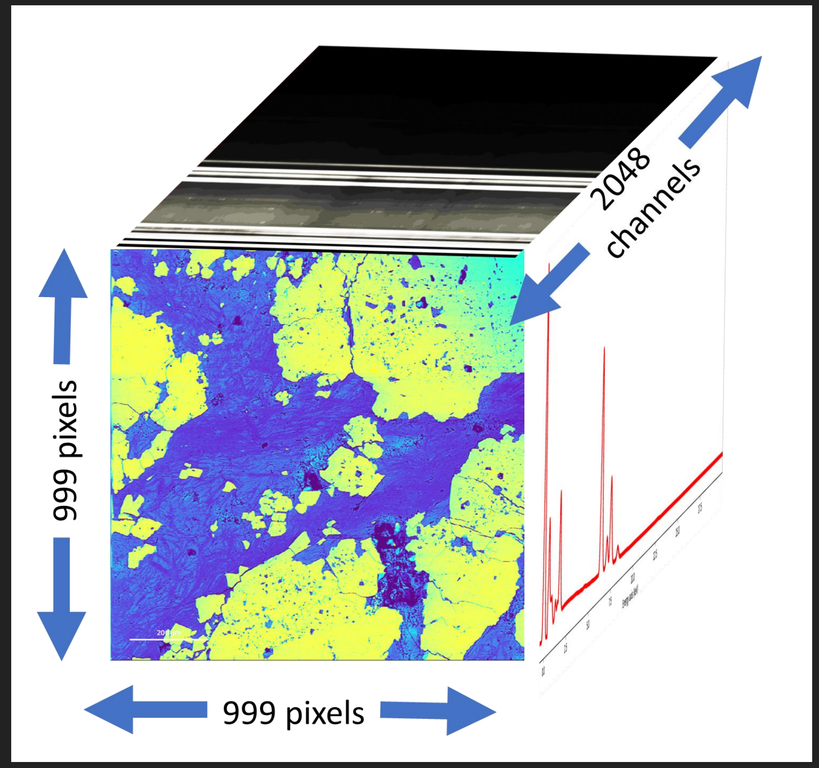
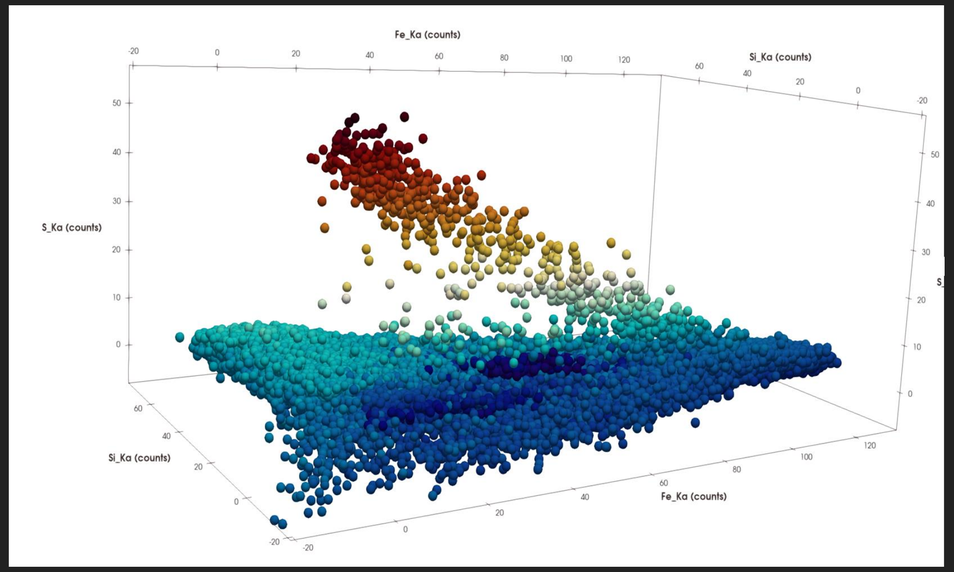
# 2D TO 4D AND BEYOND : THE (MICRO-) IMAGE DATA SPACE




