Title of the project:	**Developing an Australian-scale Characterisation Data Capture, Collection, and Collaboration Outline**
Lead organisation:	Monash University
Project leader:	Wojtek James Goscinski
Project leader details:	Email: wojtek.goscinski@monash.edu, Phone: 0412133156
Other organisations involved:	Members of the Australian Characterisation Informatics Committee, which includes: Microscopy Australia National Imaging Facility ANSTO University of Queensland University of Sydney University of Western Australia University of Wollongong Monash University MASSIVE

Approach

The project has undertaken a significant set of activities, with the goal of developing an Australian Characterisation Data Capture, Collection, and Collaboration Outline, that forms a strategic link between characterisation facilities, the standards and best practices they apply, and the research communities using instrument data to create domain-specific collections.

At the time of writing, the main outcomes of this project are:

- A national instrument survey, and an early analysis of results;
- A national Symposium and workshop with broad participation from national and international stakeholders.

The project will continue after the due date for this report, with the following activities:

- Complete survey data collection across a larger cohort of facilities and instruments;
- Compile the results of the national Symposium and workshop, form a writing group, and articulate national-scale goals as an Australian Characterisation Data Capture, Collection, and Collaboration Outline (and likely as part of an ARDC Platforms Request for Proposals response).

This section of this report addresses the items we set out to do in our project proposal.

- *Understanding Data Capture Requirements. NCRIS capabilities and institutions have developed multi-year multi-institutional instrument investment strategies. However, no long-term plan exists for the data, and quality research collections, these instruments will produce.*
    - *Activity: collate information about characterisation instruments, their data generating capabilities, their eResearch capability, and their surrounding research communities -- this information will be collated across existing Australian networks.*
    - *Activity: In-person survey research leaders to identify current and future collections.*

Working with National Imaging Facility (NIF), Microscopy Australia (MA), and local research facilities, we undertook a survey of instrument informatics requirements and opinions across Australia. We identified **399** characterisation instruments, across **29** facilities and **12** Universities. Of these, **113** were identified to be directly NCRIS-funded and a majority were identified to be available to researchers nationally through NIF or MA.

We developed an instrument survey for facility managers that addressed the following areas about their instruments:

- Basic information, such as modality, age of instrument, scientific application area,
- Metrics of usage and data generation capability,
- Data formats and data volumes,
- Informatics requirements,
- Data collection, publication and retention, and
- A range of qualitative questions that addressed informatics requirements and their suitability for these instruments and their user communities.

We worked through our facility networks to have the survey completed by as many instrument facilities as possible, within the limited timeframes of the project, and we received back

information across **111** instruments. We plan to continue collecting data after the project period.

We are working with *ARDC Discovery Activity: Bringing Long-Tail Microscopy and Characterisation Data into the Light, A.Mehnert*, which is undertaking a focused review of meta-data and data standards, to expand the responses to this survey and ensure that the two projects collect complementary data.

An early summary of the survey results and the insight collected is included in **Appendix A** to this report, and will be presented at two upcoming events:

- Data Sharing: Neuroscience, Microscopy and Experiments Collaborative Symposium on the 9th of October; and
- ARDC Data & Services Submit, on 21st Oct.

- *Developing and Implementing Data Collection Standards and Best-Practice. The Australian characterisation network provides researchers with a comprehensive suite of techniques. It is critical to develop consistency of standards and best-practises, to ensure data quality, so that researchers creating collections are able to collect across institutions, and determine correlations, spatial and functional, across modalities, so that data is more confidently shared and reused.*
    - *Activity: collation and review of the landscape of relevant data collection standards and best-practices, currently in use across Australia, and in development internationally.*

Our survey has identified a wide spectrum of data capture, handling, management, analysis and publication techniques that are being undertaken and infrastructure used across Australian characterisation instruments. This information is summarised in Appendix A, and includes publication to both international collections (e.g. AIBL) and publication repositories (e.g. Image Data Resource, PDB).

- *Underpinned by Australian and International Collaboration. Data collections formed using our instruments involve multiple owners, are generated by multiple instrument facilities and institutions. Collaboration is essential to enact best practise, and long term stewardship, leading to increased quality and opportunities for reuse and research collaboration.*
    - *Activity: invite international experts to present and form an international advisory group, as part of the Australian Characterisation Informatics Committee (ACIC).*

The project has worked with two complementary ARDC Data Discovery projects[1] to organise the Data Sharing: Neuroscience, Microscopy and Experiments Collaborative Symposium on the 9th of October. The purpose of this symposium is threefold:

- Engage international experts to present and contribute to a national discussion on the impact of data across Characterisation and Neuroscience;
- Provide a venue for open national discussions around future research software platform proposals, such as those led by the Australian Characterisation Informatics Committee and the Australian Brain Alliance, to build engagement and new collaborations;

[1] ARDC Data and Services Discovery Activities: Australian Brain Data Commons, and Australian Neuroscience Microscopy Data Sharing Platform

- Engage international experts as a first step toward forming an international advisory for future characterisation informatics projects.

Details of this event are available at https://www.massive.org.au/events/data-sharing-neuroscience-microscopy-and-experiments-collaborative-symposium/.

