

# Developing an Australian-scale Characterisation Data Capture, Collection, and Collaboration Outline

Wojtek Goscinski, Jay van Shyndel, Paula Martinez, et al

Australian Characterisation Informatics Committee: Microscopy Australia, National Imaging Facility, ANSTO, University of Queensland, University of Sydney, University of Western Australia, University of Wollongong, Monash University, MASSIVE



Australian Research Data Commons

This research/project is supported by the Australian Research Data Commons (ARDC) The ARDC is enabled by NCRIS

## **Key Issues and Our Study**

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.

**What's the scale of the data problem?**

**How do facilities rate the informatics environment that is available to them?**

**What instruments feed into international collections?**

**What is international best practice?**

1. Collate information about characterisation instruments, their data generating capabilities, their eResearch capability, and their surrounding research communities.
2. Collect and review of the landscape of relevant data collection standards and best-practices, currently in use across Australia, and in development internationally.
3. Invite international experts to present and advise Australian projects.

The project has worked with two complementary ARDC Data Discovery projects to organise the Data Sharing: Neuroscience, Microscopy and Experiments Collaborative Symposium on the 9th of October.

# Characterisation Modalities

Imaging	29	Human Focus	9	MR-human	7
				Hybrid imaging -human	1
				Magnetoencephalography	1
		Pre-clinical and Animal Focus	20	MR-preclinical	5
				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
		X-ray Microscopy Techniques	1	1	X-ray Microscopy
Other	2			Other	2

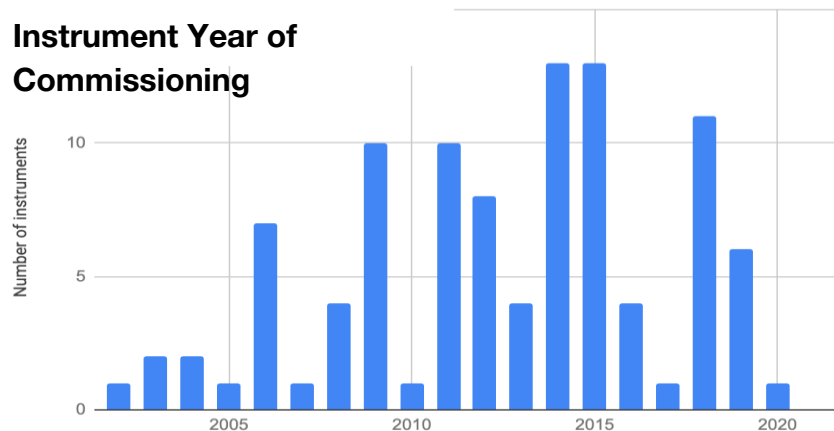
# Characterisation Instrument Survey Overview (8th Oct - information already out of date)

**Responding**  
**111** Instruments  
**17** Facilities  
**9** Universities  
 and Institutes

**Identified**  
**399** Instruments  
**29** Facilities  
**9** Universities and  
 Institutes

<b>Estimated number of projects per year</b>	Total research projects	<b>2505</b>
	Average no. per instrument	<b>22.6</b>
<b>Estimated number of users per year</b>	Total users	<b>3007</b>
	Average no. per instrument	<b>27.1</b>

## Instrument Year of Commissioning



Classified into **22 characterisation modalities**

**25** petabytes of data per year produced by responding instruments

**Eight modalities produce over 1TB per week**

However, the amount of cumulative data produced by these instruments is 95% of the total.

The most significant data producing instruments are:  
**Transmission Electron Microscopy, Cryo Electron Microscopy, Lightsheet, Hybrid imaging -human**

**42%** of instruments reported feed data into a data management system which is capable of providing global data identifiers.

