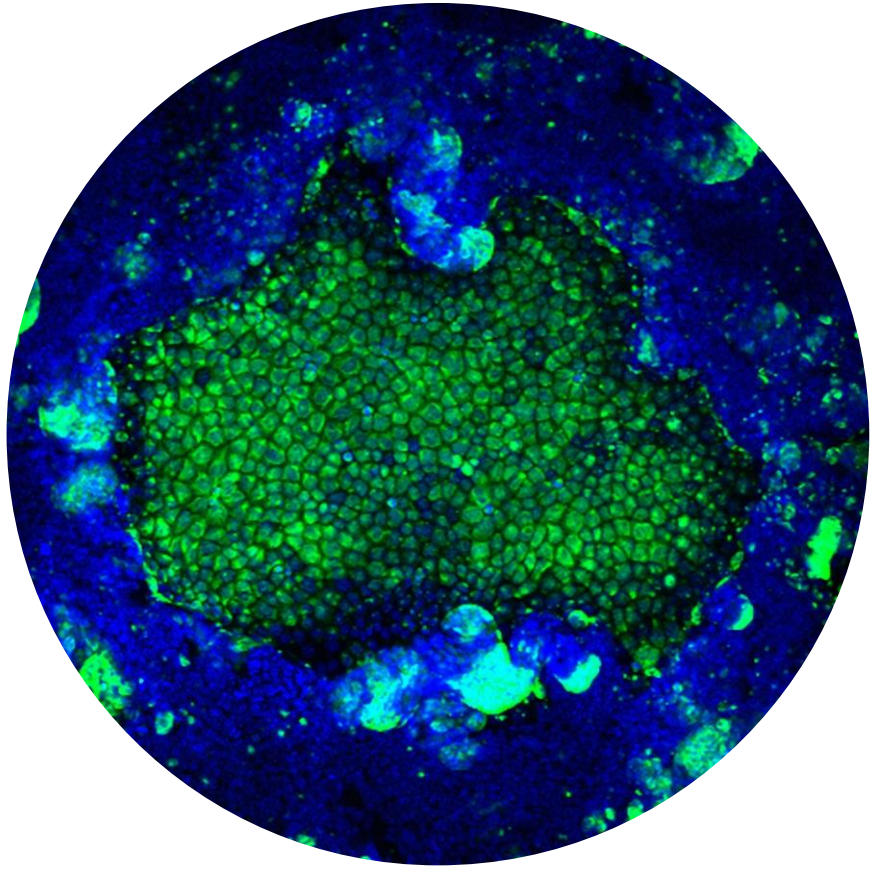
Artificial intelligence in the big-data era