GOLDSCHMIDT 2023 WORKSHOP: MANAGING, PUBLISHING, AND ARCHIVING IMAGERY GENERATED BY GEOCHEMICAL AND COSMOCHEMICAL SAMPLE ANALYSIS

JOSHUA F. EINSLE

SCHOOL OF GEOGRAPHICAL AND EARTH SCIENCES

UNIVERSITY OF GLASGOW

# ACKNOWLEDGEMENTS AND COAUTHORS




**UNIVERSITY OF GLASGOW**

Matt Divers
Sammy Griffin
Francesco Perrone

UNIVERSITY OF STRATHCLYDE

Kate Dobson

UNIVERSITY OF CAMBRIDGE

Richard J Harrison

Paul A Midgley

OXFORD UNIVERSITY

Paul Bagot

UNIVERSITY OF OTTAWA

Brian O'Driscoll

UNIVERSITY OF MANCHESTER

Alex Eggeman

NIST

Austin S. McDannald
Aaron Gilad Kusne

NHM

Tobias Salge

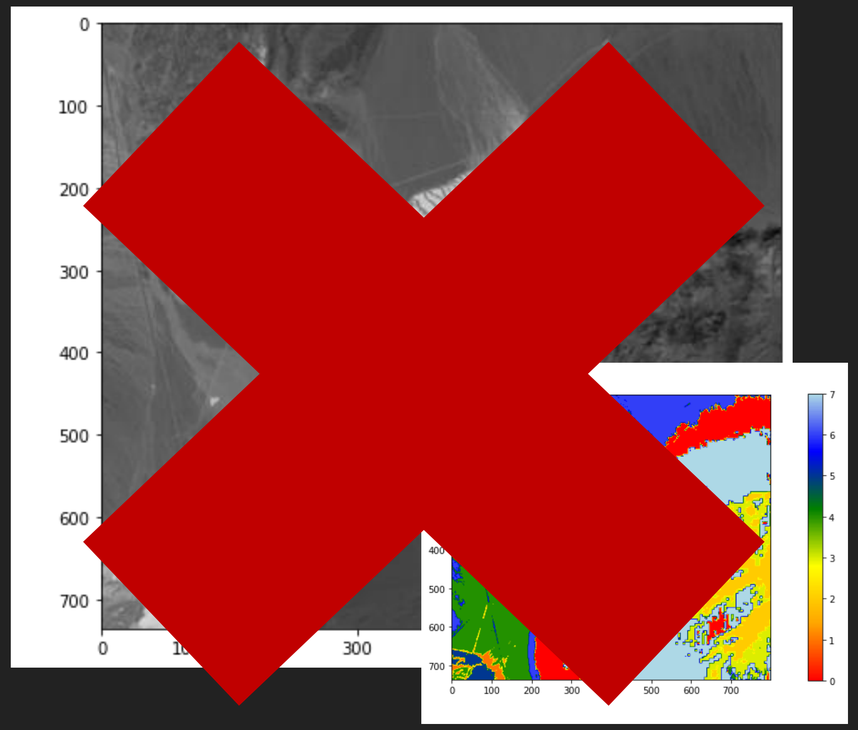
NERC SCIENCE OF THE ENVIRONMENT

# OUTLINE

- Define what  are ‘microscopy’ images
- The need for  (standardised?) open data formats
- Big data is in microscopy now
- Opportunities presented by data archives