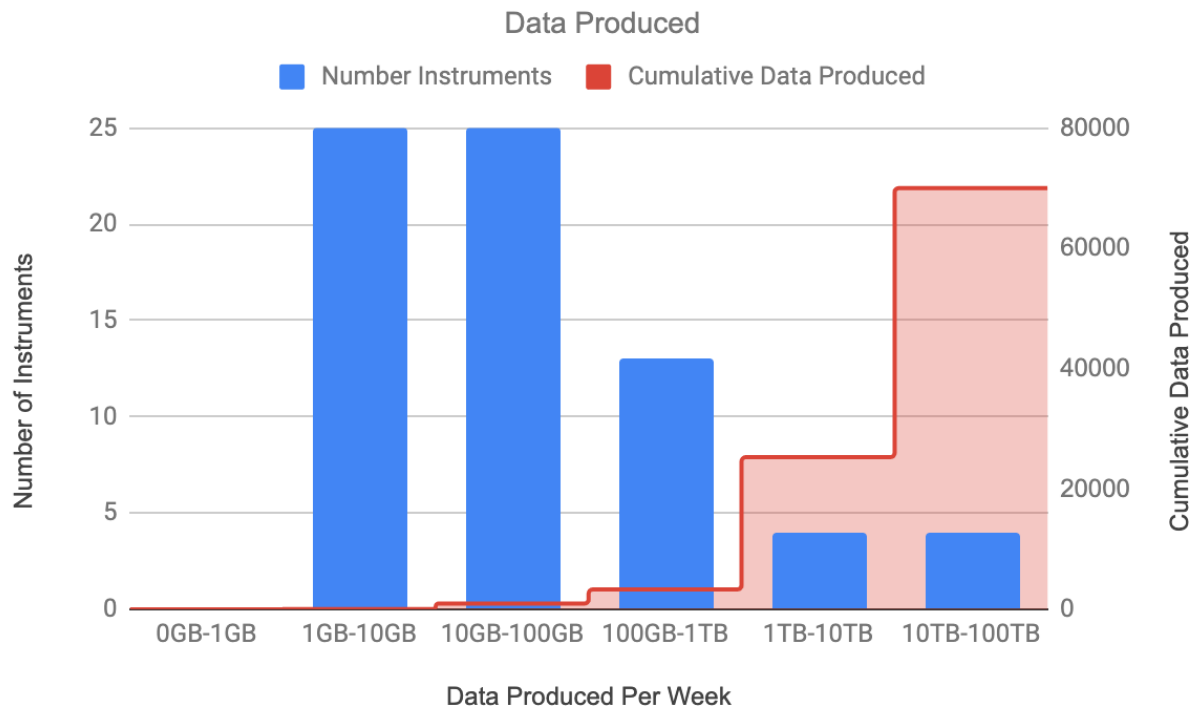
On average users required **between ½ year to 1 year to process their data**

## Data Formats

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

# Lessons Learnt: Data

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
<b>Cryo Electron Microscopy</b>	<b>30.00</b>	<b>18.50</b>	<b>2.050.00</b>	<b>4.875.00</b>
Focused Ion E Microscopy				
<b>Hybrid imagii</b>				
<b>Hybrid imagii</b>				
Laser Speckle				
<b>Lightsheet</b>				
<b>Magnetoenc</b>				
Mass Spectro				
<b>MR-human</b>				
MR-preclinica				
Optical CT				
Optoacoustic				
Other				
Radiography :				
Scanning Elec				
Super-resoluti				
<b>Transmissio</b>				
Ultrasound				
Widefield				
<b>X-ray Computed Tomography</b>	<b>17.00</b>	<b>20.75</b>	<b>23.13</b>	<b>138.75</b>



## Lessons Learnt: Collections

Data from this instrument is sometimes or more often...	
... collected and managed at the scale of small projects and PhD studies	<b>77</b>
... collected for projects across research groups and is shared across institutions	<b>41</b>
... 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	<b>31</b>

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 \(PDB\)](#)**

# Lessons Learnt: Informatics Environments

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	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%

# Lessons Learnt: Data Management

Data Management solution vs modality cross analysis								
		Modality	Number responding instruments	Data leaves with the user	Central storage	Data management system	Data management 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%





## Acknowledgements

Monash eResearch Centre, Monash University

Microscopy Australia

National Imaging Facility

ANSTO

University of Queensland

University of Sydney

University of Western Australia

University of Wollongong

Monash University

MASSIVE

**17 characterisation facilities respondents across Australia**



Australian Research Data Commons

This research/project is supported by the Australian Research Data Commons (ARDC) The ARDC is enabled by NCRIS