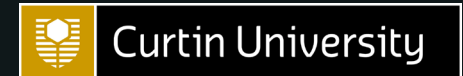


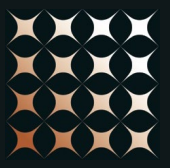


MWA Data & Archive Update

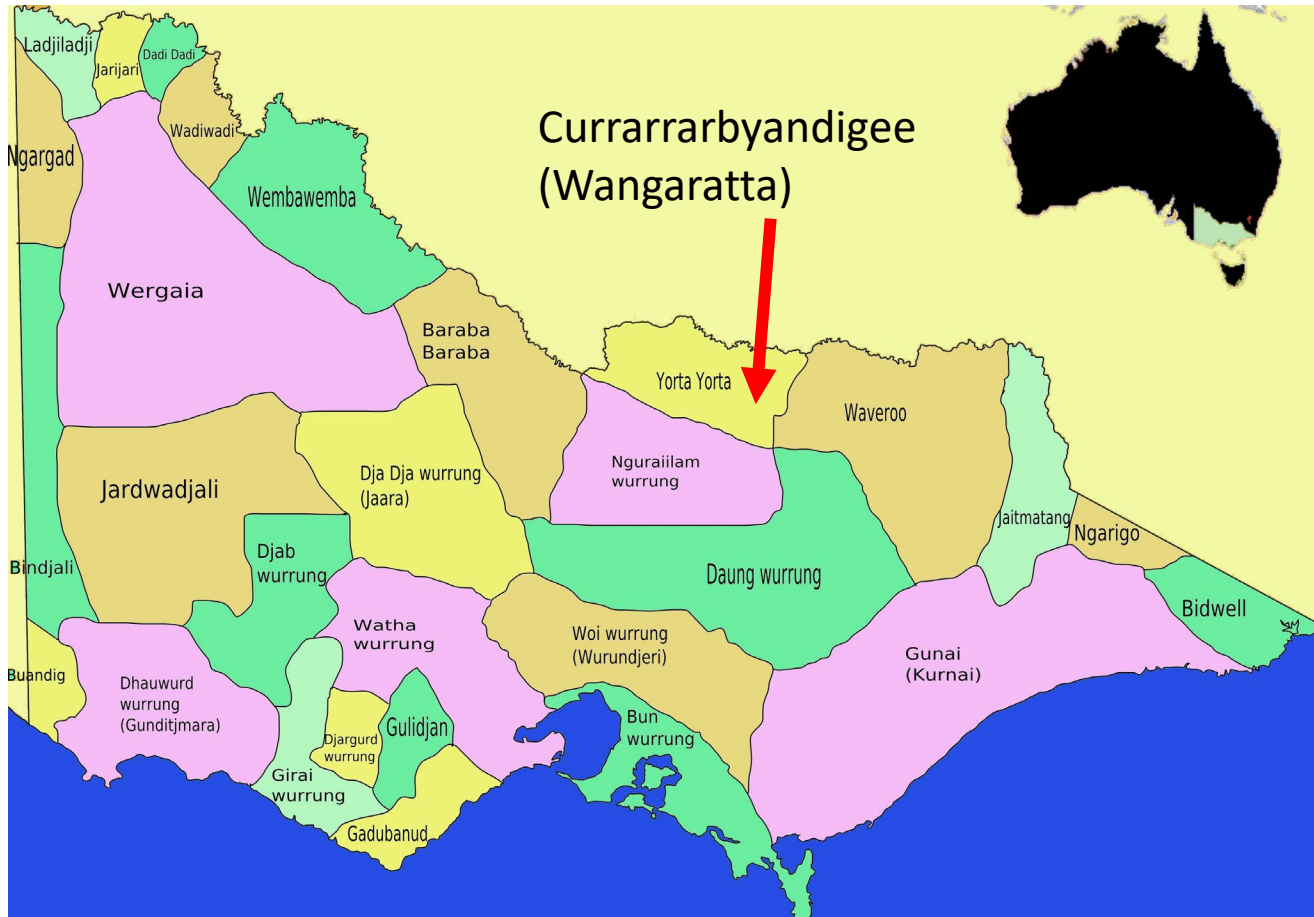
July 2022, MWA Project Meeting

Greg Slep – MWA Data Manager





Acknowledgement of Country



MAP: Tirin aka Takver - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=5541910>