International guests at this event include: Prof Alan Evans, Director of the McGill Centre for Integrative Neuroscience, McGill University, Prof Jason Swedlow, Director of the Open Microscopy Environment, University of Dundee, AProf Kaylene Sympson, Senior Editor of Nature Scientific Data, Peter MacCallum Cancer Centre, Dr Shawn Brown, Associate Director for Research Software Development, McGill University, Prof Giorgio Ascoli, Center for Neural Informatics, Structures, and Plasticity, George Mason University, David Orloff, Director of the Cell Image Library, University of California at San Diego, AProf Claire Brown, Director of the Advanced BioImaging Facility, McGill University

Appendix B is the poster for this event.

FAIR

A FAIR assessment is attached as Appendix C.

However, this project itself has identified a number of points relevant to FAIR characterisation data:

- 51% of responding instruments do not provide data through a data management system, which means that data handover is a manual process. This proportion is likely higher as this survey is skewed toward Monash facilities, which are more likely to be integrated with a data management system (supported by the survey data itself). Integrating these instruments, to ensure data is captured with full meta-data and applying quality best-practise, is a priority.
- The survey has highlighted a number of international collections. We will seek to understand how data flows from experiment to these collections and help underpin these research workflows.
- The research data repositories used by the instruments in this survey (MyTardis, DaRIS, XNAT have been identified) provide varying degrees of support for FAIR data. Other projects have addressed this specifically (*Understanding and Creating Instrument Generated Data Collections, A. Hassan*)

This is a course study designed to complement finer grain work that specifically addresses metadata standards, and data formats. The future state is not addressed as this will be established through community consultation as part of the Symposium and future discussions.

Collaboration and coverage

Our project is at a national scale:

- The project itself is supported by the Australian Characterisation Informatics Committee, which is composed of national membership from across Australia.
- The survey has received responses from facilities at 9 Universities: UNSW, USydney, Western Sydney, UQ, UniSA, UMelbourne, Swinburn, UWA and The Florey.

- The Data Sharing Symposium is a collaboration between three ARDC Data Discovery Activities, and includes experts from USA, Canada and the UK. At the time of writing this report has received over 70 registrations.

Sustainability

The data produced by the responding instruments is, for the most part, an institutional concern. Forming a community of discussion about the importance of good data management and FAIR practise increases the overall attention paid to this challenge. Our survey illustrates that there are strong differences across sites, and across instrument modalities. We believe an organised national community will improve communication of solutions, and increase the availability of these solutions across Australian Universities. Likewise, federation and national engagement will enable new modalities and facilities to deploy better (and FAIRer) infrastructure.

The Symposium has seen a strong community response from across Australia, demonstrating that the need is broad and the interest is strong, which is important to future sustainability.

The data management platforms, and their underpinning storage, that our surveys have identified, were initiated or developed under earlier national funding, they are now sustainable infrastructure and are developed, operated and managed by institutions.

We believe that any future national scale improvements to this infrastructure will likewise receive institutional support and sustainability.

Learnings

The survey has identified a Summary of Early Insights that can be used to help prioritise future national characterisation informatics activities (see Appendix A).

More broadly, our survey can be used by infrastructure providers and facilities to infer:

- The largest data producing characterisation instruments, what amount of data they produce, what proportion of this data is retained and what proportion needs archiving and for what period. This is important to understand to best support scientific instrumentation and to date this information is largely developed on a case-by-case basis;
- The instruments where informatics environments are deemed least suitable and may require increased attention. Alternatively, the survey also reveals positive facility satisfaction and this information can be used to understand the tools and techniques that are having a positive impact on the scientific process.
- The instrument modalities that contribute data to international collections. We plan to use this information to ensure that data management systems and computing infrastructure are able to support all of the required steps up to the point of publication.

Impact

This proposal addresses the broad range of techniques, instruments, storage requirements, standards, and best-practises, that are critical to researchers creating data collections to address their research challenges, and moreover, to create data collections that are FAIR.

The challenges identified in our survey, and to be discussed at the Symposium, cannot be addressed through this proposal alone, and will require long-term commitment to ongoing support.

The activities proposed will allow the Australian Characterisation community to articulate:

- The scope of challenges, drawing on an analysis of current and future instruments and the communities that use them to generate data collections.
- A high-level multi-year outline for addressing these challenges, defining challenges and the way they can be addressed through coordinated programs. This will draw on existing initiatives and resources, already committed by the partners, as well as new funding opportunities.
- Estimate the level of resourcing required to meet these challenges, based on the advice of international experts and key stakeholders.
- Opportunities offered through international collaboration.

Report prepared by: Wojtek Goscinski
Date: 8/10/2019

Appendix A
Characterisation Survey Overview of Results
DRAFT at 4th October 2019

Characterisation Survey Overview of Results

DRAFT at 4th October 2019

The following is a summary of data and insights as captured on the 4th October. Data capture is not yet complete for this project and the project team will continue to update the information presented here.

The following are important notes to

- In some cases, a particular instrument modality is only represented by one respondent or facility. Therefore, the results may be skewed toward the specific infrastructure and capability available at the University or Facility. **A larger collection of responding instruments will provide more representative conclusions and we aim to continue collecting information to achieve this.**
- There are cases where information has only been gathered about a single instrument of a certain modality. Until a broader set of instruments is captured, the authors are avoiding making conclusions for modalities that are only represented by a single instrument. These instances are marked with #.
- The survey includes future planned instruments. Survey respondents have estimated or inferred requirements from their own expert knowledge.
- Respondents were not required to answer all questions, and therefore the total number of responding instruments varies across specific questions.

Anonymised survey responses and calculations for this report will be made available in the future.