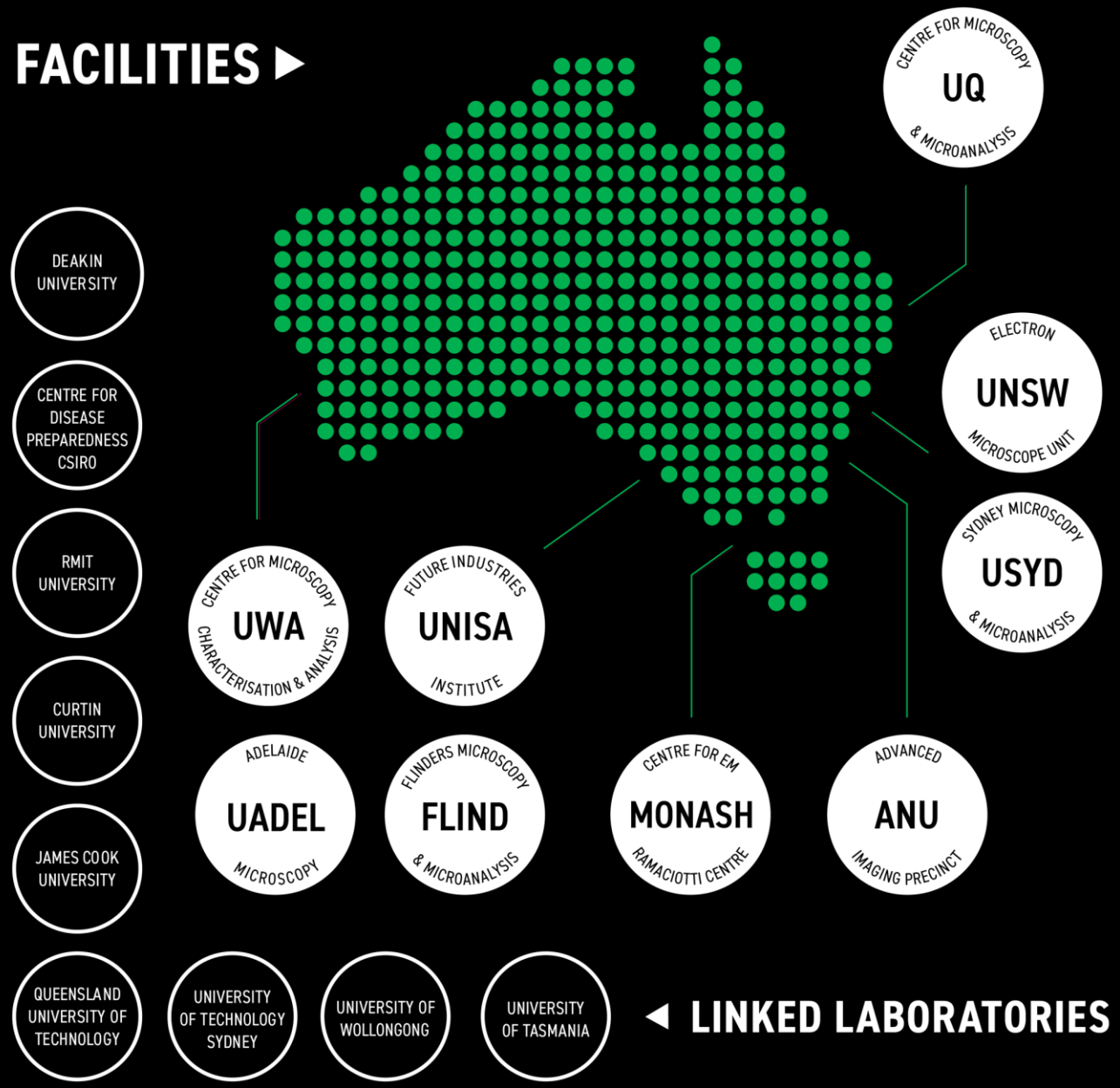
An opportunity for FAIR microscopy



Dr David Poger
Research Data Manager

MICROSCOPY AUSTRALIA

OPEN-ACCESS NETWORK OF FACILITIES



EACH YEAR WE SUPPORT...

BASED ON AVG ANNUAL DATA 2018-2021

250+
INSTRUMENTS

130+
EXPERTS

3200+
USERS

150+
INDUSTRY CLIENTS

250,000+
HRS BEAMTIME

100,000+
MYSCOPE USERS

1,400+
PUBLICATIONS

15% GEOSCIENCE & ENVIRONMENT
48% PHYSICAL & MATERIALS
37% BIOLOGICAL & MEDICAL

37% RESOURCES & ENVIRONMENT
46% MANUFACTURING
17% BIOMEDICAL

Big data in (electron) microscopy

Over the past 10 years, the rapid advances in electron microscopy (cryogenic electron microscopy/tomography, volume electron microscopy, correlative light and electron microscopy *etc*), have enabled the collection of high-volume, high-complexity and high-resolution data at an ever-increasing speed.

- A single experiment can generate TB's of data per day.
- Some facilities generate 500 TB–2 PB of data per year.

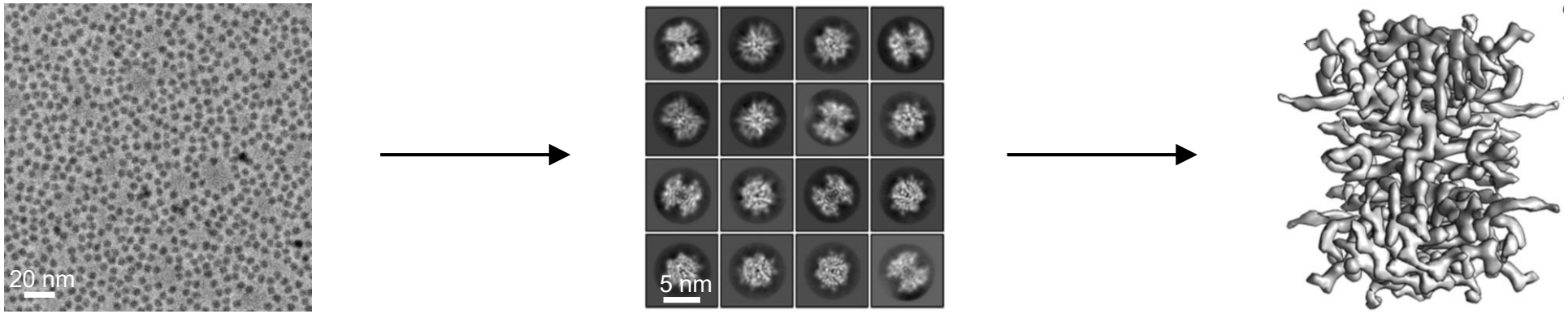
The ability to extract value from big data is key.

