Pacific Northwest
NATIONAL LABORATORY

MCPC
Materials Characterization, Prediction, and Control @PNNL

Broadening the Definition of Data Sharing: Infrastructure and Cultural Needs for Success

February 23, 2024

Erin I Barker
Jenna Pope, Julian Escobar,
Julia Nguyen, and Mohammad Taufique

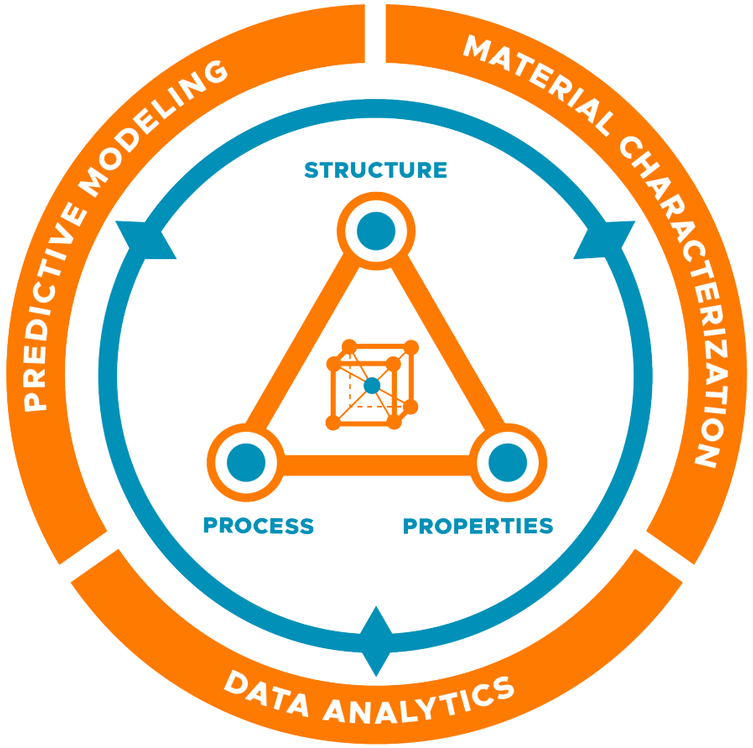
PNNL-SA-191482

U.S. DEPARTMENT OF ENERGY BATTELLE

PNNL is operated by Battelle for the U.S. Department of Energy

Accelerating the Development and Characterization of Advanced Material Processing

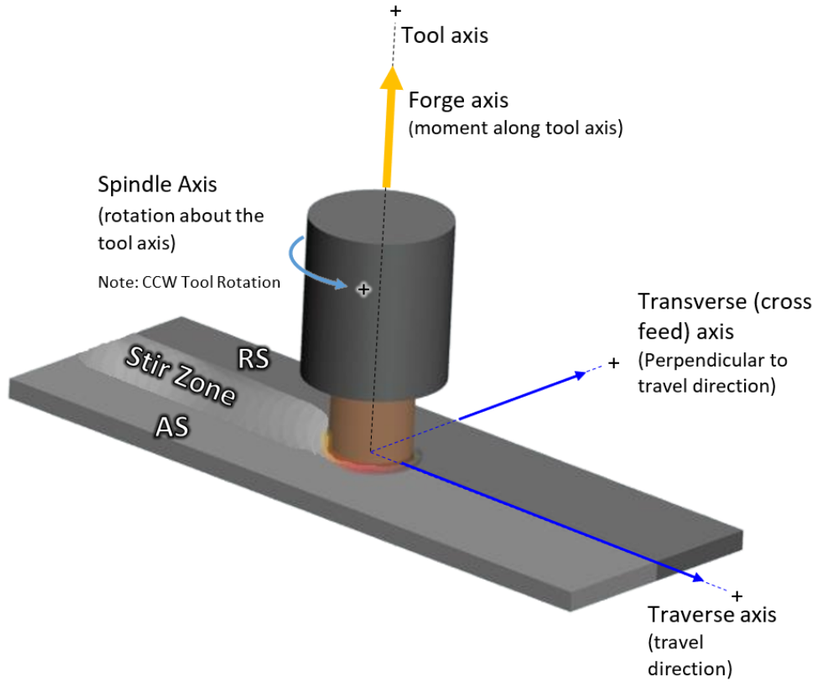
Science challenge: Predict and control process-structure-property relationships to support accelerated development and qualification of material processing systems in national security and nuclear energy.