Summary of Early Insights

Total projects	2505
Average no. per instrument	22.6
Total users	3007
Average no. per instrument	27.1

Data generation and collection (Section 5):

- The instruments produce approx half a petabyte of data per week, or 25 petabytes per year, which would significantly exceed the data storage capability of a single University.
- Eight modalities generate over 1TB per week(information collected over 36 instruments). However, the amount of cumulative data produced by these instruments is 95% of the total.
- There is a wide variation in data producing capabilities. For example, the transmission electron microscopes surveyed produce between 5GB and 8TB of data per week -- which can be attributed to the wide variety of techniques and specifically detectors.
- The most significant data producing instruments are: Transmission Electron Microscopy, Cryo Electron Microscopy, Lightsheet, Hybrid imaging -human.

Informatics environment (data capture, management and processing) available to the instrument and users of the instrument (Section 6):

- Hybrid imaging -human, Lightsheet, MR-human, X-ray Microscopy, Cryo Electron Microscopy, and Confocal have the most significant requirements[#].
- Focused Ion Beam Scanning Electron Microscopy, Optical CT, MR-preclinical, Transmission Electron Microscopy, Radiography and Absorptiometry, Scanning Electron Microscopy, and Lightsheet have been identified as the modalities where current informatics environments are the least suitable.

Data Collections (Section 7):

Survey questions were designed to identify how instruments contribute to the development of data collections. Specifically:

- Which instrument modalities were used to develop data collections for individual research groups, for broader collaboration, and for international publication; and
- Identify specific data collections that are being developed or contributed to by data generated by characterisation instruments.

As expected, the vast majority of instruments are used to create data collected and managed at the scale of small projects and PhD studies, and collected for projects across research groups and is shared across institutions.

A number of modalities were reported to contribute to a data collection that can be shared or used by others, or is published in a research repository or alongside a publication, with high (>50%) response from MR-human, Hybrid imaging -preclinical, MR-preclinical, Focused Ion Beam Scanning Electron Microscopy, Cryo Electron Microscopy.

Data Capture and Management (Section 8):

- 42% of instruments reported feed data into a data management system which is capable of providing global data identifiers.
- Cryo Electron Microscopy, Confocal, Lightsheet, Super-resolution, Widefield, and MR-human modalities are the most likely to provide users with a data management system.
- Hybrid imaging -preclinical, Transmission Electron Microscopy, Scanning Electron Microscopy, X-ray Computed Tomography, and Ultrasound, are the least likely to feed data into a data management solution[#].

Summary of Data

1. Identified instruments:

Total instruments identified: 399

Facilities identified: 29, across Universities: 12

2. Summary of Instruments in Survey Response

Instruments responding: 111

Percent of total identified: 28%

Modalities represented:

				MR-human	7
Imaging	29	Human Focus	9	Hybrid imaging -human	1

				Magnetoencephalography	1
				MR-preclinical	5
		Pre-clinical and Animal Focus	20	Hybrid imaging -preclinical	8
				Ultrasound	3
				Laser Speckle Contrast Imaging	1
				Mass Spectrometry Imaging	1
				Optical CT	1
				Optoacoustic	1
Microscopy and MicroImaging	73	Electron Microscopy Techniques	13	Transmission Electron Microscopy	8
				Scanning Electron Microscopy	3
				Focused Ion Beam Scanning Electron Microscopy	2
		Light Microscopy Techniques	59	Cryo Electron Microscopy	5
				Confocal	20
				Super-resolution	5
				Widefield	24
				Lightsheet	5
		X-ray Microscopy Techniques	1	X-ray Microscopy	1
Other	2			Other	2

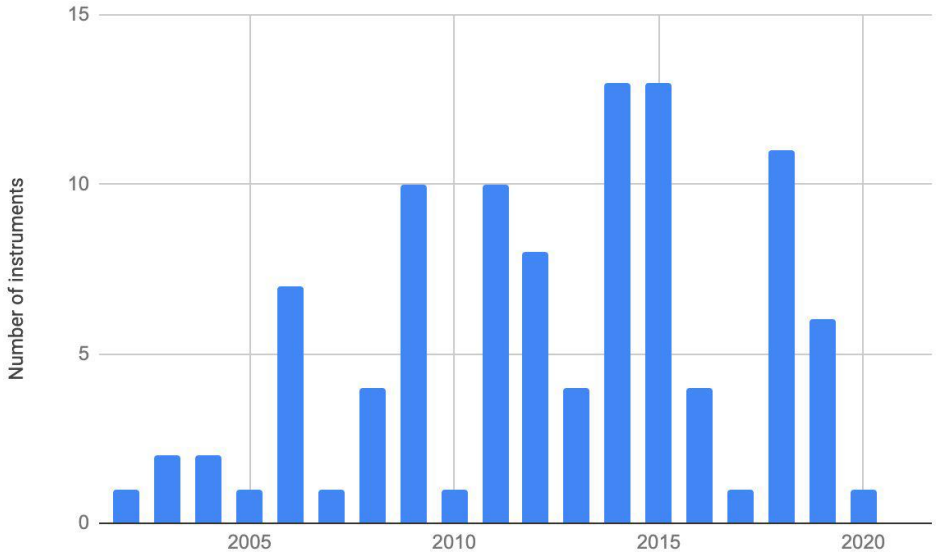


Figure 1. Age of reported instruments.

3. Projects and Users

The following gives an overview of the total number of projects and users supported by the instruments reported.

Estimated number of projects per year	Total projects	2505
	Averge no. per instrument	22.6
Estimated number of users per year	Total users	3007
	Average no. per instrument	27.1

Note: The figures should be used carefully as there is very likely overlap between users reported for each instrument at an individual facility, and likely some overlap at a national scale.

