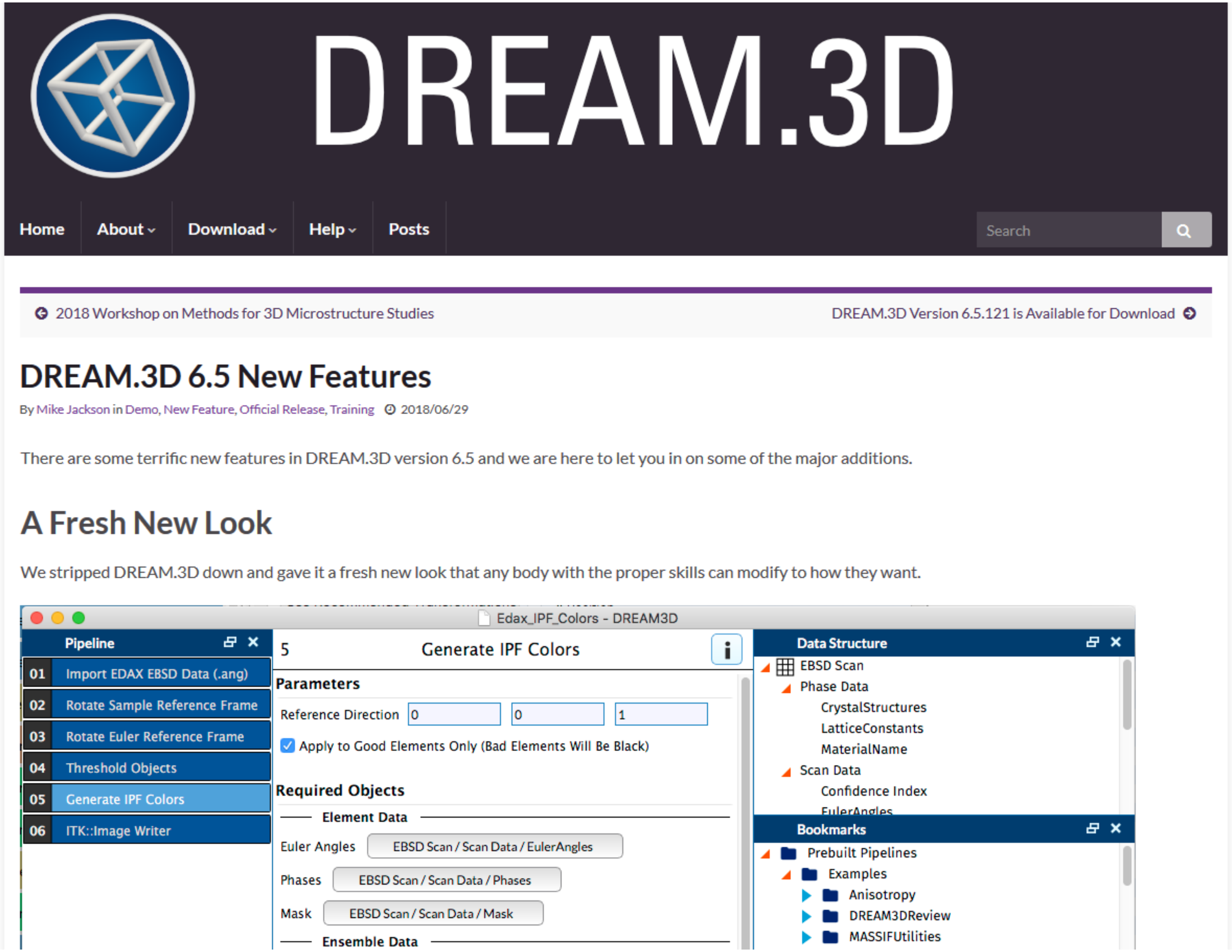
**Analyses**
- Case specific analysis tasks and pipeline
- Proprietary software + user scripts
- Reconstruction models to get crystal defects
- Grain boundary network, microchemistry
- Distill results in light of research questions
- Individual workflows

## RVE synthesis

- Possibly experimental input
- Synthesis method and solver
- Hardware and software used

## Simulation

- Parameterization
- Kinetics/thermodynamics solver
- Hardware and software used
- RVE state variable value field snapshots




## DREAM.3D

- Carnegie Mellon University + U.S. Air Force Materials science division-lead example of
- 3d microstructure synthesis and analysis tool