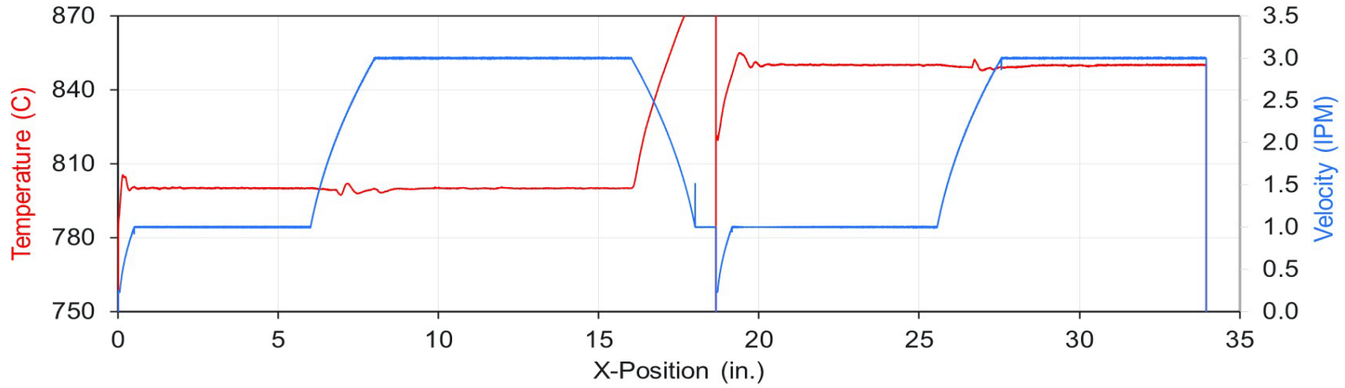


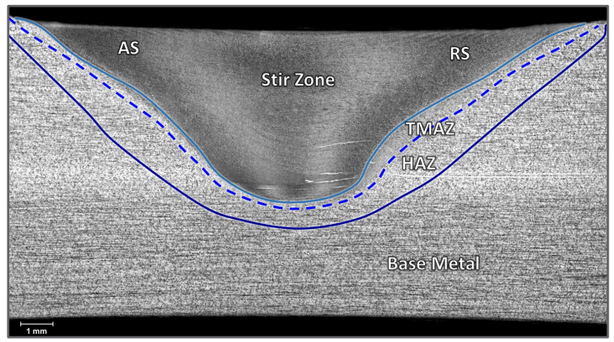
Enabling S&T

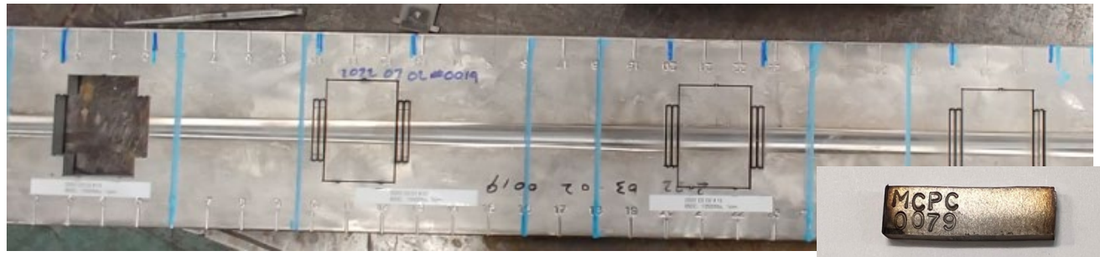
- Accelerate modeling-based prediction of material microstructure
- Exploit noninvasive measurements for real-time validation of product characteristics
- Apply physics-informed data analytics tailored to material processing

