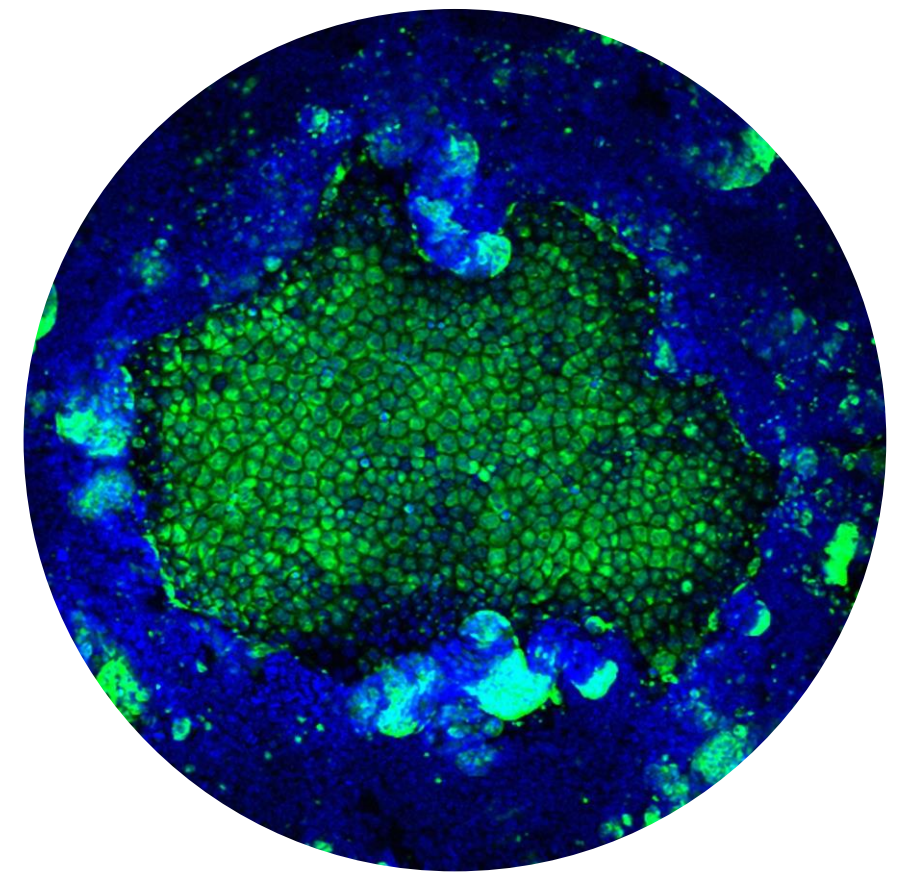


Metadata collection at microscopy research facilities

From distant goal to near future?