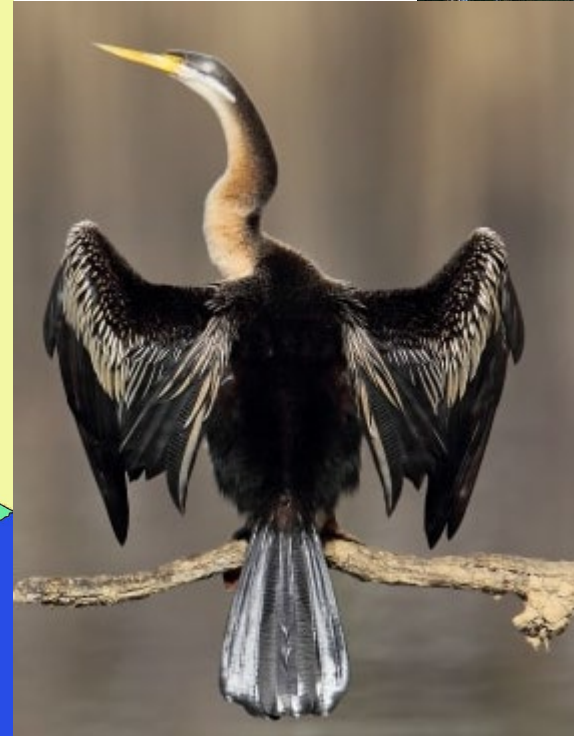


PHOTO: Chris Tzaros (Birds Bush and Beyond)





MWA Data & Archive Update

- Archive Volume Update
- ARDC Data Collections
- Data Retention Policy and Procedure Update