Material Processing and Initial Data Creation



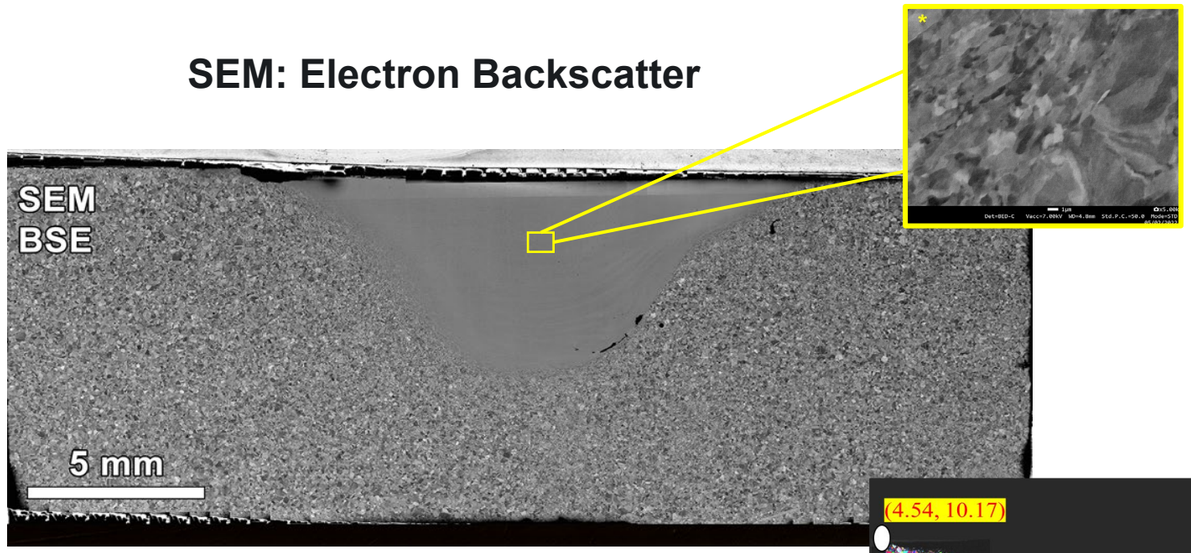




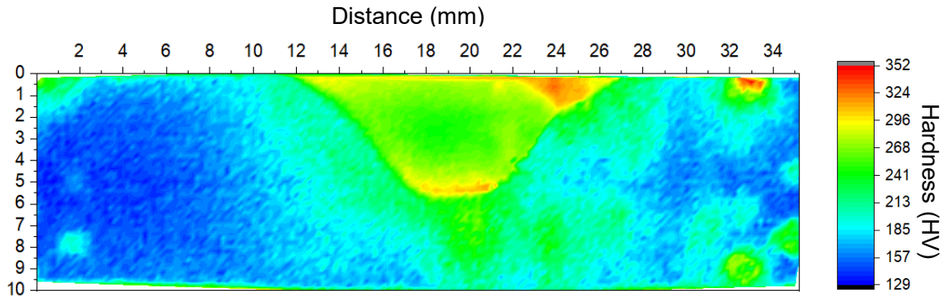


Material Characterization

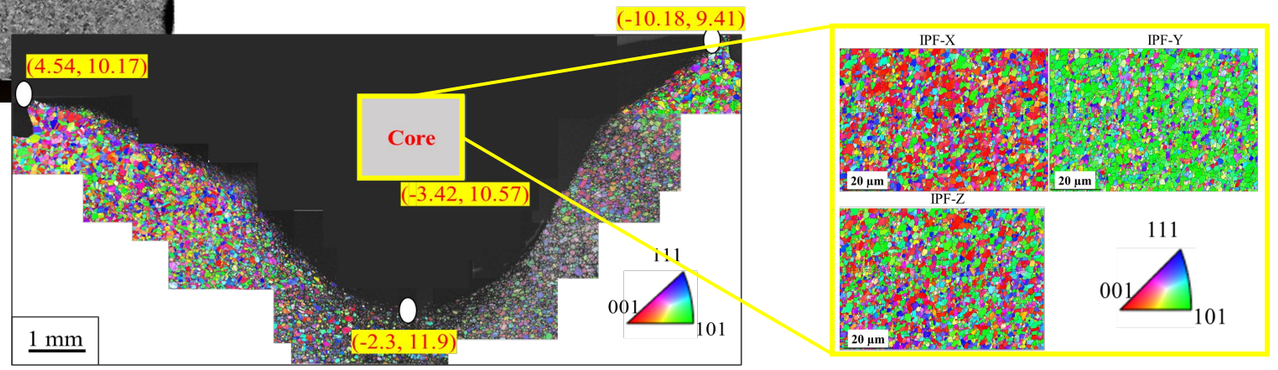
SEM: Electron Backscatter



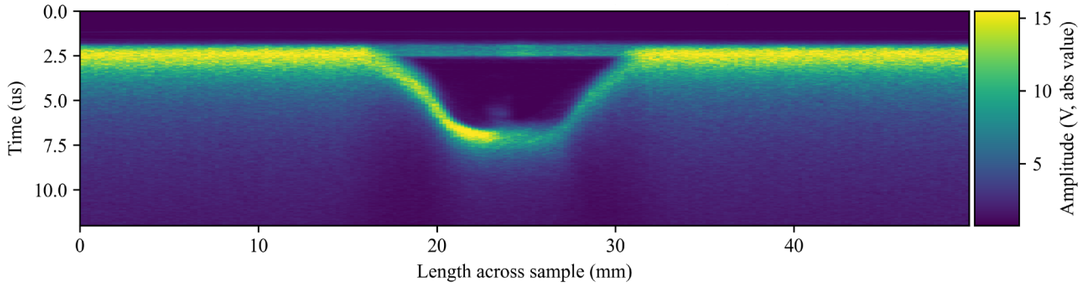
Microhardness Testing



SEM: Electron Backscatter Diffraction



Ultrasonic NDE

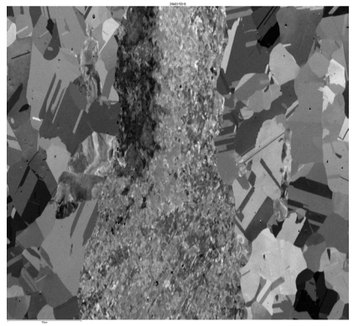
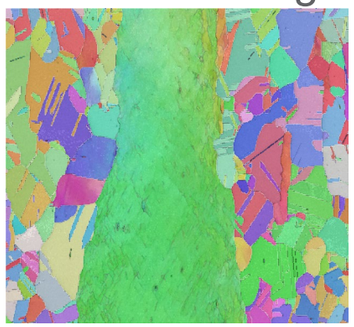


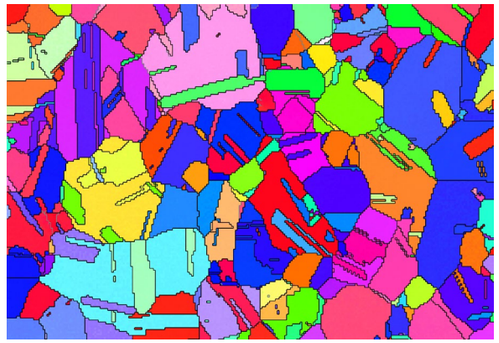
Data Management: Infrastructure Needs

- Connecting experimental equipment (safely!)
- Centralized and assessable data storage location
- Agreed-upon directory structure and naming convention
- Agreed-upon documentation (READMEs, common meta-data)
- Access control
- Revision control
- Provenance tracking
- Data storage that is scalable for different needs
- Handling of multiple modes of data