4. File Formats

The following provides an overview of the type of data collected. *ARDC Discovery Activity: Bringing Long-Tail Microscopy and Characterisation Data into the Light, A.Mehnert* is collecting this information in more detail and depth.

TIF	45	DM3	4	LOG	1
DICOM	25	ND2	3	MMF	1
JPG	11	TXT	3	NII	1
LIF	9	HDR	2	SER	1
BMP	8	BIMG	1	TXRM	1
MRC	5	IMG	1	XLS	1
RAW	5	LM	1	XML (imzML)	1

5. Data Volumes Generated

Data volume produced per session or experiment	Total data for all instruments	33,294.5 GB
	Average amount per instrument	300.0 GB
Data volume produced (est. per week)	Total data for all instruments	484,203.2 GB
	Average amount per instrument	4,362.2 GB

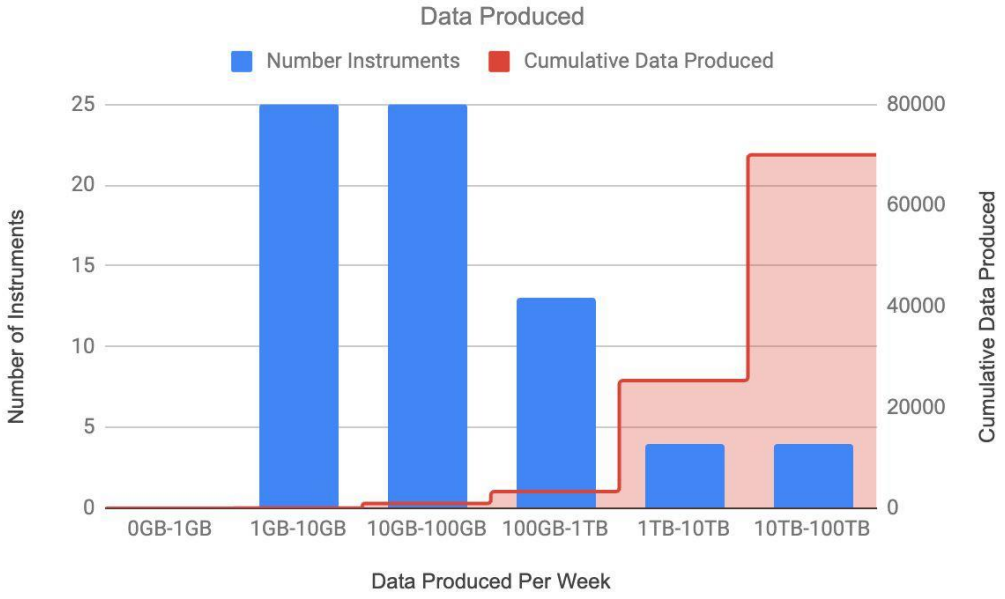


Figure 2. Chart showing the number of instruments at various data producing quantities, and their cumilative data producing capabilities.

Modality	Average estimated no. projects per year	Average estimated number of users per year	Average data volume produced per session or experiment (GB)	Average data volume produced (estimated) per week (GB)
Confocal	38.60	38.60	2.42	22.72
Cryo Electron Microscopy	30.00	18.50	2,050.00	4,875.00
Focused Ion Beam Scanning Electron Microscopy	59.00	60.50	7.50	35.00
Hybrid imaging -human	40.00	100.00	50.00	200.00
Hybrid imaging -preclinical	7.80	6.60	22.16	110.03
Laser Speckle Contrast Imaging	1.00	2.00	150.00	50.00
Lightsheet	8.50	8.50	1,080.15	3,176.55
Magnetoencephalography (MEG)	20.00	26.00	10.00	110.00
Mass Spectrometry Imaging	15.00	30.00	0.50	1.00
MR-human	35.50	126.83	18.83	120.33
MR-preclinical	10.40	13.00	12.02	75.00
Optical CT	40.00	20.00	0.02	0.10
Optoacoustic	0.00	0.00	50.00	0.00
Other	50.00	20.00	0.00	0.00
Radiography and Absorptiometry	3.33	3.33	0.67	3.38

Scanning Electron Microscopy	100.33	105.00	1.67	6.33
Super-resolution	14.00	14.00	0.20	0.30
Transmission Electron Microscopy	105.29	94.57	2,750.75	6,927.63
Ultrasound	4.00	5.33	25.03	10.25
Widefield	21.09	21.09	11.17	55.46
X-ray Computed Tomography	17.00	20.75	23.13	138.75
X-ray Microscopy	20.00	20.00	50.00	50.00

6. Informatics Environment

How mature is the informatics environment ?	Mature or stable	16
	Evolving slowly	28
	Evolving very quickly	16
	Unknown	2
	No answer provided	49

What level of informatics resourcing does this instrument require ?	Heavy	28
	Medium	17
	Light	14
	None	3
	No answer provided	49

What level of informatics resourcing does this instrument require vs modality cross analysis

Modality vs Informatics Resourcing	Respondent Instruments	Heavy	Medium	Light	None
Hybrid imaging -human	1	100%	0%	0%	0%
Lightsheet	2	100%	0%	0%	0%
Magnetoencephalography (MEG)	1	100%	0%	0%	0%
MR-human	7	100%	0%	0%	0%
X-ray Microscopy	1	100%	0%	0%	0%
Cryo Electron Microscopy	5	80%	20%	0%	0%
Confocal	2	50%	50%	0%	0%
Focused Ion Beam Scanning Electron Microscopy	2	50%	50%	0%	0%
X-ray Computed Tomography	4	50%	25%	25%	0%
MR-preclinical	5	60%	0%	20%	20%
Transmission Electron Microscopy	8	38%	25%	38%	0%