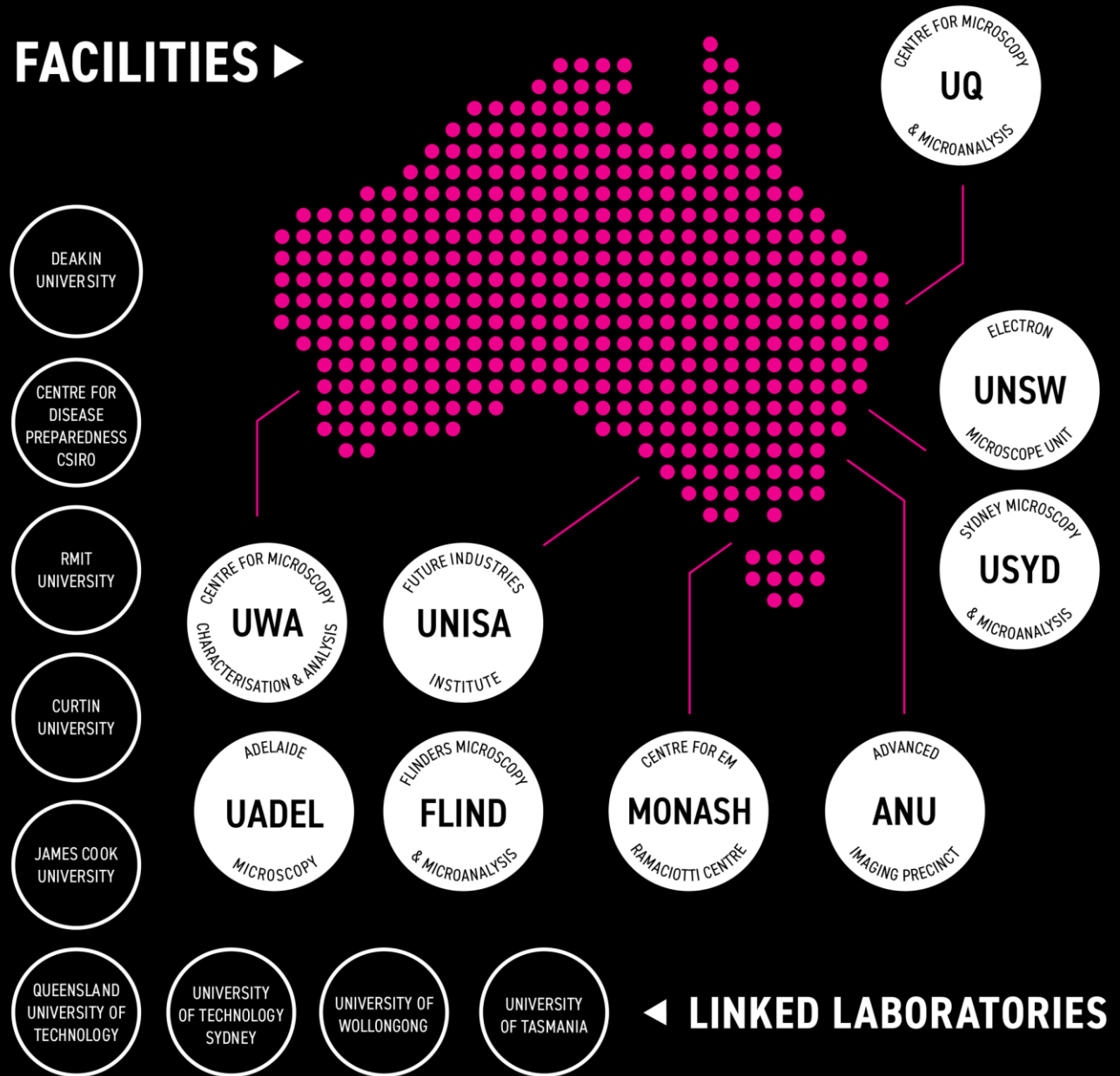
Dr David Poger
Research Data Manager



OPEN-ACCESS NETWORK OF FACILITIES



FACILITIES ▶



◀ LINKED LABORATORIES

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+
MYSOPE USERS

1,400+
PUBLICATIONS



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