Archive Current Status

Status

38.5 PB / 49.0 PB

disk: 2.8 / 9.0 PB

tape: 35.7 / 40.0 PB

2022A so far: 2.8 PB

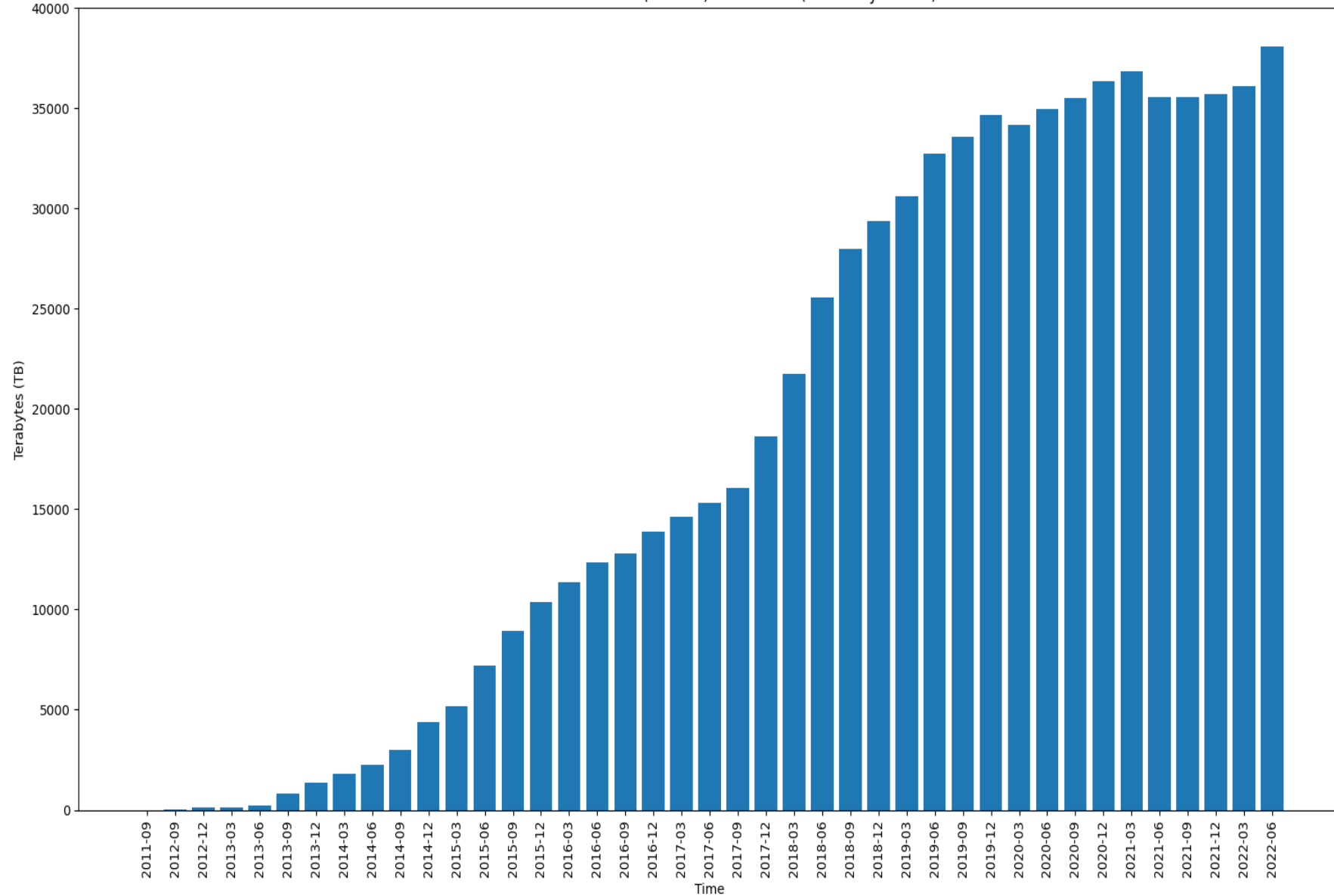
Predicted: 5-10 PB

688K+ observations

132 Project Ids

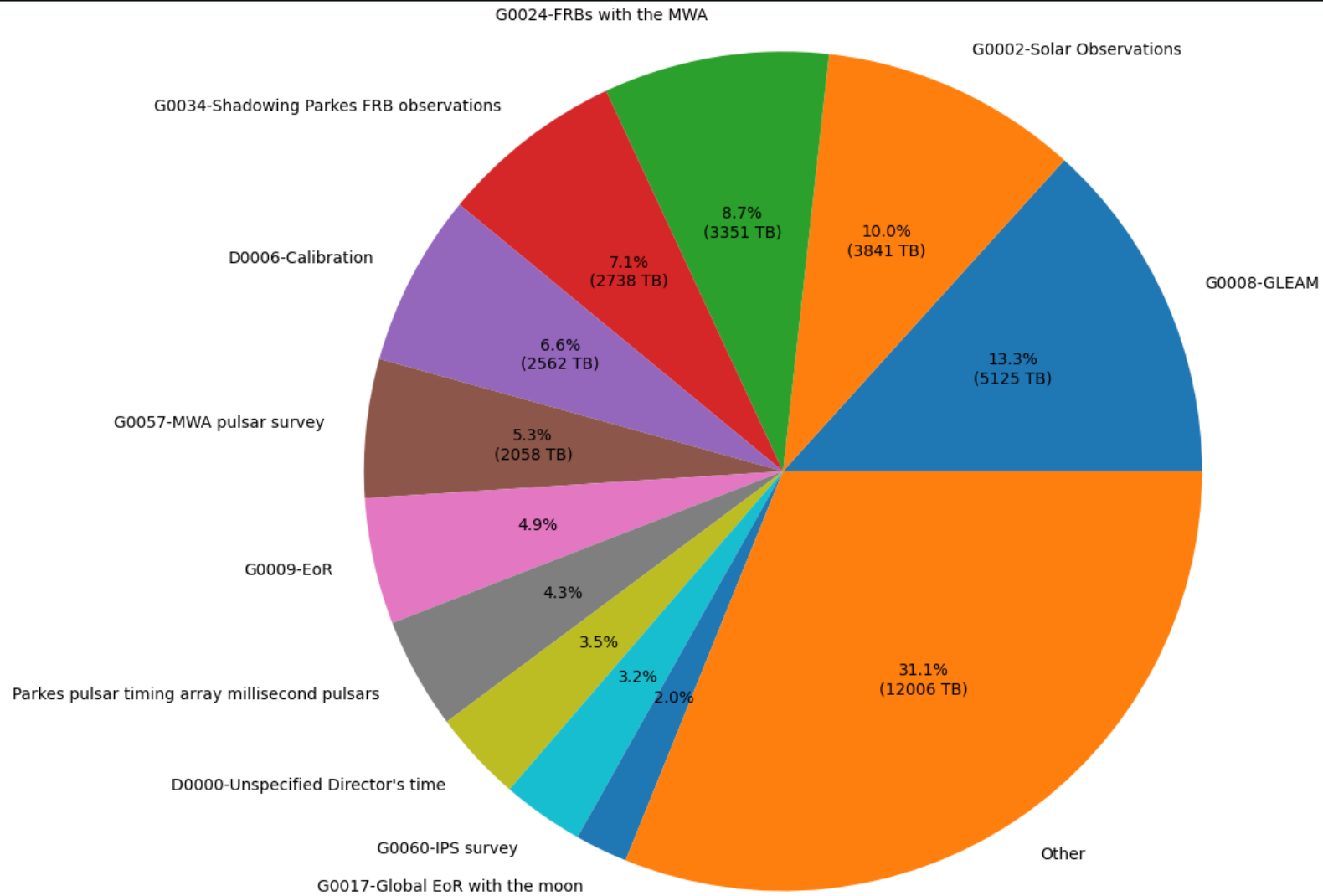
2022A MWA Project Meeting

MWA Archive Volume (all time) = 38.54 PB (as at 14-Jul-2022)





Archived Volume by Project





ARDC Data Collections

- Australian Research Data Commons (ARDC)
 - “ARDC enables the Australian research community and industry access to nationally significant, data intensive digital research infrastructure, platforms, skills and collections of high quality data”
- In 2021 call for EOI for “Significant National Data Collections”
 - Identification of collections
 - Improve metadata and FAIR
 - ARDC will fund storage
- Astronomy Australia Limited (AAL) approached Pawsey/MWA plus other ASVO nodes to complete the EOI for a specified \$ amount

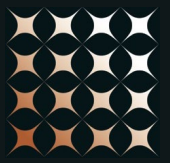




ARDC Data Collections (cont...)

- Collections:
 - EoR Raw Data (2014 - 2015) ~0.6 PB
 - GLEAM Survey Raw Data (2013-2015) ~0.9 PB
 - SMART Survey Raw Data (2018-2019) ~ 1.0 PB
- ARDC requirements:
 - Provide detailed metadata for collections ✓
 - Using metadata, mint DOIs for the collections ✓
 - Create permanent data links for the collections ✓
 - <https://asvo.mwatelescope.org/collections>





ARDC Data Collections (cont...)