Hybrid imaging -preclinical	8	25%	50%	25%	0%
Mass Spectrometry Imaging	1	0%	100%	0%	0%
Optical CT	1	0%	100%	0%	0%
Optoacoustic	1	0%	100%	0%	0%
Widefield	2	0%	100%	0%	0%
Scanning Electron Microscopy	3	0%	33%	67%	0%
Laser Speckle Contrast Imaging	1	0%	0%	100%	0%
Radiography and Absorptiometry	3	0%	0%	100%	0%
Ultrasound	3	0%	33%	33%	33%
Other	1	0%	0%	0%	100%
Super-resolution	0	No data provided			

How would you classify the informatics environment that your researchers have access to for managing or processing data from this instrument?

Highly suitable or capable	12
Suitable or capable	28
Somewhat unsuitable or incapable	21
Inadequate	2
No answer provided	48

How would you classify the informatics environment that your researchers have access to for managing or processing data from this instrument vs modality cross analysis

Modality vs Informatics Environment	Respondent Instruments	Highly suitable or capable	Suitable or capable	Somewhat unsuitable or incapable	Inadequate
Focused Ion Beam Scanning Electron Microscopy	2	0%	0%	100%	0%
Optical CT	1	0%	0%	100%	0%
MR-preclinical	5	40%	20%	0%	40%
Transmission Electron Microscopy	8	0%	25%	75%	0%
Radiography and Absorptiometry	3	0%	33%	67%	0%
Scanning Electron Microscopy	3	0%	33%	67%	0%
Lightsheet	2	0%	50%	50%	0%
Confocal	3	0%	67%	33%	0%
Ultrasound	3	0%	67%	33%	0%
MR-human	7	14%	43%	43%	0%
Cryo Electron Microscopy	5	20%	40%	40%	0%
Magnetoencephalography (MEG)	1	0%	100%	0%	0%
Mass Spectrometry Imaging	1	0%	100%	0%	0%
Optoacoustic	1	0%	100%	0%	0%

Widefield	2	0%	100%	0%	0%
X-ray Microscopy	1	0%	100%	0%	0%
Hybrid imaging -preclinical	8	38%	63%	0%	0%
X-ray Computed Tomography	4	50%	50%	0%	0%
Hybrid imaging -human	1	100%	0%	0%	0%
Laser Speckle Contrast Imaging	1	100%	0%	0%	0%
Other	1	100%	0%	0%	0%
Super-resolution	0	No data provided			

7. Data Collections

Data from this instrument is sometimes or more often...

... collected and managed at the scale of small projects and PhD studies	77
... collected for projects across research groups and is shared across institutions	41
... collected for the purposes of creating a data collection that can be shared or used by others, or is published in a research repository or alongside a publication	31

The following collections were specifically listed by respondents:

Imaging collections:

- Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing (AIBL)
- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Federal Interagency Traumatic Brain Injury Research (FITBIR)
- Australian Dementia Network (ADNeT)
- Human brain healthy control database - NIF
- OpenNeuro

Protein Structure Collections

- Electron Microscopy Data Bank (EMDB)
- "Protein Data Bank (PDB)

What instruments are used to create data collections?

Modality vs Collections	Number of responding instruments	... collected and managed at the scale of small projects and PhD studies (% of responding instruments)	... collected for projects across research groups and is shared across institutions (% of responding instruments)	... collected for the purposes of creating a data collection that can be shared or used by others, or is published in a research repository or alongside a publication (% of responding instruments)
Hybrid imaging -human	1	100%	100%	100%

Mass Spectrometry Imaging	1	100%	100%	100%
X-ray Microscopy	1	100%	100%	100%
MR-human	7	100%	100%	71%
Hybrid imaging -preclinical	8	100%	63%	63%
MR-preclinical	5	100%	20%	60%
Focused Ion Beam Scanning Electron Microscopy	2	100%	100%	50%
Cryo Electron Microscopy	5	60%	80%	40%
Lightsheet	5	40%	40%	40%
Transmission Electron Microscopy	8	100%	100%	38%
Scanning Electron Microscopy	3	100%	100%	33%
Ultrasound	3	100%	0%	33%
X-ray Computed Tomography	4	100%	25%	25%
Confocal	20	50%	10%	10%
Widefield	24	42%	8%	8%
Laser Speckle Contrast Imaging	1	100%	0%	0%
Magnetoencephalography (MEG)	1	100%	100%	0%
Optical CT	1	100%	0%	0%
Optoacoustic	1	100%	0%	0%
Other	2	50%	0%	0%
Radiography and Absorptiometry	3	100%	0%	0%
Super-resolution	5	20%	0%	0%

8. Data Retention, Management and Analysis

Please estimate what percentage of data generated by this instrument is retained beyond 1 week after the initial experiment?	0 %	12	instruments
	10 %	2	instruments
	99%	1	instruments
	100%	68	instruments

		Retain for analysis (days)
Of the data that is retained please estimate, as best you are able, on average, how long researchers require to analyse this data? (e.g. 1 month, 1 year, 2 years etc)	Avg minimum range	195
	Avg maximum range	390