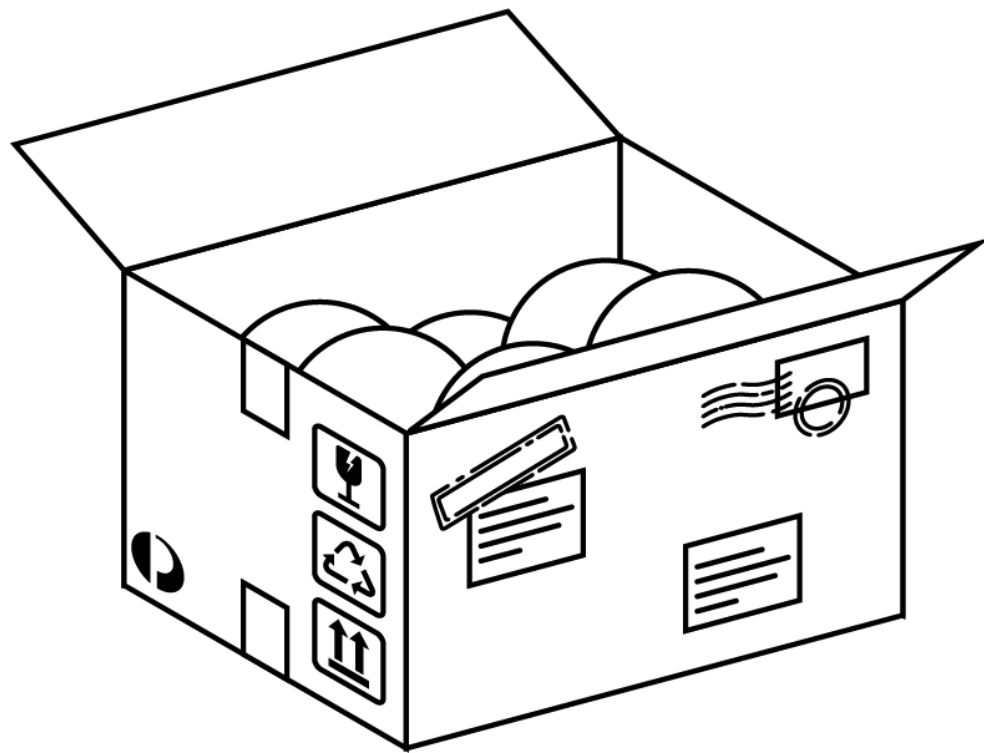


37% RESOURCES & ENVIRONMENT
46% MANUFACTURING
17% BIOMEDICAL

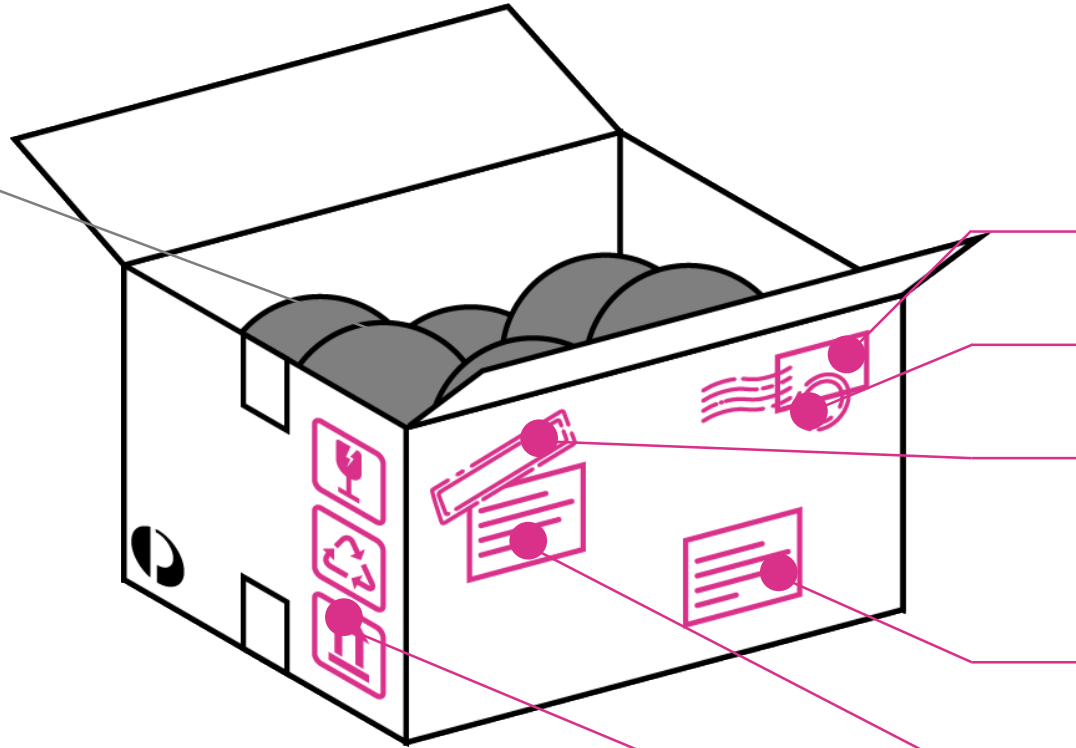




metadata: data about data



content



postage

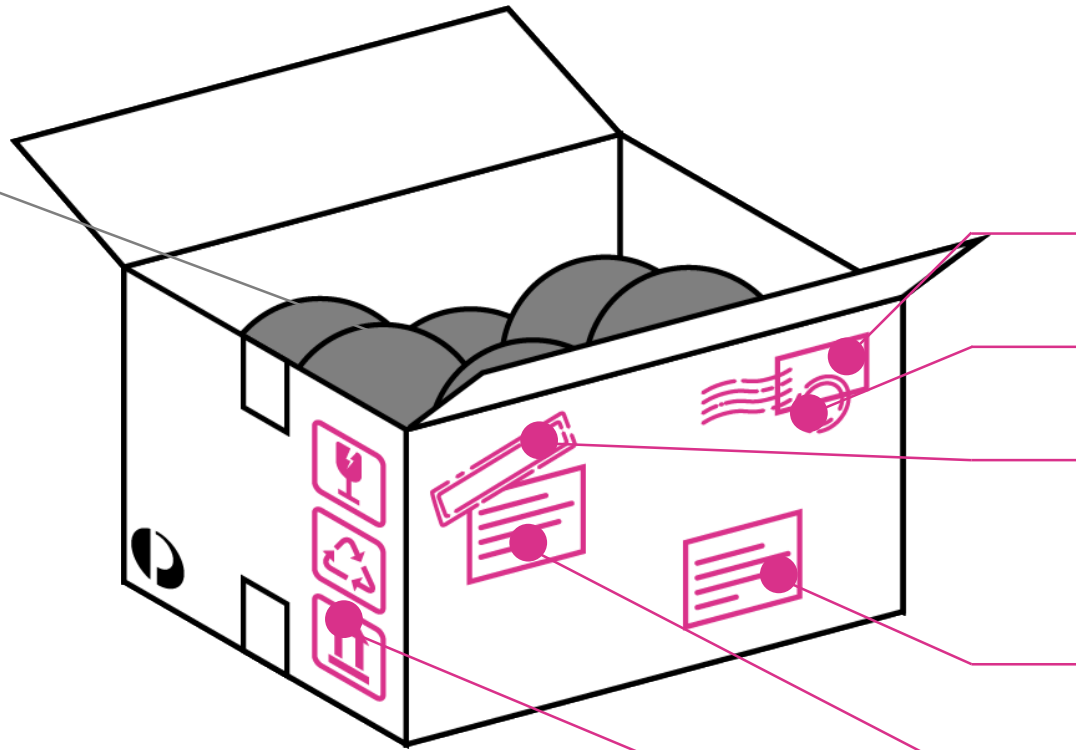
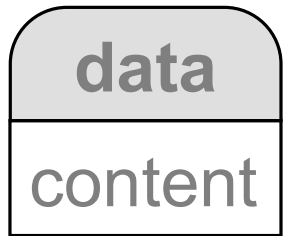
mailing date and place