# WHAT KINDS OF IMAGES?

## Remote Sensing

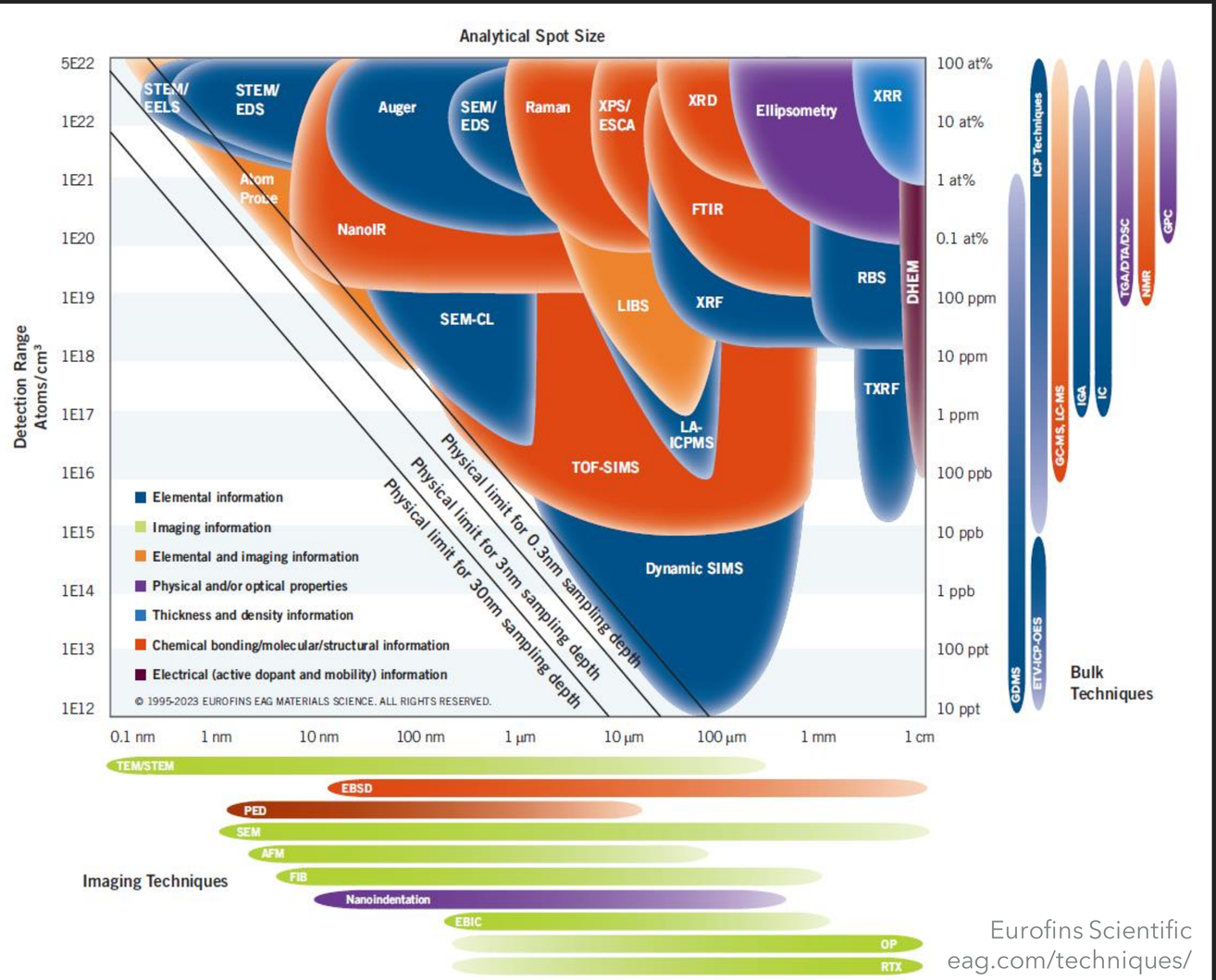


- Already a large community with archiving and sharing standards

## Photographs


- Again, a large community with archiving and sharing standards this is also the domain of AI/ neural networks.

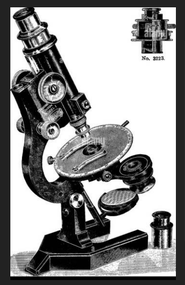
# FOCUS ON MICROSCOPY TECHNIQUES




Eurofins Scientific
eag.com/techniques/

- Microscopy describes a wide range of methods for ‘imaging’ structures in a sample.