	Minimum	0
	Maximum	2555

		Retain in archive (days)
Please estimate, as best you are able, on average, how long researchers retain data generated by this instrument in archive?	Average	646
	Minimum	24
	Maximum	12000

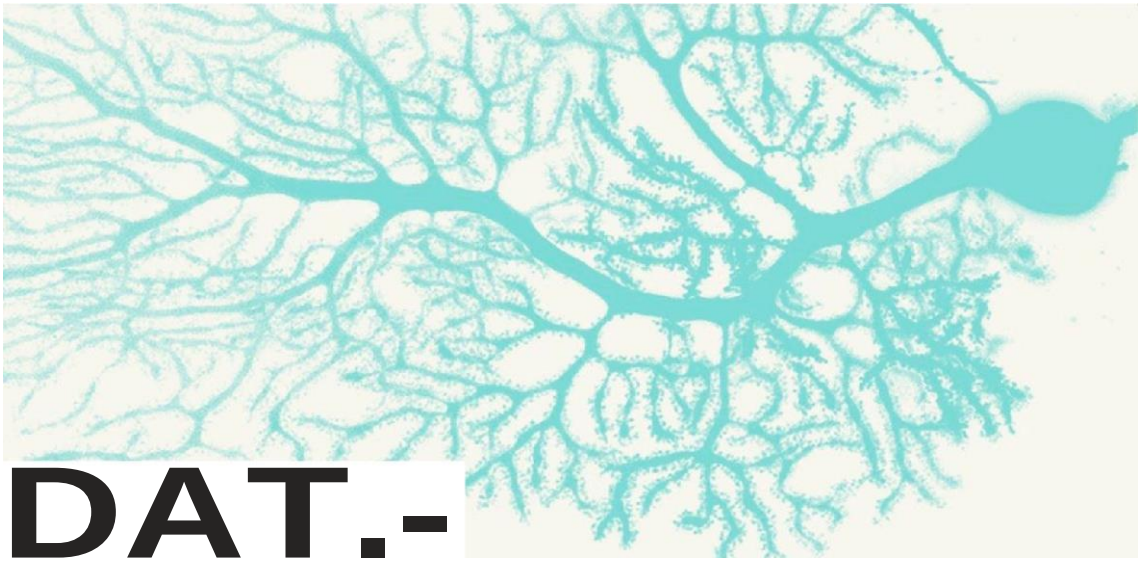
Data Management solution vs modality cross analysis

	Number responding instruments	Data leaves with the user	Central storage	Data manage-ment system	Data manageme-nt system, with identifiers	No answer
Cryo Electron Microscopy	5	0%	0%	0%	100%	0%
Confocal	20	0%	0%	0%	70%	30%
Lightsheet	5	0%	20%	0%	60%	20%
Super-resolution	5	0%	0%	0%	60%	40%
Widefield	24	0%	0%	0%	58%	42%
MR-human	7	14%	0%	29%	43%	14%
Focused Ion Beam Scanning Electron Microscopy	2	50%	0%	0%	50%	0%
Hybrid imaging -human	1	0%	0%	100%	0%	0%
Hybrid imaging -preclinical	8	25%	25%	13%	38%	0%
Transmission Electron Microscopy	8	63%	0%	0%	38%	0%
Scanning Electron Microscopy	3	67%	0%	0%	33%	0%
Laser Speckle Contrast Imaging	1	0%	100%	0%	0%	0%
Optoacoustic	1	0%	100%	0%	0%	0%
X-ray Computed Tomography	4	50%	50%	0%	0%	0%
Ultrasound	3	33%	33%	0%	0%	33%
MR-preclinical	5	80%	20%	0%	0%	0%
Magnetoencephalography (MEG)	1	0%	0%	0%	0%	100%
Mass Spectrometry Imaging	1	100%	0%	0%	0%	0%
Optical CT	1	100%	0%	0%	0%	0%
Other	2	0%	0%	0%	0%	100%
Radiography and Absorptiometry	3	100%	0%	0%	0%	0%
X-ray Microscopy	1	100%	0%	0%	0%	0%

9. Confidence of Respondent

How confident are you in your answers ?	highly confident	21
	moderately confident	25
	low confidence	1
	not confident	0

Appendix B
Data Sharing Symposium Poster



DAT.-SHARING:

Neuroscience, Microscopy and Experiments Collaborative Symposium

Speakers and Invited Guests

A combined symposium to plan how we enable interoperability and collaboration between researchers in a data-driven world.

Prof Alan Evans
Direct0< of rl7e M cGill Centre fo, Imegrnrive NWrOS.d9 nC.9 , Mc.Gm Unfve~/$1ty

Prof Jason Swedlow
Direc f0t of the Open Microscopy Environment , Un rirs !ty of D un

AProf Kaylene Sympson
S6nior Editor of Narure *Scientific* Dara.
Peter MacCa/AJm *Ctrncer Cenlre*

Dr Shawn Brown
Associ;,ite *DirfICtor* for *ReS68FGh SoffWBr6' Development, McG1/f Vnf,.-er$ily*

Prof Giorgio Ascoli
Cfinf9r JQ/' N,e1.1n! Inf orm:tr lc:s, Structu s. ơŋ0 Pillstl clry, *Go.orge MBSOIl Universiry*

David Orlof f
Director of the Cell *Ima ge Ubra,y, University or Ca!J'lorn/a at* Sa ri *Diego*

AProf Claire Brown
Direc tor of the *A r:Jv3 n cl EJiolm.99i119F* acifity, *Mci3,II University*