sensitivity and
importance

recipient's address

sender's address

packaging and handling
information



? | **why collecting metadata**

Metadata are a source of diverse information

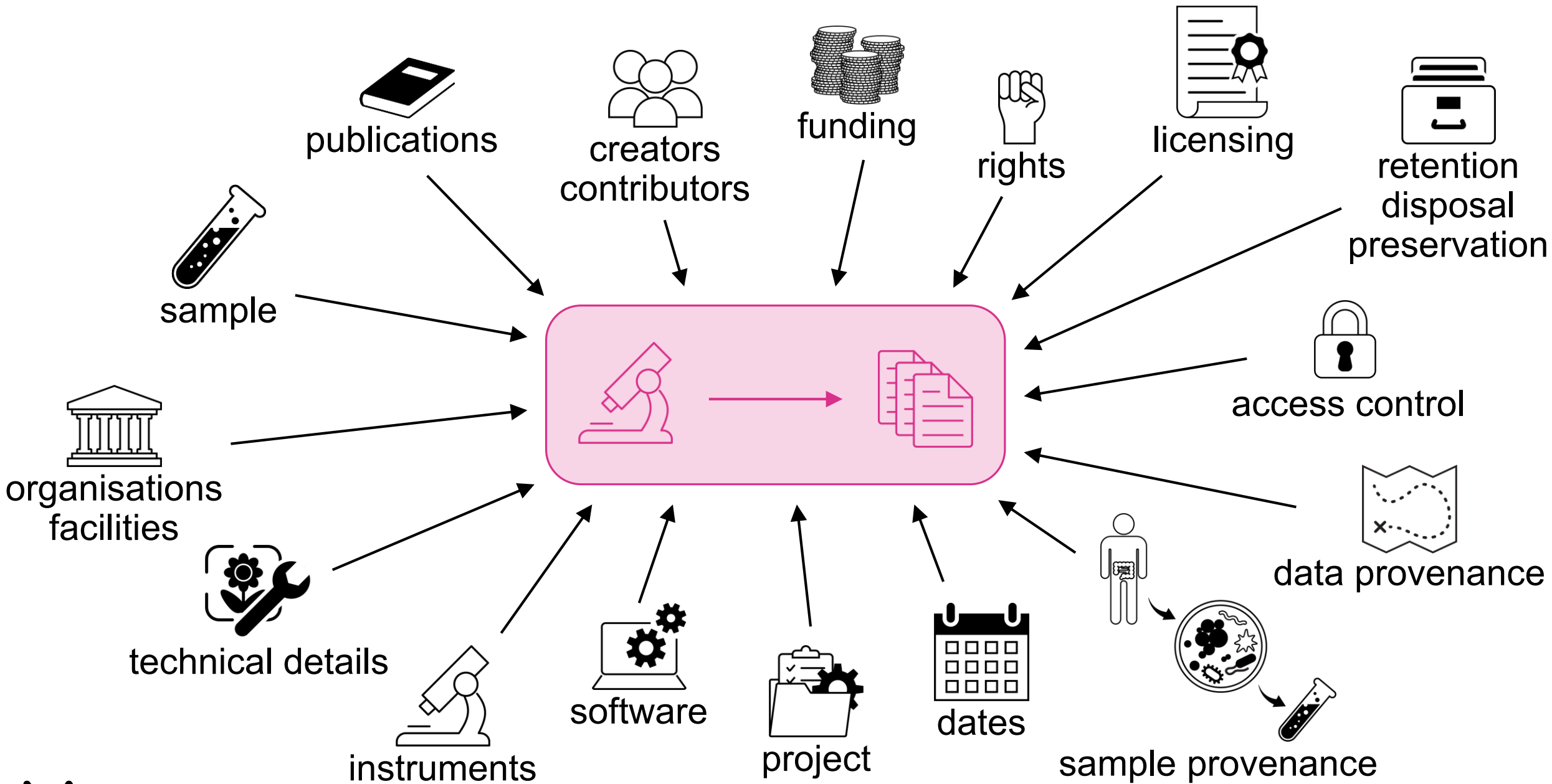
Metadata provide information on:

- the resource itself: title, description, creator(s), keywords *etc*
- technical details: file format, resolution, hardware, software *etc*
- details associated with the resource: funding, organisation *etc*
- management of the resource: preservation, quality control, licensing, rights, access control *etc*

There are several classifications of metadata.