- MWA: metadata and minted DOIs; Pawsey: contract/reporting
- ARDC: fund Pawsey storage (with co-funding provided by AAL)
- Pawsey will provide dedicated **~2.5 PB** for collections (Q3 or 4 2022)
- MWA Collections are not eligible for deletion
- MWA ASVO search filter coming soon

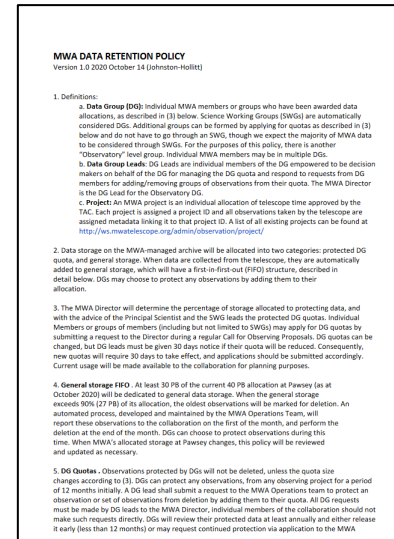


pawsey



Data Retention Policy

- MWA Data Retention Policy (v1) Published late 2020
 - <https://www.mwatelescope.org/team/policies>
- Implementation delayed: Available archive space + MWAX commissioning + Pawsey LTS
- Findings during early development:
 - Complex process, difficult to understand
 - Policy and procedure were tightly coupled
 - Large administration overhead (Ops and PS)
 - Ambiguity and various scenarios/edge cases not covered
 - Complicated software changes
- A new version of the policy was written, in consultation with the Director, Principal Scientist, Data Manager, and MWA-ASVO Developer & goes to the Board Friday



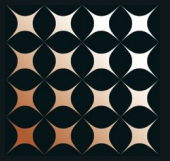


Core Data Retention Policy Principles

- MWA Archive is a rich resource for reprocessing archival data in novel ways
 - e.g. ‘A radio transient with unusually slow periodic emission’, published in *Nature* on January 27th, 2022.
- Aim is to keep as much data for as long as possible, BUT
- MWA Archive capacity at Pawsey is limited
- There is no expectation that data can be kept indefinitely*
- Researchers are responsible for their own institution’s data retention policies
- Delete data **before semester** primarily based on age, but also *utility*
- Allow MWA Members to apply to temporarily prevent deletion
- Transparency for MWA Members

All points except those in red are core to both policies

(* **ARDC Collections excepted**)