Event Workshops

Australian Characterisation Informatics Planning Wor kshop

Ch.iracta, l$ation hes become an ,are-a v1here 1,11 o rrnmlcs In I rastructu re. Al(pArtlse end b&s t practice is c,ltica t. The Australian Charact!ɪ isation con-un un ity $r'18r f:'$ *e*,MtJl'n t:if:'r o1 $ig niti ¢8nt Info ,ma tfcs cha lleriges . and ttie community Is working together to plar1arx:I Im p[emen t strat ies toovercome these cha!len ij1E!S. Thi!5work15hop will bring (Qge h1:1r stralitrn :oim;! 1n1em, ul orHl1 @-X r1$ TD rY!8r, tr!A sta 1teg le lin ks btitwoori ch:aracte ris at ioo faciliti es. in formatics 5tande rd5 .an d tle5 t pract ice s, an d research comrnunitles, with Ihe OVl'.?ra D goa l o i p la nn i ng lor tt,o fulure.

Aus tralian Neuroscience Microscopy Data Sharing, Standards and Best-Practice Workshop

Rec,er t ao va, r ces i n o pti c$, com pu t:ttio n o!l : ll Is ond outomatlori tiove lner@osed t ti ,e Impac t of modern optica l micro5COpy not oo ly ill basic $Cie-nee t:iut aiso iii m JStry <'fn(I t1 ink: .er1 doma i ns . The Ausl ratl::mop tica l rri c rosco py community is actively work.i ng to gether to torm B d ynamic network. of re searchers wtlO wil eo:oib1e d<11tri5h<11ring'<ind:5,hind <1rd i5allioo, The Ail"r'l of tl'lls In itiat ive i Š lo Ine e&Sl! rt)E>\JI,ality and reproducibili tr. and boost the overall Yalue of op ti cal m icroscopy data . T hl 5 5ymp o5ium will discuss ttie curmn1 ctia Jenges ot ca t.a harino aOO protoco ls with the aim of mapping out a M tlonal strategy to integ r.ste c u rre nt syste ms Into :i vnilied Austf -.illa n data :r.h a,ing p la t10rm.

Aust ralian Brain Data Commons Workshop

By its Y'=ff'lf nat tH , trle broiin ls ar 11-e (O r-e 0,1 ,M wa arll and crackIng hs "codA• (h ow it works) Is one of the greal e8l scieritific chall!!ges racing hum& ity. To crttck th.e bra!m !1- eOde- neuro "ic ience h as become one of the most tilghlyinterdisclplinat)' scie n t if ic res earch dis c iplines. U nders t and ing the brai n' s code - the nernal circuits and proce&Se1> u nd e-rl ying co gllition - i5 cuuentlr r:it 1ne 1 1ro n1 of brain reseai(:h ruitlM allyana internationally. Central to cn:icking the code will be a oahonal Strategy tor brBJn r ese arch data shari ng and curaUon. This wa rk stlop will bring together brain rcsca rchc:rsand ncurolnfOfm atician s to map the nati on al bm In data commons e.,ir, a,t!il ties . 11o1Yi"1.vl t1e comm un iry u lrame,r11s, end develop a nation al plan for nAUrosc le nce data sha ring arx:I cura t ioo,

9th October 2019

h1tp s.:1/b i1.fyf21m7 dW
Rttglstra tio n sclos8
CO B Monday 7tt,October

Schedule
ReQistm t io n from 8:30am
Openi ng ar'ld inv it ed ss:.,eake r'S I rOrn 9am
Work.shop5 trom1pm
Combined dlscussronand soclal from 4pm

Orgaa is| og| CoromLtt .
P,ro Ll.l'\d.a i ch ard 5
Prol Tony H arm an
DrNclaDu rl slc
Dr Rumelo Amor
Dr Khaled Ctiakli
Dr ull::I Mdr ire t Villeg i'lS
Dr Wo jto l<. Jamas G osc1nsl<.I

The Shine Dome Canberra

A ;Ji:> C M ONASH THE UNIVERSITY OF WESTERN AUSTRALIA W'i;lli:-, ll @REY M_ MACQUARIE University Walter+Eliza Hall @=-= TRIO A t tl?l9 MASSIVE

Appendix C
FAIR Assessment

	FAIR self assessment for project:			*Please delete Example answers as you fill out the assessment for your own data*		
	Completed ../../2019					
	Questions for each FAIR component ↓	**Answer options: Increasingly FAIR -->**				
	FINDABLE					
Q1	Does the dataset have any identifiers assigned?	No identifier	Local identifier	Web address (URL)	Globally unique, citable and persistent identifier (e.g. DOI, PURL, or Handle)	
A1	Start of project	**51% of responding instruments do not provide data through a data management system and likely support no Identifiers**		**42% of responding instruments feed data into a data management system that supports unique identifiers and is capable of supporting gloabl identifers**		
	End of project					
	Two years time			**To be established through consulation. A goal should be to ensure 100% of characterisation data is ingested into data management systems capable of global identifiers.**		
Q2	Is the identifier included in all metadata records or metadata files describing the data?	No	Yes			
A2	Start of project		**Yes, for 42% as above**			
	End of project					
	Two years time		**To be established through consulation**			
Q3	Is the data described by a metadata record?	The data is not described	Brief title and description	Brief title and description, and multiple other fields filled out, albeit briefly.	Comprehensively (a min metadata template will be provided) using a formal machine-readable metadata schema.	
A3	Start of project	**Likely true for remaining 58% of responding instruments**			**Yes, capability exists for 42% of responding instruments**	
	End of project					
	Two years time				**To be established through consulation**	
Q4	What type of repository or registry is the metadata record in?	The data is not described in any registry or repository	Local institutional repository	Domain-specific repository	Generalist public repository	Data is in one place but discoverable through several places (i.e. other registries, RDA, Google Data Search)