- here the focus is on generic metadata, not instrument metadata.





** non-exhaustive list*



Metadata are a source of useful information

Metadata enhance discovery, accessibility, use and provenance of data.

- They underpin the FAIR and CARE data principles
- They support quality control and data reproducibility
- They contribute to research integrity.

Metadata are useful to all stakeholders in research : researchers, institutions, facilities, funders, publishers *etc*, in particular for:

- reporting
- impact tracking and return-on-investment assessment
- storage and records management.





when to collect metadata

Many opportunities to collect metadata

Metadata can be captured at various stages of a research project and throughout the research data lifecycle.



- during an experiment at a facility
- creation of a research data management plan
- creation of a project at a facility
- booking of an instrument at a facility
- description of data methods in an electronic notebook
- data annotation (e.g. OMERO)



current metadata collection

The beginning of a metadata journey

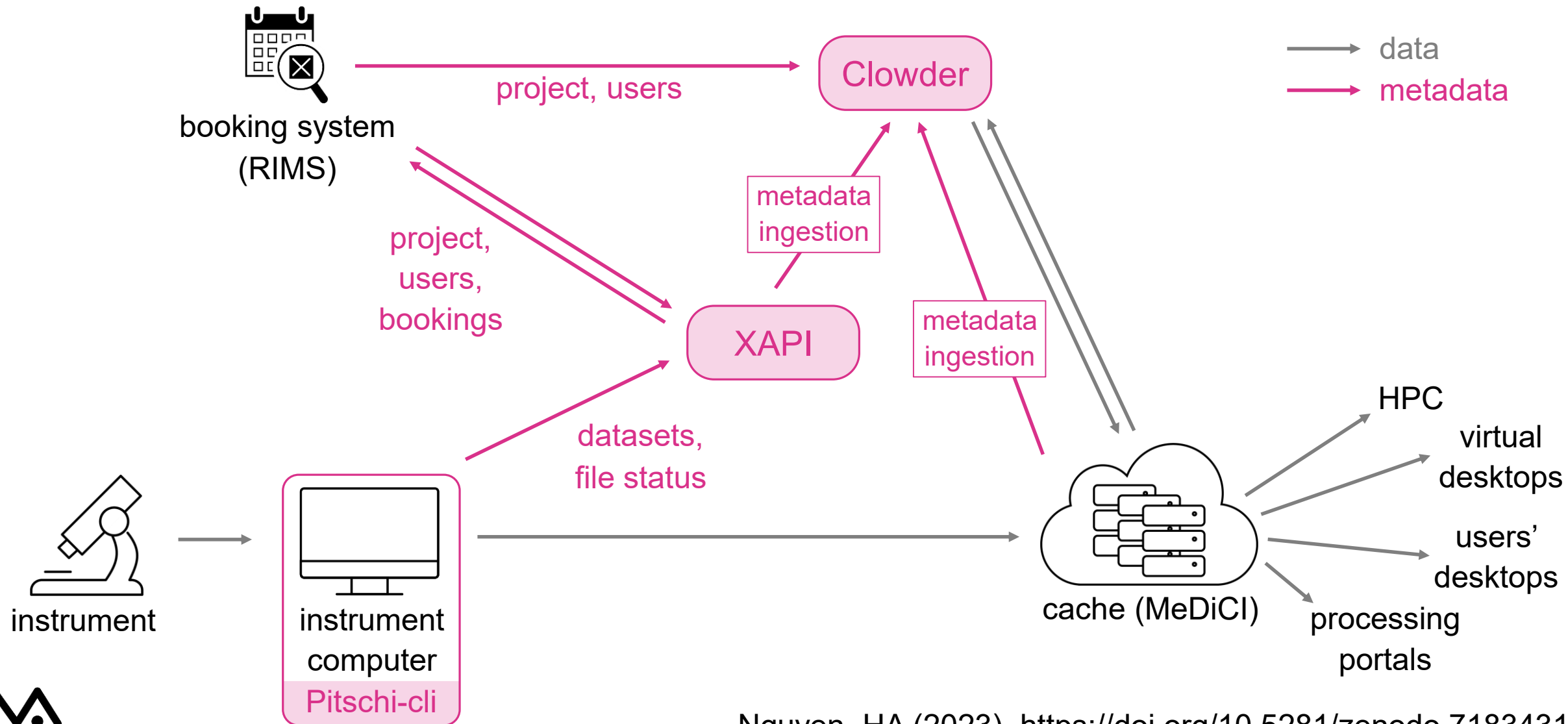
Currently, only the Centre for Microscopy and Microanalysis (The University of Queensland) can collect metadata through systematic and automatic extraction and ingestion of metadata using Pitschi:

- Pitschi is based on the data management framework Clowder
- Pitschi supports the entire research data lifecycle by storing, indexing and annotating data
- Pitschi extracts metadata from various sources: booking system, instrument computer (or third-party computer).

Other facilities interested in testing or deploying Pitschi using their current booking or facility management system (ACLS, EZBooking).



Pitschi (Particle Image depoT using Storage caCHing Infrastructure)



What are the barriers to metadata collection?

Sustainable metadata collection must be seamless with limited human intervention. This implies:

- systematic and automated metadata ingestion from various sources
- interoperability between several systems (e.g. booking system of a facility, institutional data repository, electronic laboratory notebook, data-management system or laboratory information system).

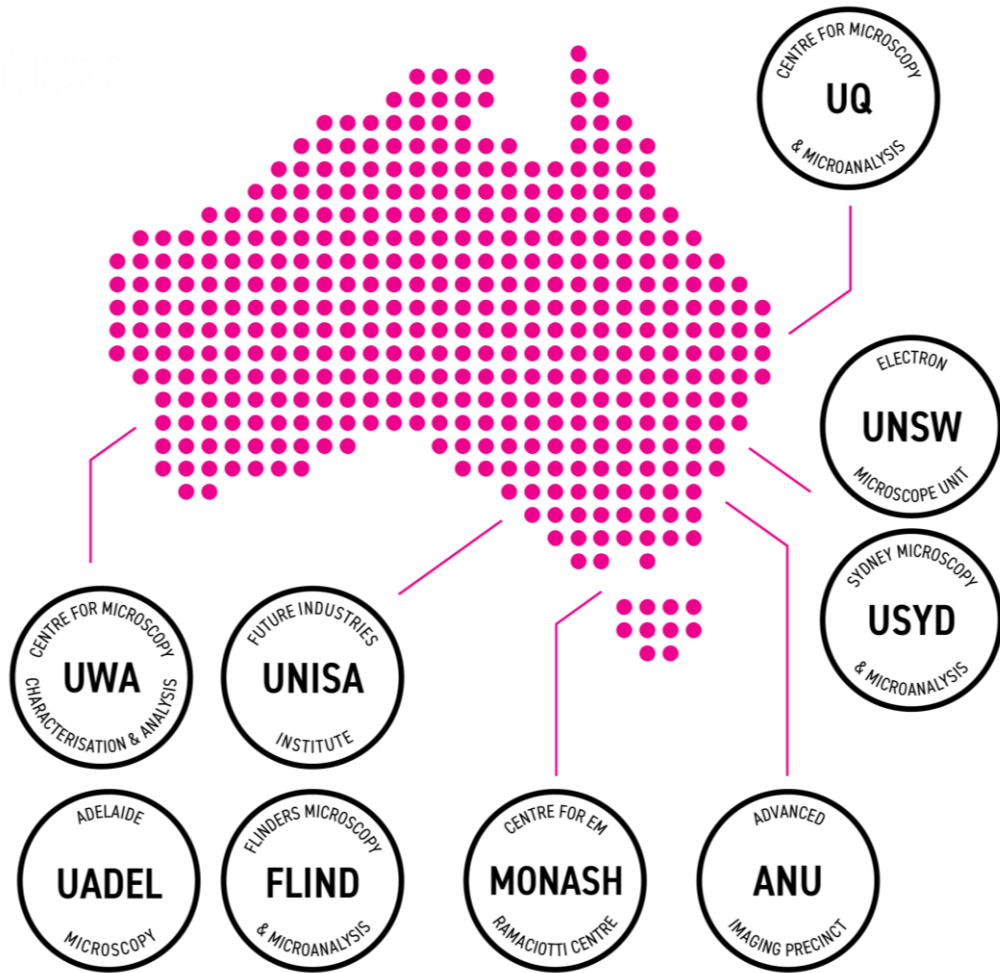
The institutional environment is critical:

- many of the operations of a facility depend on the systems used by their host university
- institutional awareness on research data management and FAIR is essential.



 | **the metadata journey**

Research data management at Microscopy Australia



9 institutional environments

Each university:

- has its own data governance
- is on its own journey (data management, FAIR, CARE).