# CONNECTING THE DATA TO THE METADATA



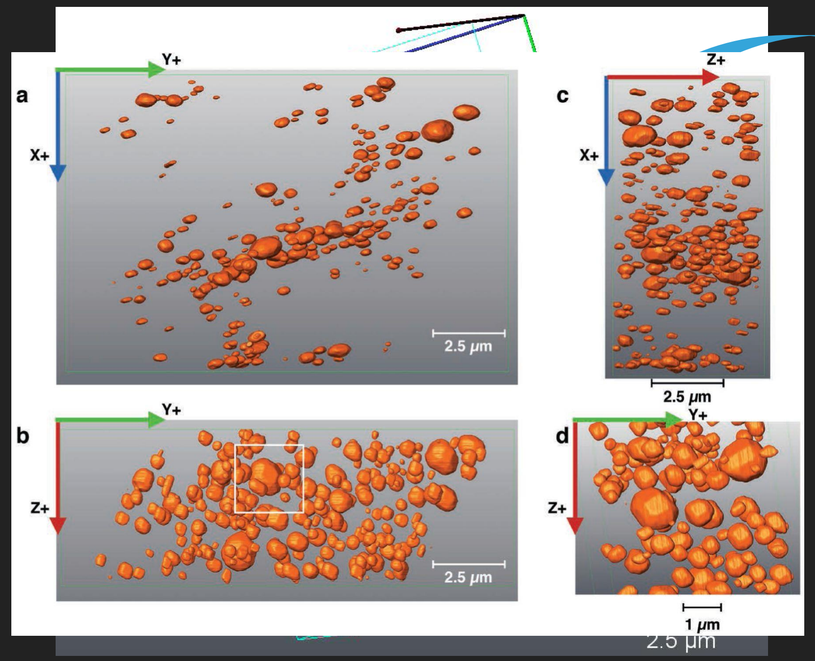


| Optical | X-Rays | SEM | FIB | TEM | APT |
|---|---|---|---|---|---|
| transmitted PPL | XCT/XRM | SE | SE | bright field | Mass Spectroscopy |
| transmitted XPL | XRF | BSE | SI | dark field | field ion microscope |
| reflectedXPL | XRD | CL | FIB-nT | EELS | |
| reflected PPL | LabDCT | hyperspectral-CL | SIMS | EDS | |
| z-stack | XANES (2D & 3D) | EDS | HIM | hyperspectral-CL | |
| bertrand lens | XMCD | WDS | 3D-EBSD | Electron Diffraction | |
| spectroscopy | Holography | EBSD | 3D-EDS-EBSD | Holography | |
| Raman | Family of Diffraction Techniques (Sync) | Raman | TEM Prep | DPC | |

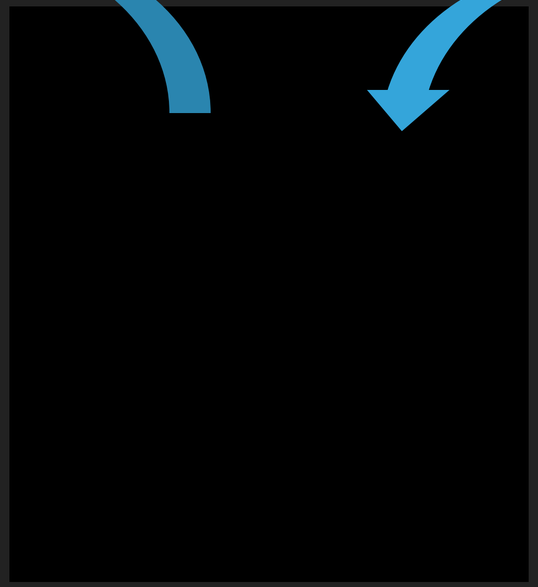
- Key challenge is preserving the link between the data (intensity / counts) and the metadata which gives it meaning (scale, energy, time etc)
  - Contrast imaging techniques often result in simple greyscale images which are saved as Tiff, JPEG, or BMP
    - metadata is not always preserved/ can be easily lost
- Lab based analytical techniques (spectroscopy, diffraction) produce data in proprietary formats which do preserve the meta data but require likened software.
  - Increasingly vendors are supporting open access through API's - which are not currently standardized (both for commercial as well as domain reasons)
- Open files formats allow for packaging of these two data objects together - raw data and the metadata
- Multidimensionality drives data volume – What is essential for reproducibility?