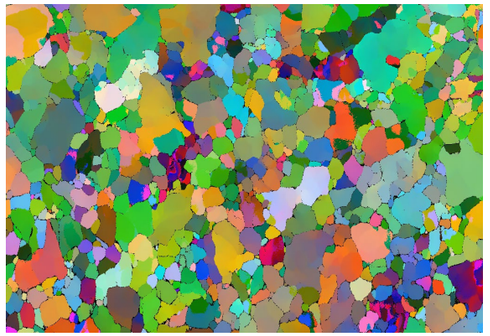
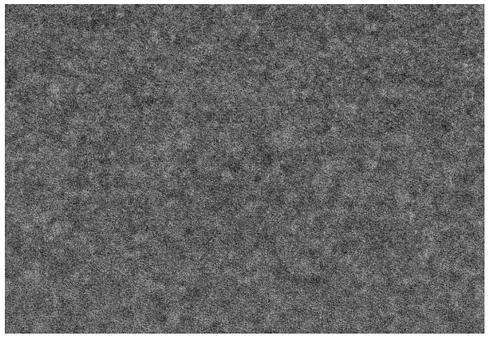
Building on Previous Experience

- EBSD-like Segmentation using SEM
  - EBSD provides more detail but is more time consuming to collect than SEM
  - Train a segmentation model to identify grain boundaries from SEM images



- Transfer to similar material with different processing history
  - Attempted to reuse images collected to calculate grain sizes and distributions
  - Similar material (different types of stainless steel)
  - Very different processing histories
  - Problems encountered by data scientists:
  - Images had been post-processed differently based on original intended use
    - ✓ SEM images were too noisy

eXtremeMAT
Accelerating the Development of Extreme Environment Materials

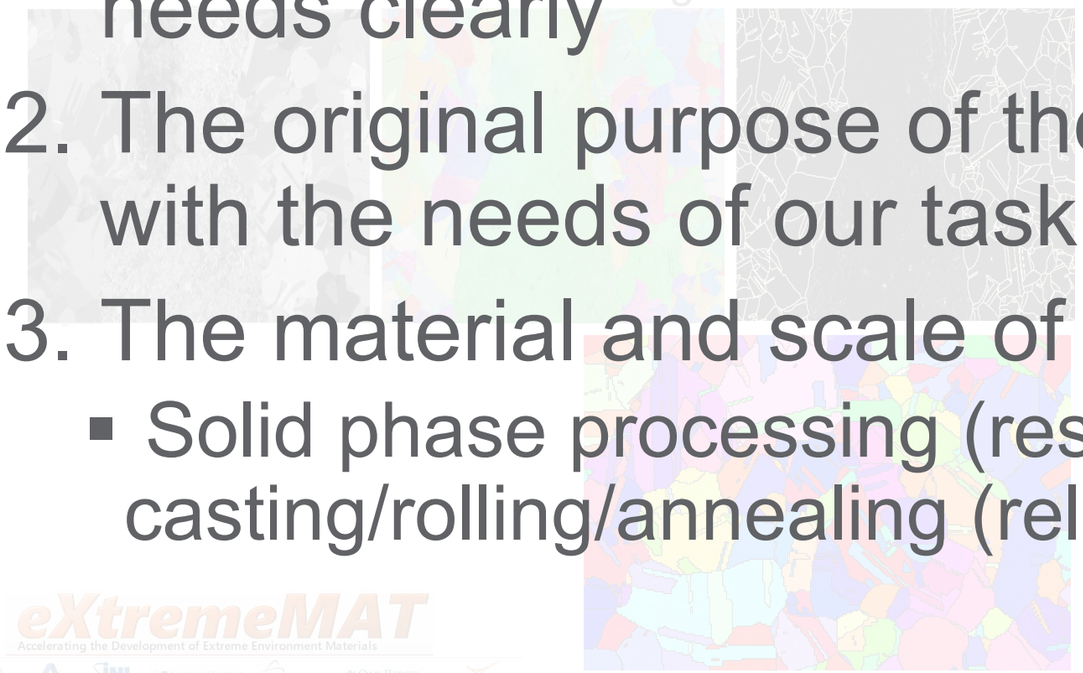
Recently accepted to IMMI
https://arxiv.org/ftp/arxiv/papers/2305/2305.07790.pdf