Various community and disciplinary standards (or lack thereof).

Microscopy Australia promotes:

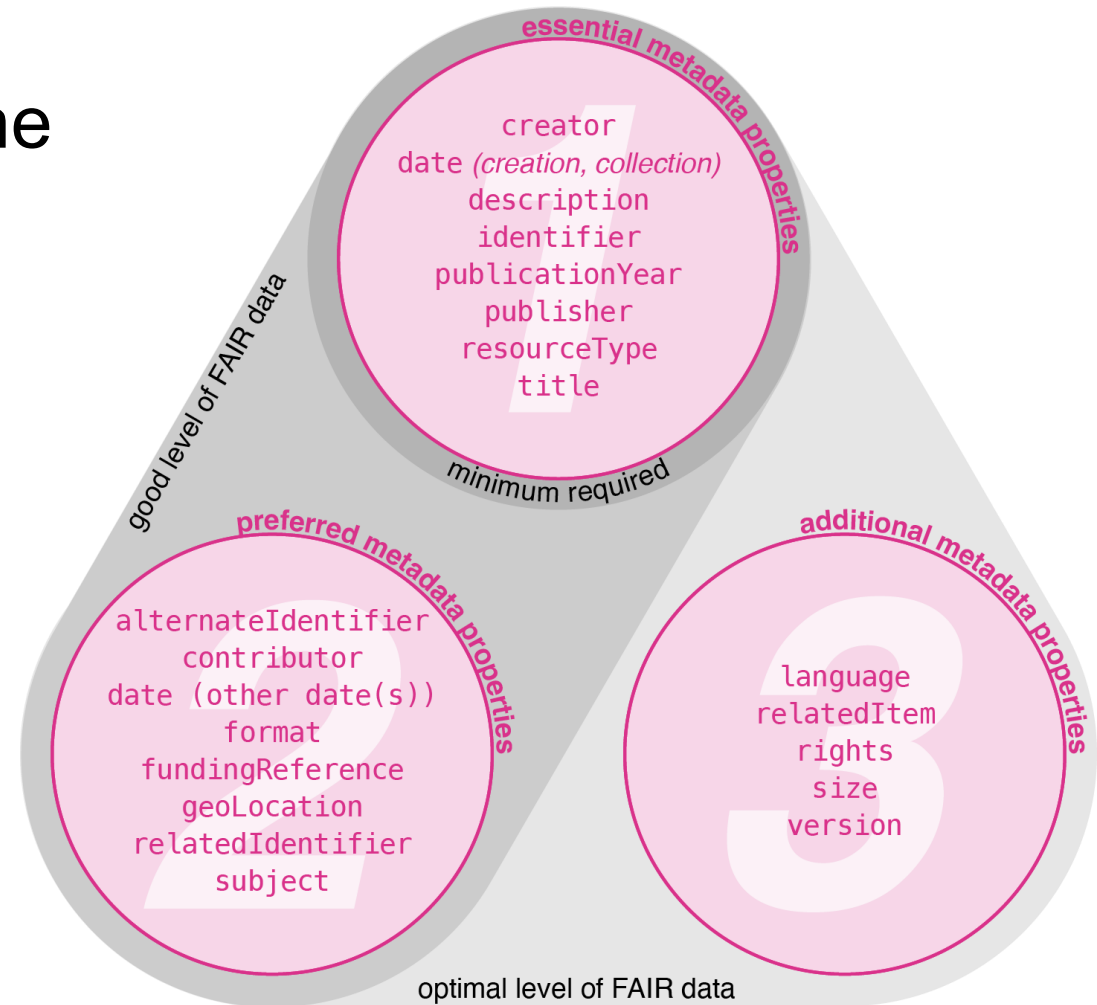
- best practices
- convergence in practices.

What does Microscopy Australia do?

Developing guidelines so metadata collection can be consistent across all the facilities:

- DataCite Metadata Schema
- minimum metadata requirement
- metadata publication by the data creator.

Engaging with the host universities is central.



Persistent Identifiers (PIDs): yes, let's do it!

Microscopy Australia supports the national PID strategy with the main focus on:

- PID for instruments (DOI)
- PID for organisations and facilities (ROR ID)
- background focus: ORCID ID (CReDiT) and RAiD.

The main barrier is metadata collection:

- what's the use of creating PIDs if they can't be used to enrich data or don't lead to increase discoverability of the identified resources?



Persistent Identifiers (PIDs): ... wait a minute

It isn't as easy as it seems...

- what is an instrument?
- what instrument needs to be identified uniquely?
- can institutional repositories support PIDs for instruments?
- have universities ever minted DOIs for instruments?
- do all university (data) librarians know about PIDs for instruments?
- who to ask for a ROR ID for our facilities?