# DATA IS NOT THE PUBLICATION FIGURE

**3D Visualisation derived from….**

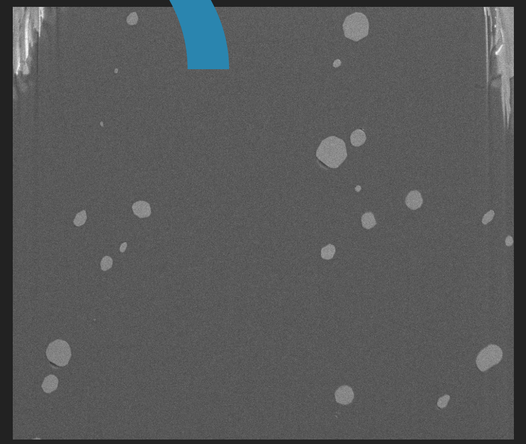



**A binary image stack, where bright Fe grains are segmented from olivine matrix…**



**Can only be done once the image stack is aligned…**

**From the raw data**

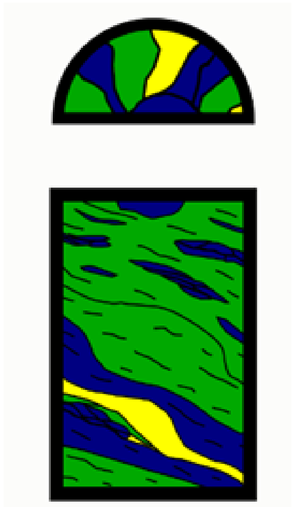


**Movie of Fig 4 from Einsle et al Am Min 2016**

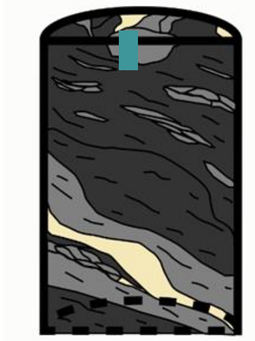
- We process and manipulate the raw data at each step along the way.
  - Even if the figure is an image, the publication process present interpolated and modified version of the original micrograph.
  - For reuse the ‘raw’ data is the most useful as it allows analysis innovation as technology advances.