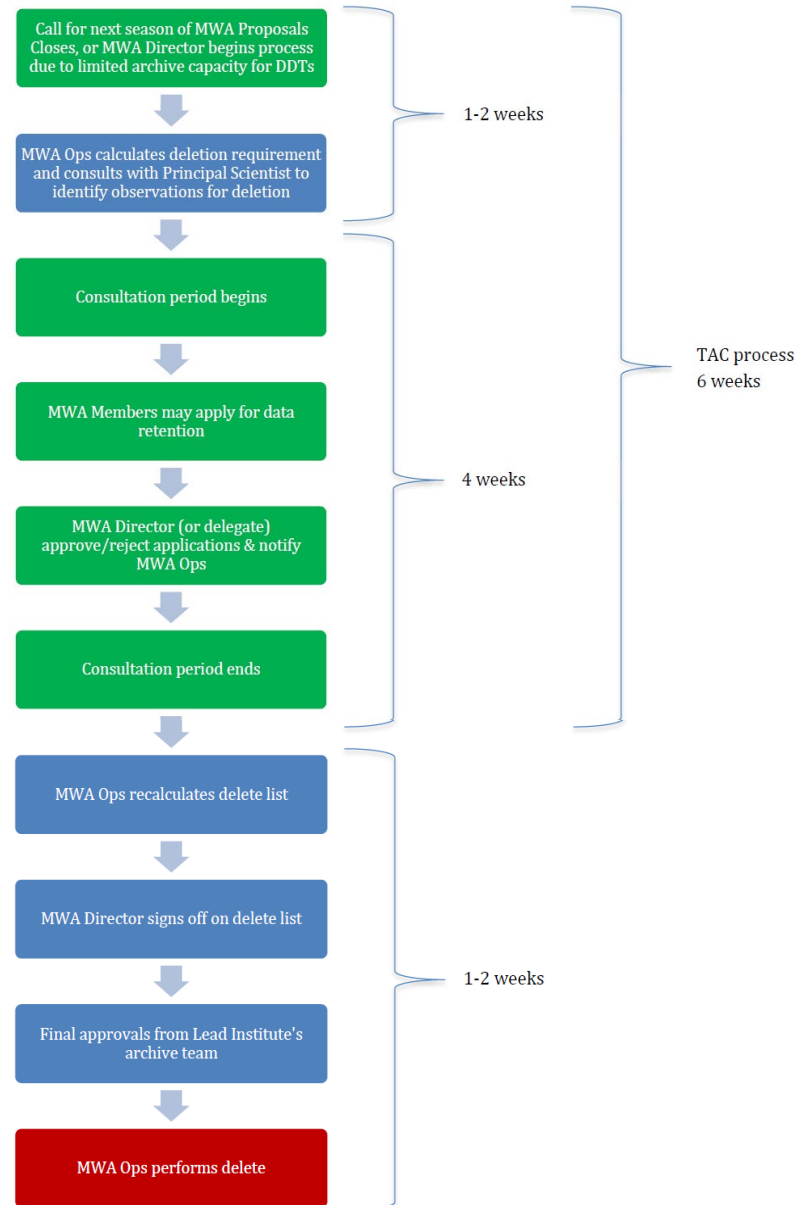


Data Retention Policy v2

- PI's may nominate bad/useless data for early deletion
- Data is deleted to make room for upcoming season observations
- Delete list is based primarily on age but also utility
- MWA Collaboration notified with summary of delete list
- MWA Members have 4 week consultation period
- MWA Members may apply to Director for data retention
- At end of review period, data is deleted (except retained data)
- Data retained this time is subject to deletion next time



Data Retention Policy v2 - Procedure





Data Retention Policy v2 – Procedure (cont...)