Dedicated techniques, technologies and infrastructure are needed.

Artificial intelligence in electron microscopy

AI is increasingly being used in data processing and analysis to extract meaningful information from large datasets:

- to improve the resolution and sensitivity of microscopes
- to reduce noise in datasets
- to identify and classify patterns and shapes
- to count and classify cells.



Fan et al (2019). *Nature Communications* **10**, 2386

Artificial intelligence in electron microscopy

AI is increasingly being used in data processing and analysis to extract meaningful information from large datasets:

- to improve the resolution and sensitivity of microscopes
- to reduce noise in datasets
- to identify and classify patterns and shapes
- to count and classify cells.

So far AI has mostly been used:

- to speed up data processing and analysis
- to make data processing and analysis methods more user-friendly and more accessible to new users.

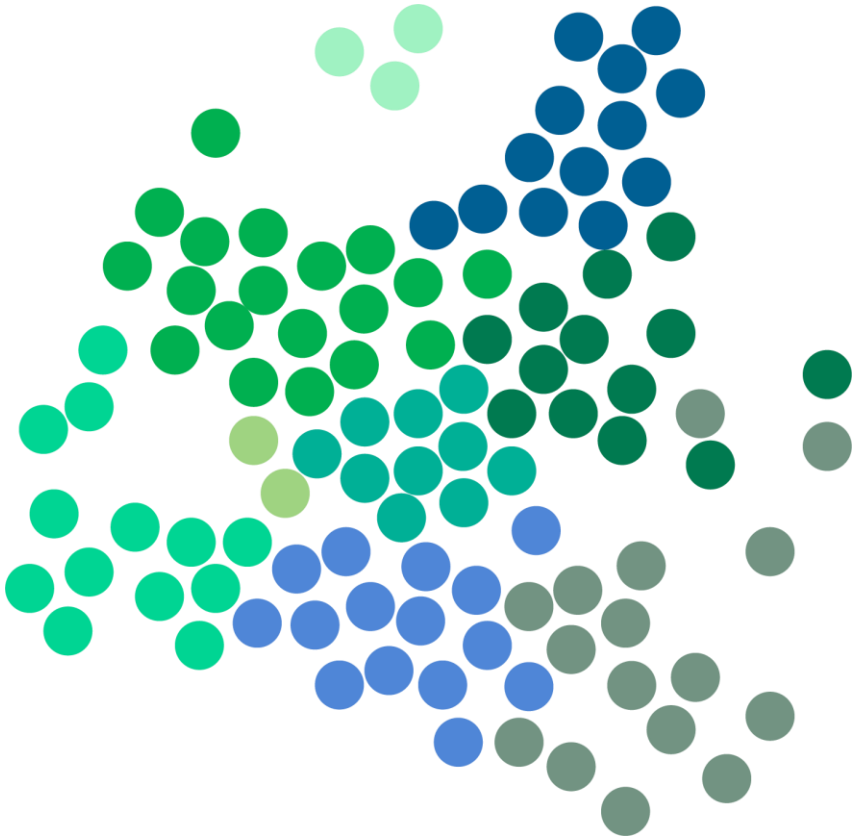
Standardisation in data annotation

Data annotation is tedious, repetitive, time-consuming.

Annotation is often non-standardised:

- it may change from a dataset to another
- it may change from an experiment to another
- it may change from a researcher to another.

This is a barrier to data sharing and the use of standard tools for analysis and visualisation.



Standardisation in data annotation

Current efforts to standardise data annotation:

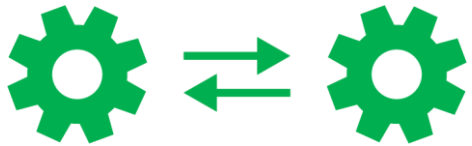
- AI4LIFE project: development of standards through harmonised and interoperable tools and methods
- Recommended Metadata for Biological Images (REMBI): metadata guidelines
- Minimum Information about Highly Multiplexed Tissue Imaging (MITI): data/metadata in genomics and microscopy of tissue images
- QUality Assessment and REProducibility for instruments and images in Light Microscopy (QUAREP-LiMi): development of standards, guidelines, metadata models and tools