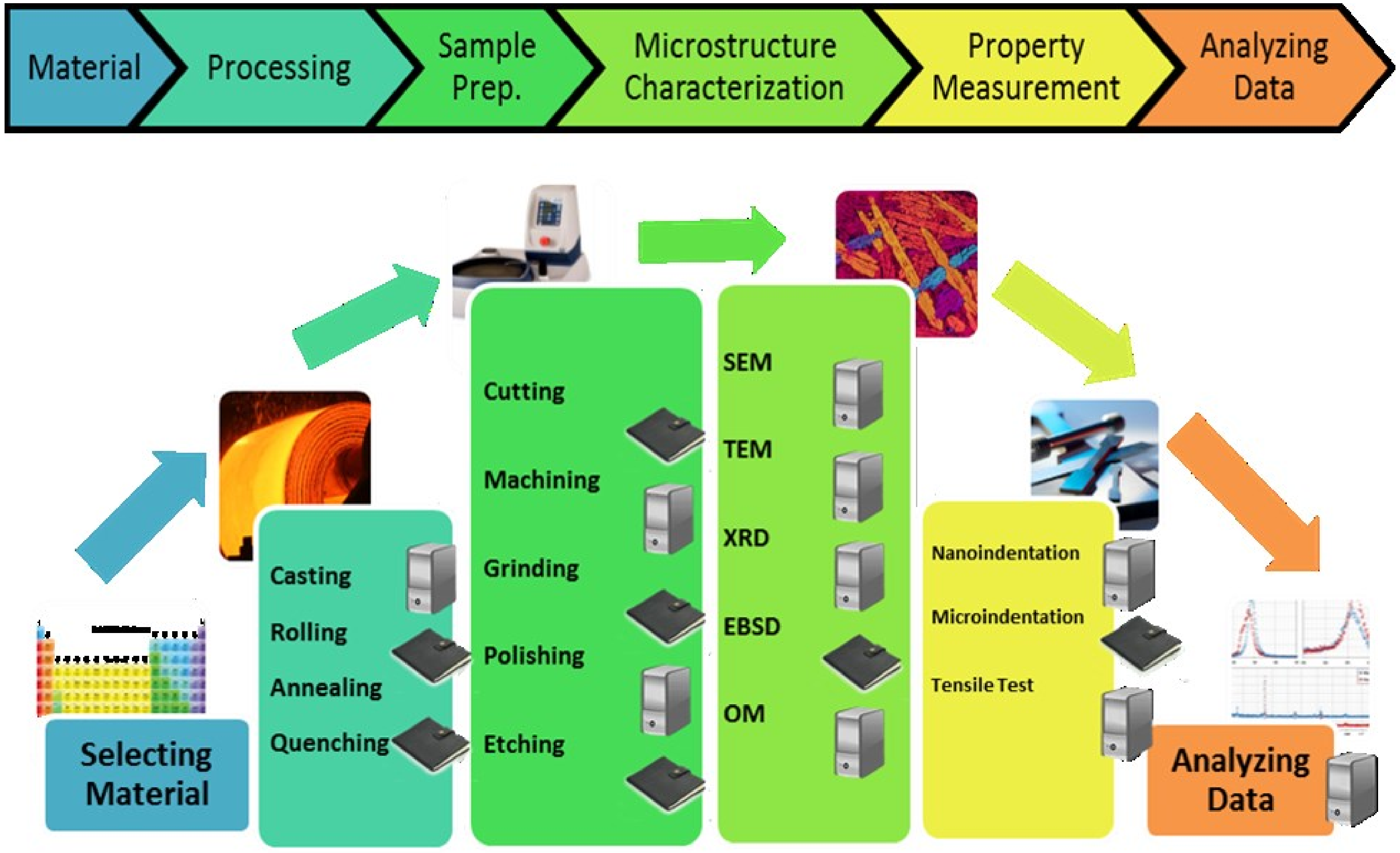
- Pipeline-based workflow
- JSON metadata document pipeline settings for each microstructure realization

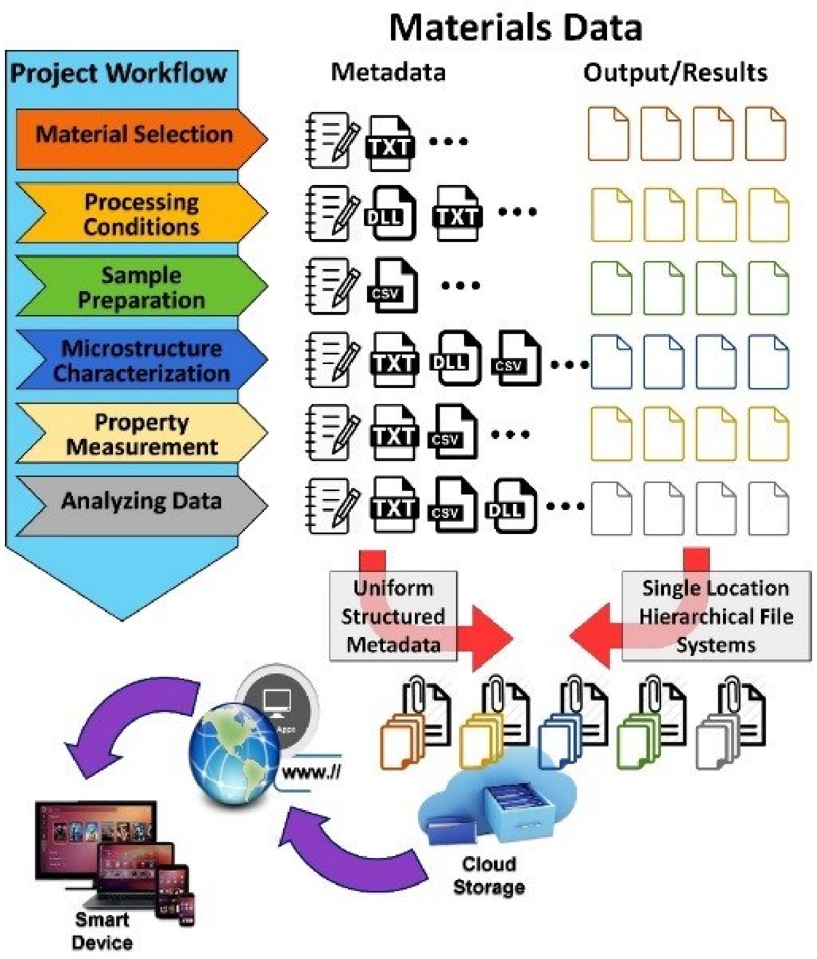
- One well-documenting tool of a pipelines of different mentalities