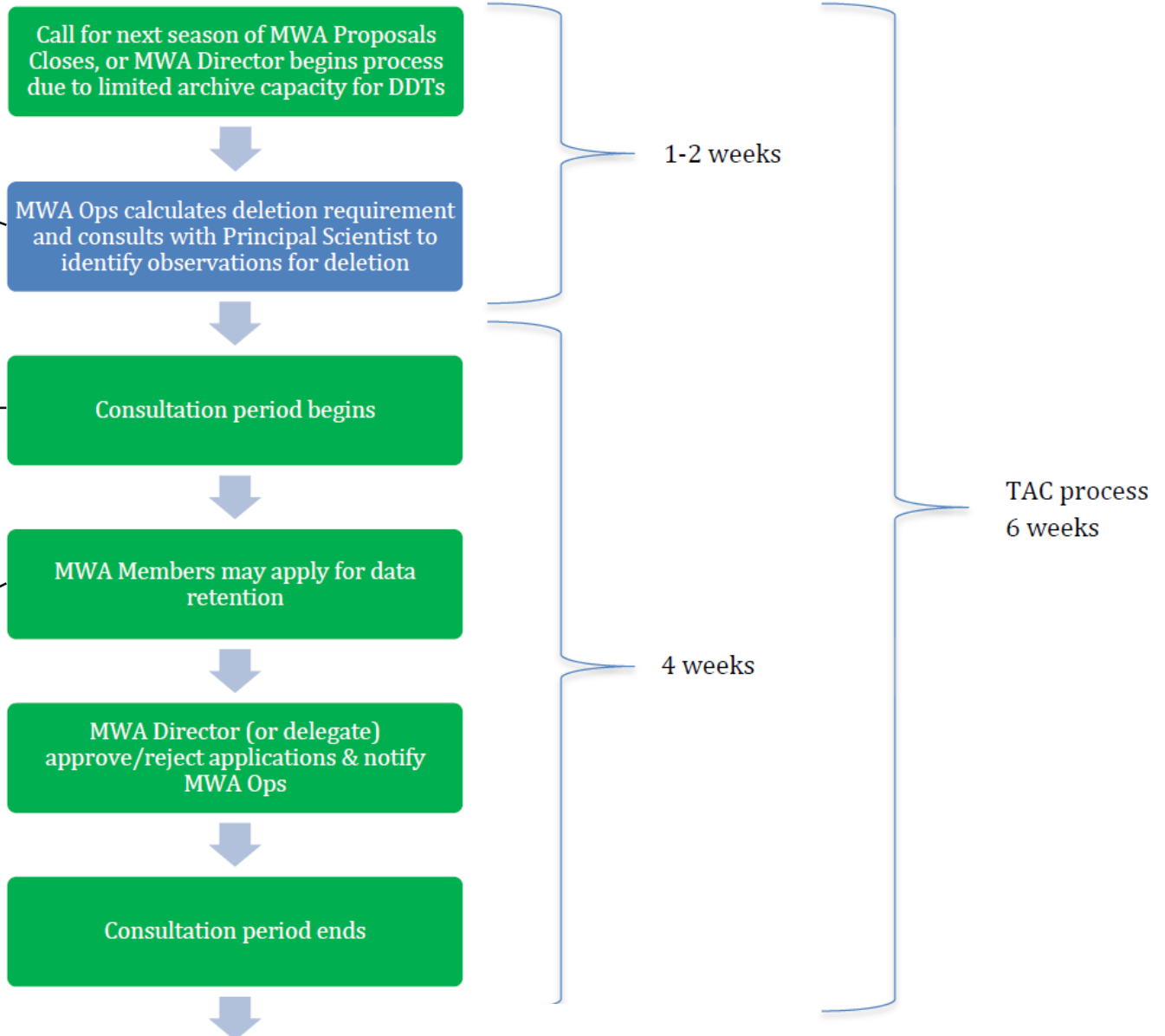
Delete list:

- oldest, with PS review
- Collection data exempt
- + X % extra margin for error

Summary email to collaboration of delete list, including PI nominated

Application to Director:

- Describes what is to be retained
- Significance & justification
- Plan for copying / own retention