A4	Start of project	**51% of respondent instruments**	**Significant ingestion into local repositories that are capable of global identifiers. These repositories are both generalist and domain-specific.**			**Data identified to be flowing to:** **Imaging collections:** *** Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing (AIBL)** **Alzheimer's Disease Neuroimaging Initiative (ADNI)** *** Federal Interagency Traumatic Brain Injury Research (FITBIR)** *** Australian Dementia Network (ADNeT)** *** Human brain healthy control database - NIF** *** OpenNeuro** **Protein Structure Collections:** *** Electron Microscopy Data Bank (EMDB)** *** Protein Data Bank (PDB)**
	End of project					
	Two years time		**To be established through consulation**			
	ACCESSIBLE					
Q5	How accessible is the data? Note: The access method (s) must be explicitly stated in the metadata record, e. g. if any authentication is needed, or there are any restrictions to access.	No metadata record	Access to metadata only	Unspecified access conditions e.g. "contact the data custodian to discuss access"	Embargoed access after a specified date; or A deidentified version of the data is publicly accessible	Fully accessible public, or to persons who meet and follow explicitly stated conditions and processes, e.g. ethics approval for sensitive data
A5	Start of project	**As the study is specific to experiment data or raw data, the vast majority of data identified does not provide public access to meta-data, or public access to data.**	**Specific to specific collections, as identified in A4.**			
	End of project					
	Two years time		**To be established through consulation**			
Q6	Is the data available online without requiring specialised protocols or tools once access has been approved?	No access to data	By individual arrangement	File download from online location	Non-standard web service (e.g. OpenAPI/Swagger/informal API)	Standard web service API (e.g. OGC)
A6	Start of project			**True for the 49% of responding instruments feeding data directly to a data management system**		
	End of project					
	Two years time			**To be established through consulation**		
Q7	Does the repository/registry agree to maintain the persistence of the metadata record, even if the data product is no longer available?	No (or not applicable, if no metadata record exists)	Unsure	Yes		
A7	Start of project		**Unsure**			
	End of project					
	Two years time			**To be established through consulation**		
	INTEROPERABLE					
Q8	Are the data available in (an) open (file) format(s)?	Data are mostly available only in a proprietary format	Data are available in an open format	Data are available in an open, documented, widely-used standard format (i.e. NetCDF, CSV, JSON, XML, etc)		

A8	Start of project	**Some proportion of instruments still generate proprietary data formats. Further work with ARDC Discovery Activity: Bringing Long-Tail Microscopy and Characterisation Data into the Light, A.Mehnert**				
	End of project					
	Two years time	**To be established through consulation**				
Q9	Are the data machine readable?	The data are unstructured	The data are structured and machine-readable (i.e. csv, JSON, XML, RDF, database files, etc)			
A9	Start of project		**100%**			
	End of project					
	Two years time		**100%**			
Q10	What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?	Data elements are not described (i.e. fields or objects are labelled with codes or not at all)	Data elements are described (so that a human user can correctly interpret the data), but no standards have been used in the description	Recognised standards have been used in the description of data elements, but no published vocabularies with resolvable URIs are used	Published vocabularies using resolvable identifiers linking to explanations are used, so that the data can be read and understood by machines as well as humans.	Published vocabularies using **persistent** resolvable identifiers linking to explanations are used, so that the data can be read and understood by machines as well as humans.
A10	Start of project	**This project did not address this question. However, ARDC Discovery Activity: Bringing Long-Tail Microscopy and Characterisation Data into the Light, A.Mehnert will likely address this in more detail. This is expected to be a major component of future work across the Australian Charactersation community.**				
	End of project					
	Two years time	**To be established through consulation**				
Q11	How is the relationship to other data and resources (e.g. related datasets, services, publications, etc) described in the metadata, to provide context around the data?	There are no links to other metadata or data	The metadata record includes URI links to related metadata, data and definitions	Qualified links to other resources are recorded in a machine readable format, e.g. a linked data format such as RDF		
A11	Start of project	**This project did not address this question. However, ARDC Discovery Activity: Bringing Long-Tail Microscopy and Characterisation Data into the Light, A.Mehnert will likely address this in more detail. This is expected to be a major component of future work across the Australian Charactersation community.**				
	End of project					
	Two years time	**To be established through consulation**				
	REUSABLE					
Q12	Which of the following best describes the license (usage rights) attached to the data?	No license is applied	Non-standard license applied, without a license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Non-standard license applied, WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), without a license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record
A12	Start of project	**Not addressed under this study**				
	End of project					
	Two years time	**To be established through consulation**				
Q13	How much provenance information has been captured to facilitate data reuse? i.e. project objectives, data generation/collection (including from external sources) and processing workflows.	No provenance information is recorded	Partially recorded	Comprehensively recorded in a text format (i.e. TXT or PDF)	Comprehensively recorded in a machine readable format (i.e. in metadata record's schema or PROV, or in RDF, JSON, NetCDF, XML, etc)	
A13	Start of project	**Not addressed under this study**				
	End of project					
	Two years time	**To be established through consulation**				