# ERA OF CORRELATIVE MICROSCOPY

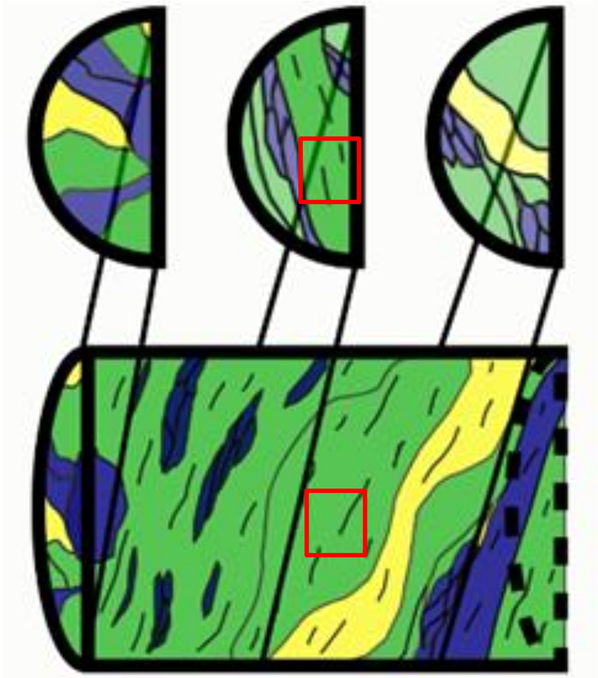
2D EDS Mapping



XRM

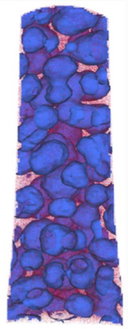


**‘Correlative’ Microscopy hints at a samples 3D Minerology**



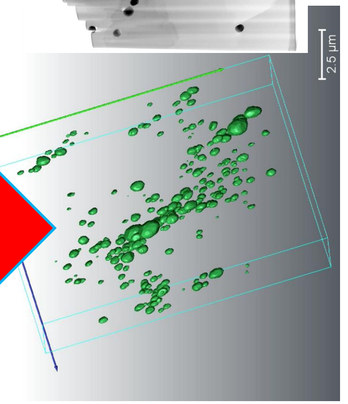
**Guess to Identify Regions of Interest**

APT



S/TEM

FIB-nT



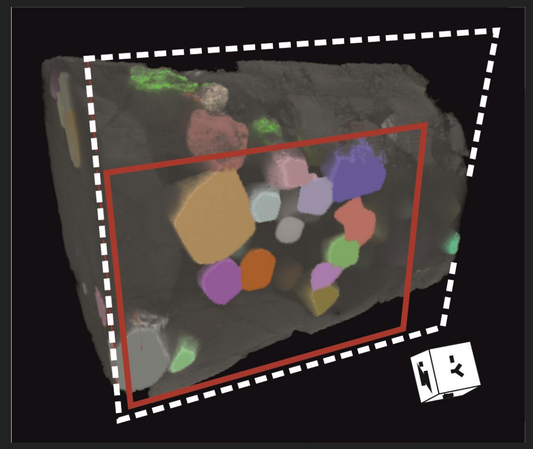

- Assuming a linear discovery pathway from low resolution to high
- 'Correlation' is little more than overlaying two images and squinting
- High resolution experiments are time consuming
- Conclusions are not always ( rarely?) based on a statistically representative sample size
- The connection between features across length scales is based on 'expert knowledge' or 'experience' not clear from the outset.

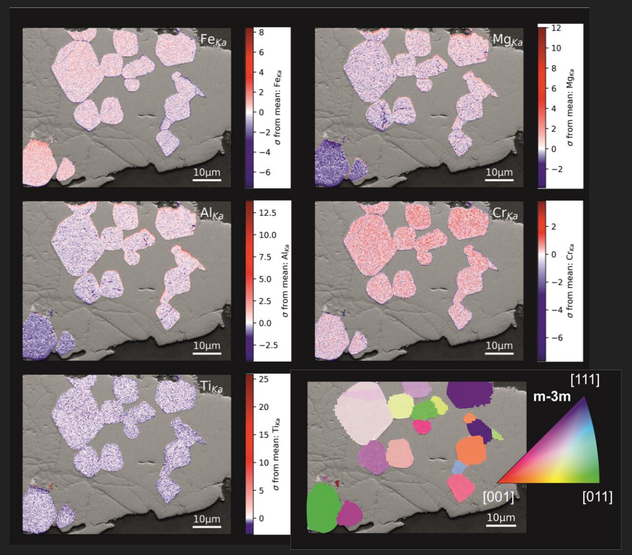
# WE ARE IN THE ERA OF BIG MICROSCOPY DATA