It's a cultural change: some facilities are only realising that FAIR data are underpinned by PIDs.



Why persistent identifier for instruments?

host organisations:

- facility: ROR ID
- university: ROR ID

funding:

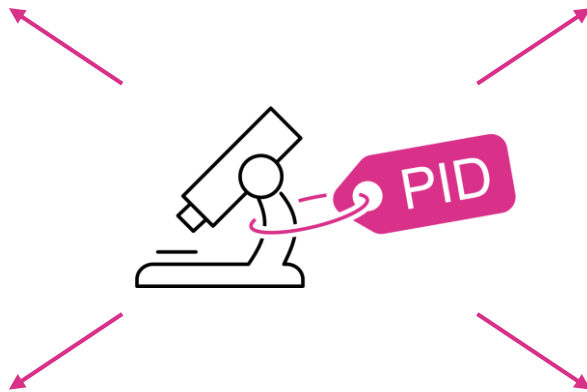
- grants (Grant ID)
- funders: ROR ID

standardisation of information/vocabularies

- instrument model and series names
- manufacturers' names
- technique names

reporting: link PIDs to the facility LIMS

- usage (booking)
- number of users
- volumes of data
- impact tracking





Microscopy Australia facilities and linked laboratories, and their host universities



Dr Rubbiya Ali
Dr Hoang Nguyen
Prof Roger Wepf
Prof David Abramson
Jake Carroll



Dr Emily Barker
A/Prof Martin Saunders



Dr Richard Webster
Dr Yin Yao
Dr Luc Betbeder-Matibet
Duncan Smith
Dongming Zheng



A/Prof Jamie Riches
Prof Matthew Bellgard
Craig Windell



Australian Research Data Commons

Siobhann McCafferty, Matthias Liffers, Max Wilkinson

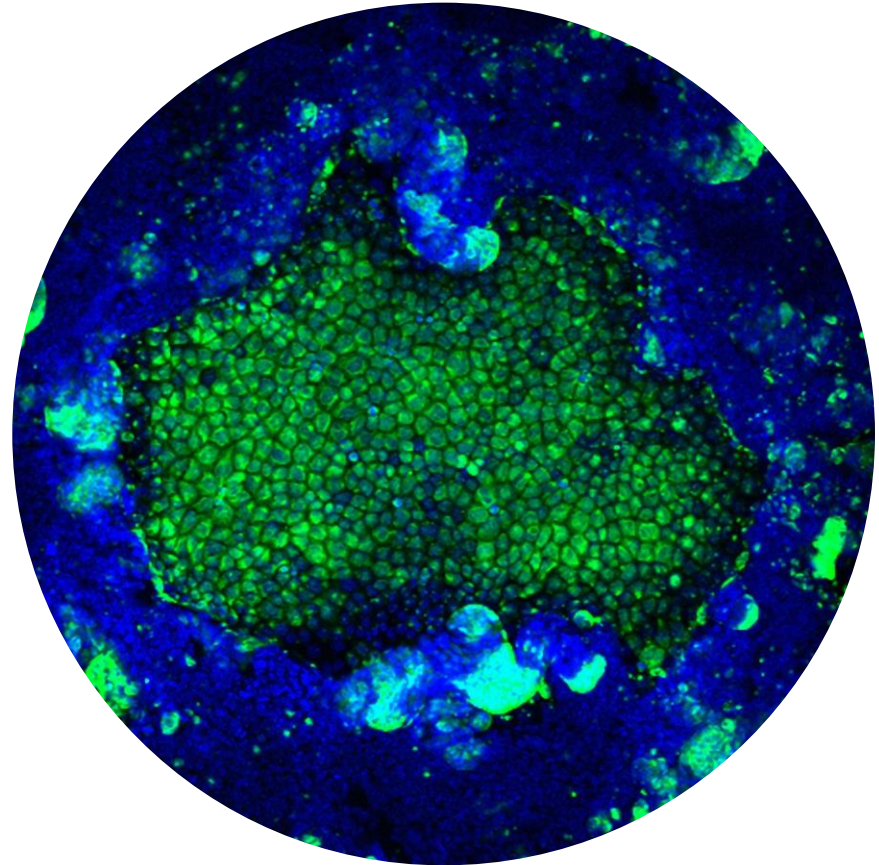
Australian Characterisation Commons at Scale (ACCS)

Various thematic groups: i4iOZ, Data Management Planning Interest Group, Queensland Characterisation, Global BioImaging Image Data Working Group



Thank you.

david.poger@sydney.edu.au



Further reading:

Poger, D (2023) Metadata collection at microscopy research facilities: Overview and recommendations. *Zenodo*. <https://doi.org/10.5281/zenodo.7949932>

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>