- DREAM.3D is one accepted and increasingly used tool in ICME
- Motivates community members to interface their tools to

http://dream3d.bluequartz.net

# To not be Wearing Blinkers: MATIN and ICME

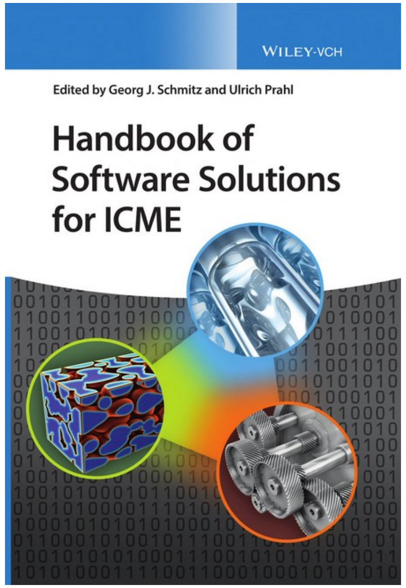



- MATIN – Materials Innovation Network
- Georgia Tech University (S. Kalidindi)-lead initiative
- Develop fully-integrated materials processing and characterization toolbox

https://matin.gatech.edu/groups/mined_super/facilities/data_science_tools1

Microstructure evolution modeling community has seen the train of ICME winding through the literature: ICME is *„linking of individual microstructure evolution models to enable simulations of a processing chain with which to probe (structural) materials using virtual laboratories*"




## Role of experiments for ICME:

- Key ingredient to parameterize models and quality
- Many experiments within the microscopy communities addressed in FAIRmat should be seen in context of the microstructure models the experiments should support

## Status quo consensus on metadata and file formats:

- Wide consensus that HDF5 is method of choice to carry through/manage FAIR open science workflows
- A community moving to customized HDF5
- Not only this workshop showed, this alone is not enough: ontologies? workflow hierarchies? reproducibility?

## Open questions:

- How to convince proprietarizing software developers within the community to open up file formats and provide interfaces for docking open science community developed workflows ?
- How best to support these domain-specific communities instead of telling them just you have to implement all the tools to a FAIR-compliant level yourself?

# Take Home 2: Microstructure Evolution Exp/Sim Using RVEs

☞Cross-scale/correlative diffraction/microscopy experiments which probe physical mechanisms and are used to initialize full-field ensemble models define the status quo.

☞Zoo of solver and implementations whose FAIR state is similar like it was in the ab-initio community prior to FAIR activities shown in this workshop.

☞Despite considerable ICME activities, there is still no consensus nor completed tool modifications and hence most simulations and experiments get still documented non-FAIR.

☞What is necessary to achieve stronger permeation of FAIR principles in these communities ?

- Tools which automatize documenting of exp/sim and connect/integrate these in ICME pipelines
- Opening of proprietary software algorithms and exporting to open source file formats (like in APT case)
- Necessity for further success stories to convince community that efforts in moving to FAIR is worth it
- Necessity for a cultural change from monolithic software solutions to flexible pipeline-based approaches with rigorously implemented auto-documentation functionality
- Stronger communication of technical side + benefits of using open source data repositories across groups
- Software engineering, programming, data science technique training in graduate and PhD curricula

# Thanks for your attention!

## Questions ?

☞ Development of parallelized software research tools
☞ Strategies for research data volume handling
☞ Metadata management

☞ Email m.kuehbach@mpie.de
☞ Skype kaiobach
☞ Phone +49-211-6792-385

Dr.-Ing. Markus Kühbach
BiGmax software engineering consultant
Research scientist in the group MA/Theory and Simulation
Max-Planck-Institut für Eisenforschung GmbH
Max-Planck-Str. 1
D-40237 Duesseldorf
Room 649

YEARS 1917–2017

MAX-PLANCK-INSTITUT FÜR EISENFORSCHUNG