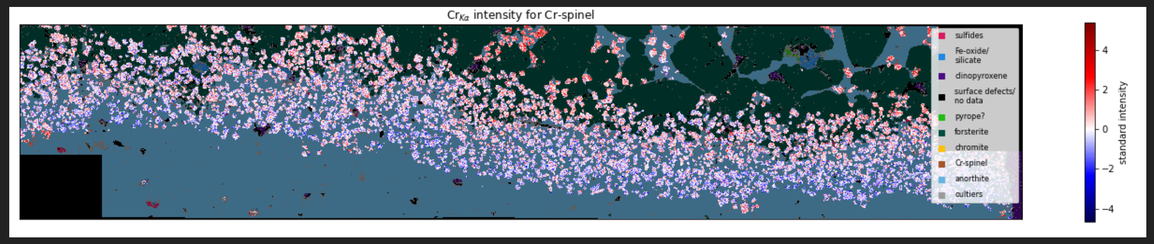
## Chromite cluster formation within the Rum Layered Igneous Complex, NW Scotland:

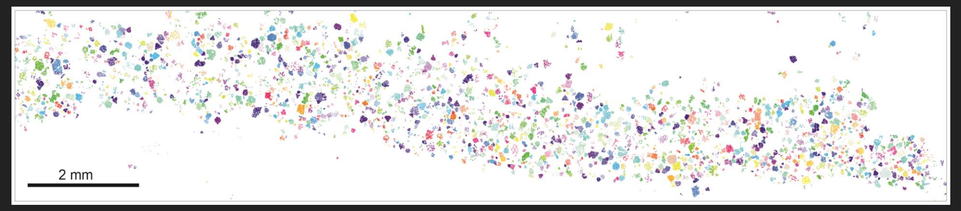
A multi-modal approach

Sammy Griffin



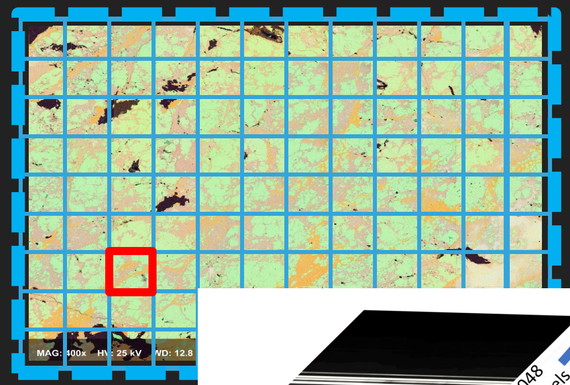


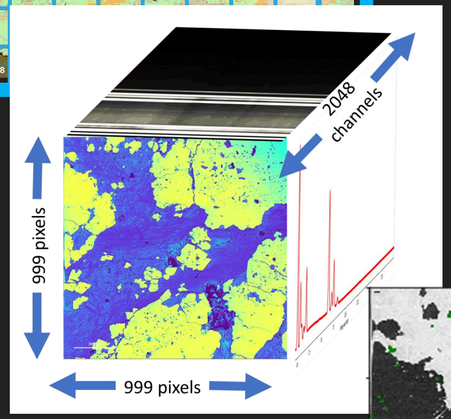



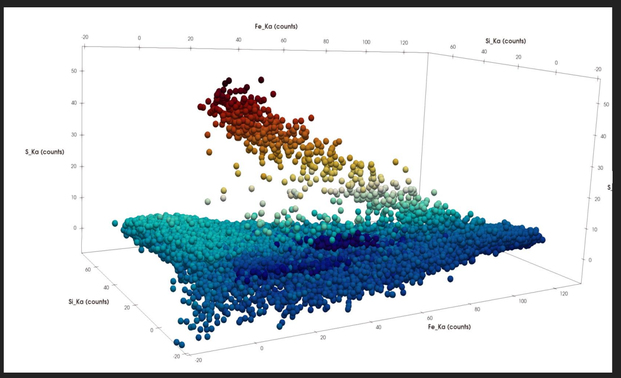



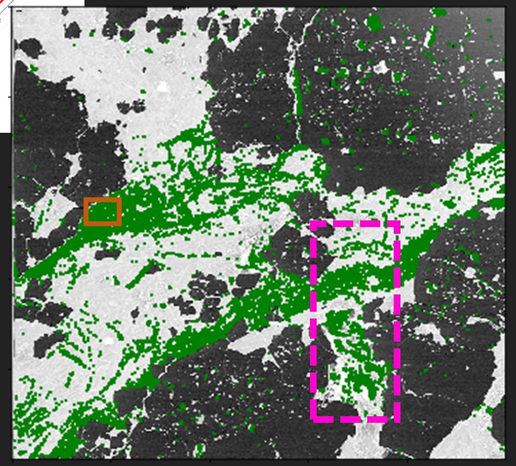

## Working around microanalytical data gaps to make multiscale predictions:

Exploring the Bon Accord nickel mineral species



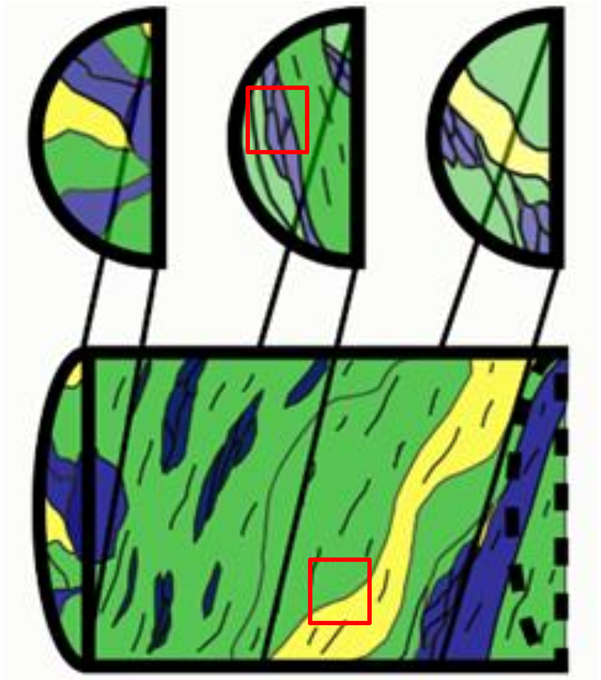
# CONNECT ALL THE DATA

THE NEED FOR OPEN DATA REPOSITORIES

2D EDS Mapping

XRM

**Machine Leaning**
**Data Fusion = 3D**
**Minerology**

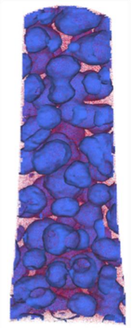


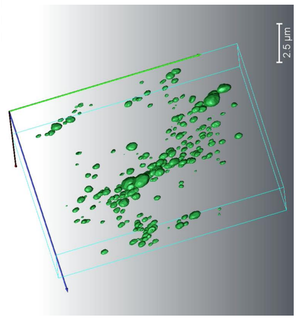
**APT**

**S/TEM**

**FIB-nT**

**Fill in new information to modify multiscale understanding**






- Open data repositories would allow the data to be connected…
  - Not just in one sample (as shown)
  - But also across similar samples collected in different labs.
- Result in a connected virtual sample (Digital twin) from the millimetre to the nanometre.
- Initial experiments have a specific end goal in mind.
  - Many experiments collect ‘companion data’ (similar to companion metals)
  - With archives these digital twins allow for re-mining of the data for new insights
- Need tools for connecting the repositories (PANGEO??)

# RESTATING THE CHALLENGES

- Due to the variety of microscopy data formats, there is a clear need for standard open data formats which bundle the raw data with the metadata
  - This should be supported by vendor distributed API's for export / sharing of data.
- We are now in the era of big data- automation has made collection of large multidimensional datasets routine
  - This presents challenges both in processing and archiving of data
  - Real opportunity for the community to leverage advances in distributed (/cloud) computing and storage.
- Development of microscopy archives would allow for correlative approaches to take on new meaning
  - Digital twins of a sample become increasingly meaningful as correlative techniques enable the connection of multi-modal experiments.
  - Correlative studies across samples collected by different labs
    - Examples: same samples – different labs, related samples – same lab, related samples – different labs
  - Connecting measurement types from different labs