Data annotation: an opportunity for AI

AI can automate data annotation.

automated and standardised image annotation and workflows

controlled vocabulary and common machine-readable metadata standards



interoperable programs

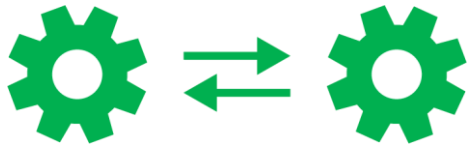
towards a FAIRer microscopy

Data annotation: an opportunity for AI

AI can automate data annotation.

automated and standardised image annotation and workflows

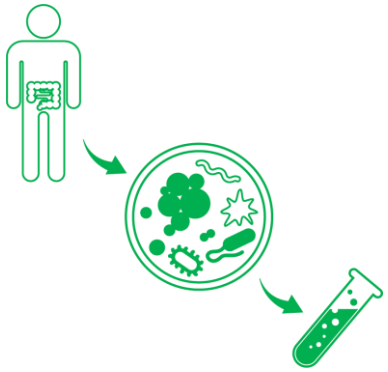
controlled vocabulary and common machine-readable metadata standards



interoperable programs

FoR codes

towards a FAIRer microscopy



provenance

AI-assisted data analysis for FAIR data

The development and integration of AI methods in microscopy workflows can promote FAIR data.

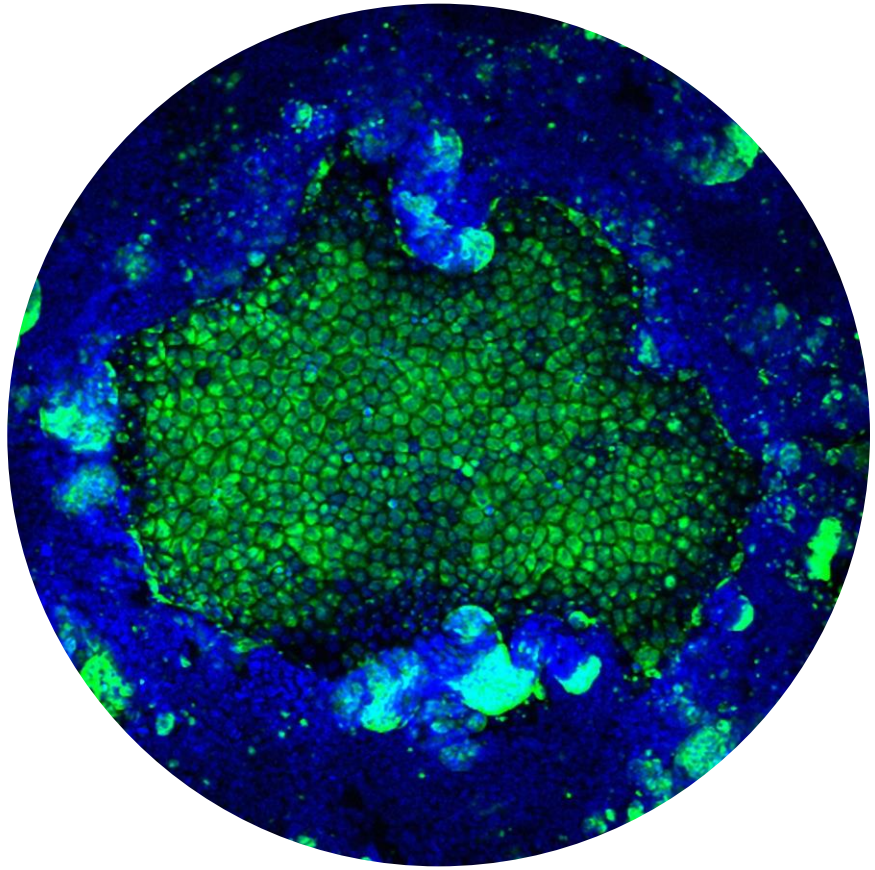
It’s happening at the right time: convergence of the operationalisation of FAIR, big-data challenges and availability of AI tools.

Microscopy Australia is working with partners across the world to standardise/harmonise vocabularies to describe data (Global BioImaging, Euro BioImaging, BioImaging North America).

Thank you.

david.poger@sydney.edu.au

MICROSCOPY AUSTRALIA



Further reading:

Poger, D et al (2023) Big data in contemporary electron microscopy: challenges and opportunities in data transfer, compute and management. *Histochem Cell Biol*. https://doi.org/10.1007/s00418-023-02191-8