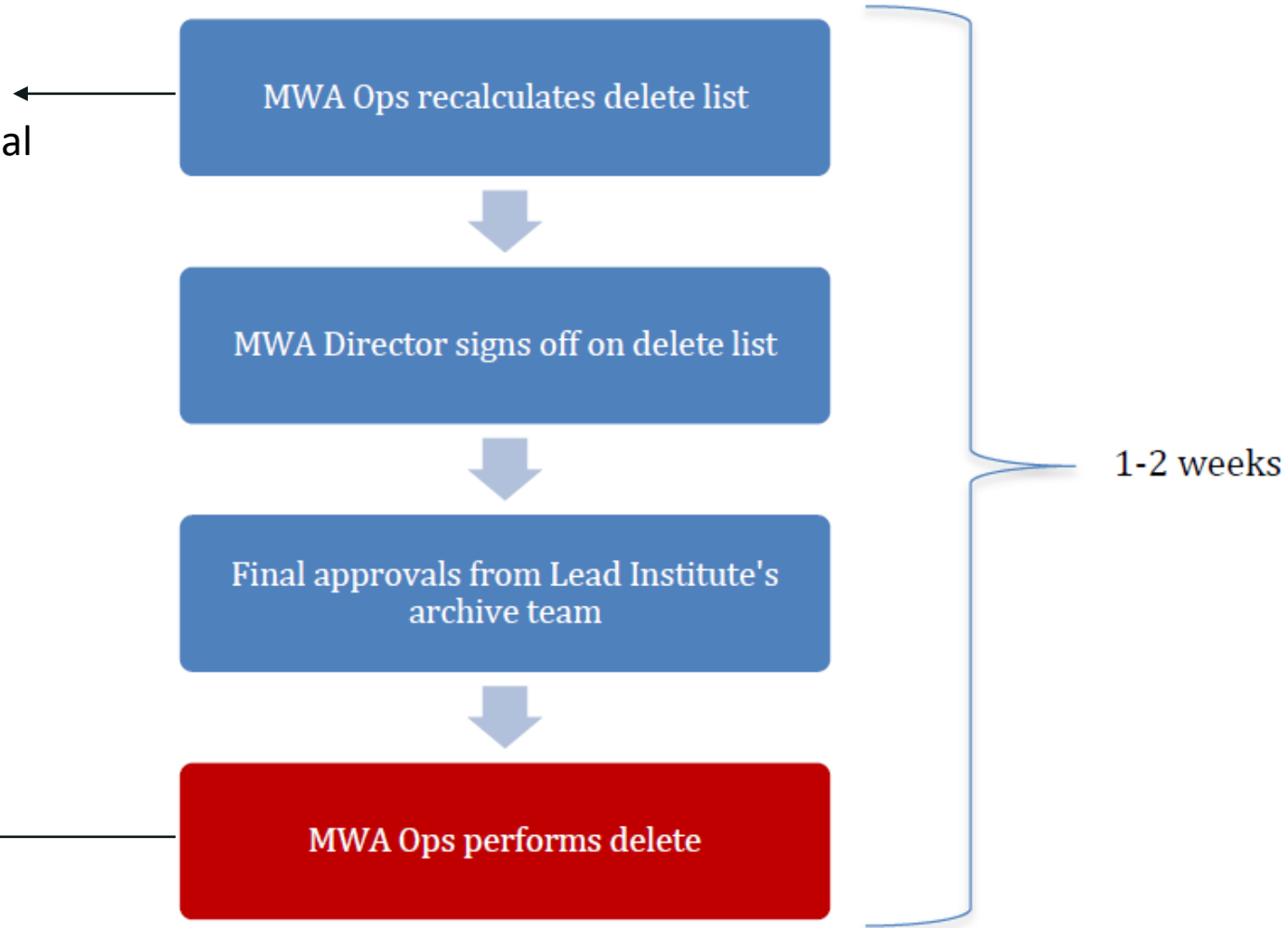




Data Retention Policy v2 – Procedure (cont...)

Update delete list based on:

- Archive space required for final TAC approved proposals
- exclude retained data



Data files deleted. Metadata retained



Data Retention Policy v2 Summary

- New policy and procedure up for review by MWA Board
- Intent of new policy remains the same
- Balances taking new data while removing oldest, low value data
- Separation of policy and procedure
- Simpler, easier to understand process
- More feasible to implement
- Adheres to MWA lead organization records management policies



Thank You!

Questions?

Stay tuned: after the coffee break, we will open a discussion:
“Adapting to a Post-Limitless Archive”

Contact me:

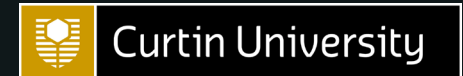
- greg.sleap@curtin.edu.au
- MWA Users Slack: @Greg Sleap



Adapting to a Post-Limitless Archive

July 2022, MWA Project Meeting

Greg Slep – MWA Data Manager





Adapting to a Post-Limitless Archive...

	Receivers	Tiles	Voltages vs 128T	Baselines	Visibilities vs 128T	2022A Size (VCS+Vis PB)
	16	128	100%	8,256	100%	5.0+5.0
We are here!	17	136	106%	9,316	113%	5.3+5.6
	18	144	112%	10,440	126%	5.6+6.3
	20	160	125%	12,880	156%	6.2+7.8
	24	192	150%	18,528	224%	7.5+11.2
	32	256	200%	32,896	398%	10.0+19.9