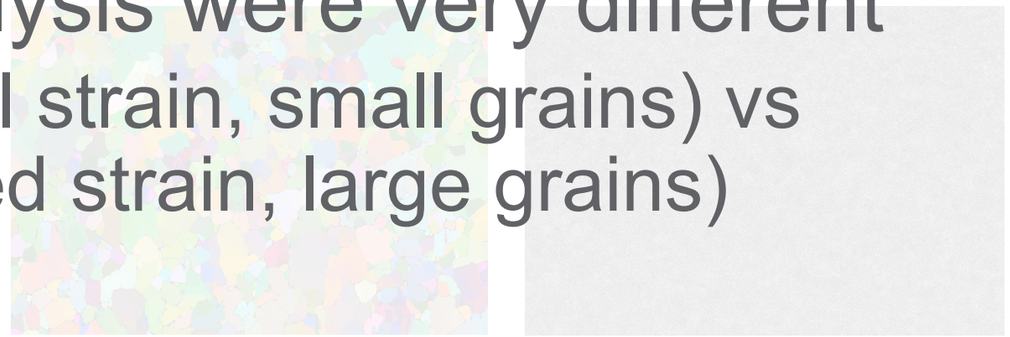
Can we replicate the process for different materials?

- EBSD-like Segmentation using SEM
  - EBSD provides more detail but is more time-consuming to collect than SEM
  - Train a segmentation model to identify grain boundaries from SEM images
- Transfer to similar material with different processing history
  - Attempted to reuse images collected to calculate grain size distribution
  - Similar material (different types of stainless steel)
  - Very different processing histories
  - Images had been post-processed differently based on original intended use
    - ✓ SEM images were too noisy

The *Real* Problems

1. We – as data scientists – didn't communicate our data needs clearly
2. The original purpose of the data collection didn't align with the needs of our task
3. The material and scale of analysis were very different
   - Solid phase processing (residual strain, small grains) vs casting/rolling/annealing (released strain, large grains)





eXtremeMAT
Accelerating the Development of Extreme Environment Materials

Recently accepted to IMMI
https://arxiv.org/ftp/arxiv/papers/2305/2305.07790.pdf

Data Management: Cultural Needs
Communication Mismatch

AI/ML task:

*We need **good** data:*

- Single set of instrument parameters
- Low noise level
- No drift
- Pixel-to-pixel matching of images and EBSD maps
- No borders or scale bars on images
- Pixel-to-micrometer conversion ratio readily given

Microscopist task:

*We can provide **good** data:*

Imaging:

- Sharp images
- Balanced contrast
- Low noise
- High image resolution

EBSD:

- Fine step size
- High indexing rate
- Balance between quality and time

Did you get what you needed?

No.

Why?

Why?

Data Management: Cultural Needs
Communication Iterations

Data Scientists

Jenna Pope

Julia Nguyen

Mohammad Taufique

We need ~1000 grains per image

We need sharp grain boundaries

We need continuous boundaries

We need grain boundary skeletons

Yes!

Microscopists

Ok let's varying the resolution based on grain size (1000x, 2000x, 5000x)

Ok let's vary the step size (50, 100, 200, 500 nm)

Ok let's try cleaning/indexing

Ok let's deconvolute information

Is it working better?

Julian Escobar

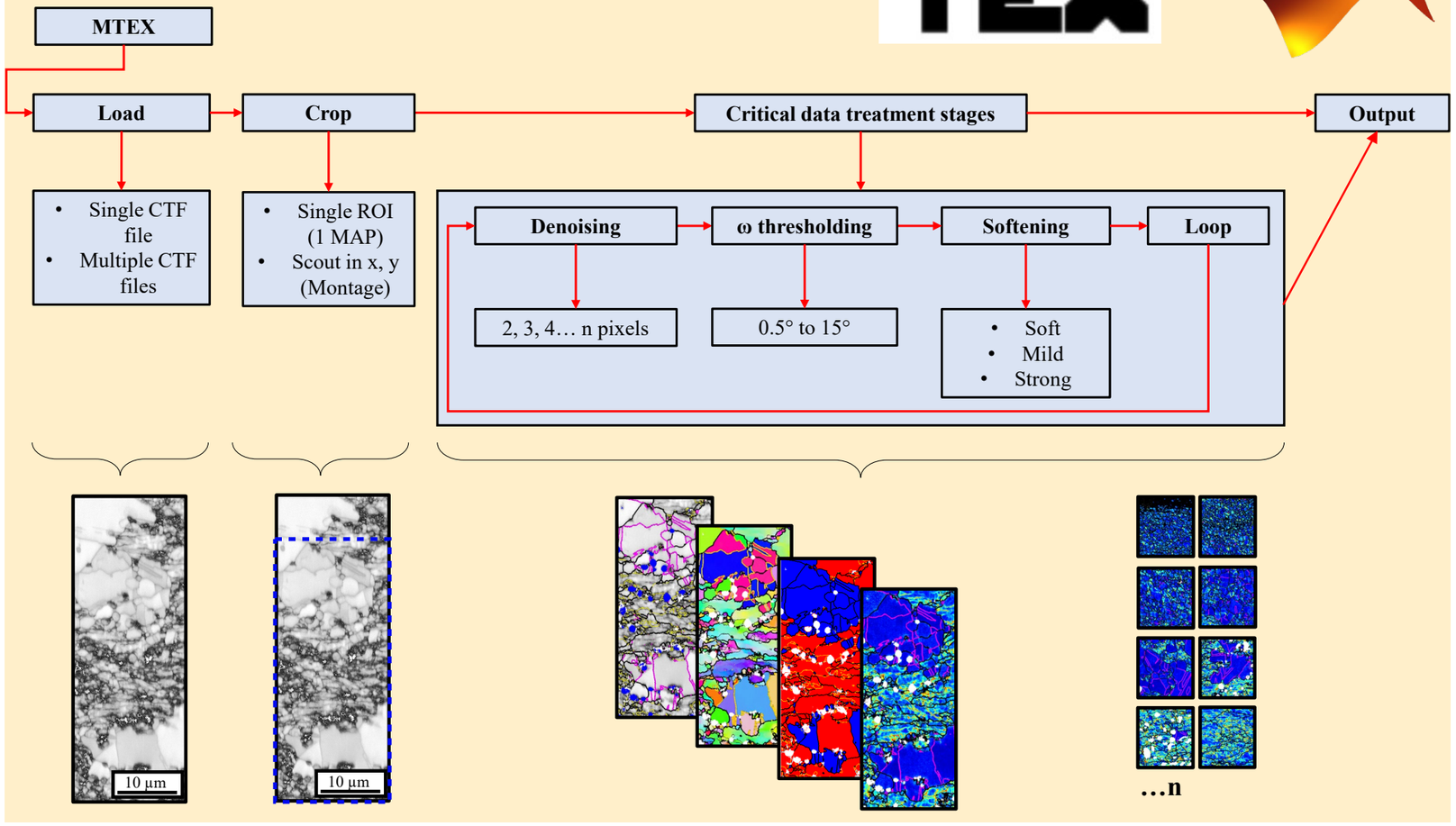
Outcome:
A standard method for BIG EBSD data

AI/ML task:

We need ***more*** *good data:*

- ✓Single set of instrument parameters
- ✓Low noise level
- ✓Sharp and continuous boundaries
- ✓No drift
- ✓Pixel-to-pixel matching of images and EBSD maps
- ✓Pixel-to-micrometer conversion ratio readily given

Microscopist task:



Conclusions

- Data management is more than just hardware and software infrastructure.
- Cultural components need to be considered for effective collaboration across domains.
  - A bridge must be built for effective communication between microscopists (*domain scientist*) and data scientists to assure high-quality research outputs.
  - Data curation ***must*** be tailored for the scientific problem as well as the data science tool.
  - Need a continuous and fast feedback loop.
  - Develop shared ownership of project data.
- Then useful tools and workflows can be developed to speed analysis and discovery
  - Generated a standard procedure for data acquisition, cleaning, and curation for EBSD data, especially useful for large numbers of files.
- Propagating these cultural shifts to other projects and teams remains a challenge.
  - Lessons learned have aided this and other tasks on the project.
  - Working to transfer the lessons to related projects.
  - *Still a work in progress.*

Pacific Northwest
NATIONAL LABORATORY

Thank you