How to save $24M a year in researcher time

A cost-benefit analysis of persistent identifiers in Australian research systems

https://doi.org/10.5281/zenodo.7100578

CLIENT LOGO HERE

Josh Brown
Co-founder, research & strategy
josh@morebrains.coop

Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

The big picture - wasted time and money

- Admin tasks take **30-40% of researchers' time*** - an unsustainable burden
- Total time cost of manually rekeying metadata about funded grants, projects and publications into computer systems is nearly **38 thousand person days** per year
- Total financial cost is $**24 million** per year

*J. Miller, 'Where does the time go? An academic workload case study at an Australian university', *Journal of Higher Education Policy and Management*, vol. 41, no. 6, pp. 633–645, Nov. 2019, doi: 10.1080/1360080X.2019.1635328.

Sacha Jafri - Journey to Humanity
(Worlds largest canvas painting)

There is more work to be done...

I note the higher education sector's concern regarding the workload required for the current mode of delivery of the ERA assessment.

...

Streamlining the processes undertaken during National Competitive Grant Program funding rounds must be a high priority for the ARC ... I ask that the ARC identify ways to minimise administrative burden on researchers.

- Hon Jason Clare MP
  Minister for Education
  Statement of Expectations 2022

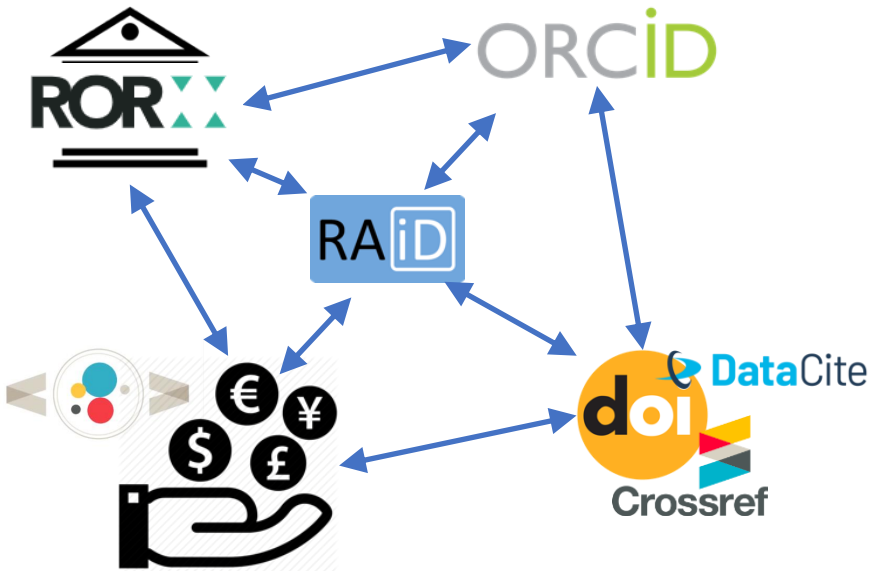
ARDC
Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MORE BRAINS

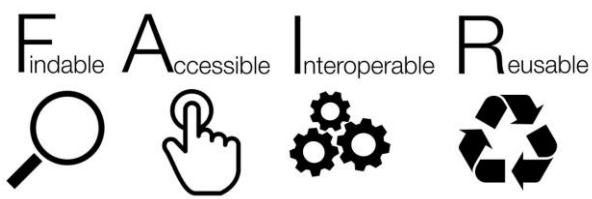
Why PIDs are important

Characteristics that make PIDs invaluable



PIDS…

- are unique
- don't change
- are connected to each other
- can be read by computers
- save effort and time when finding and compiling information
- Support a range of practical and policy priorities



Australian Research Data Commons

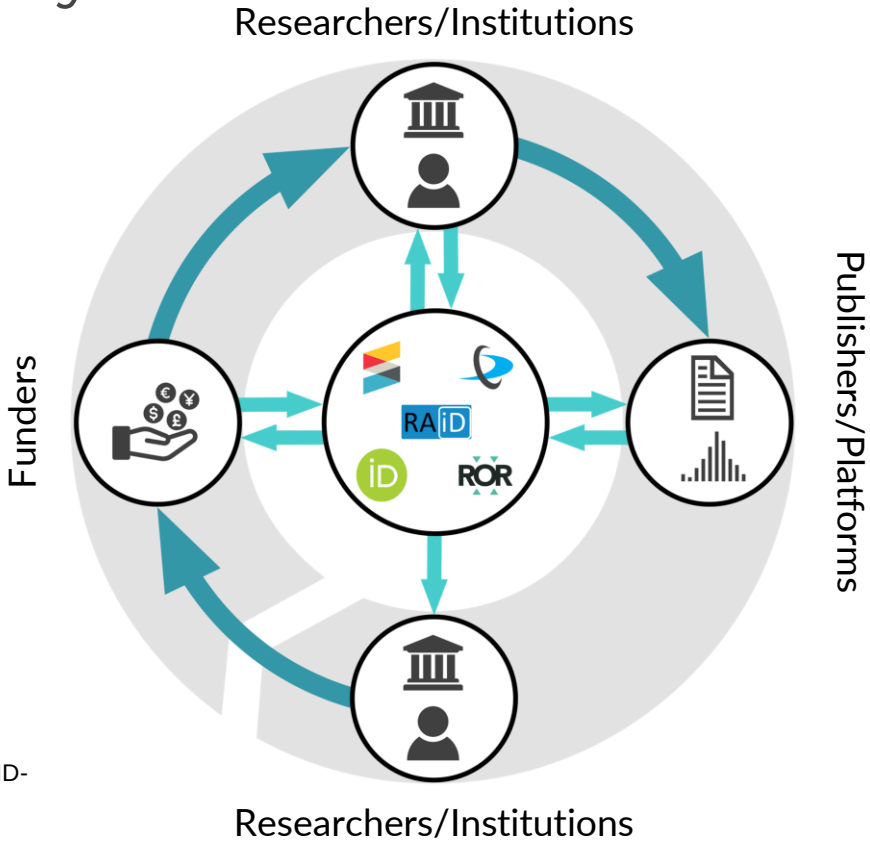
AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

PID-optimized research lifecycle

Three areas of benefit:

1. Movement of data between systems
2. Automation creates efficiencies
3. Better strategic decision making



https://resources.morebrains.coop/pidcycle/

Brown, Josh, Jones, Phill, Meadows, Alice, & Murphy, Fiona. (2022, September 16). PID-optimised workflows: A vision of a more efficient future. Zenodo. https://doi.org/10.5281/zenodo.7085489

Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

Five priority PIDs

Grants - Crossref DOIs for grants allows unique identification and certainty of referencing

People- ORCIDs represent a verifiable, reusable record of the employment, funding history, and research outputs of an individual

Projects - A portable container of research project activities connecting the people, publications, instruments and institutions involved

Organisations - Unique identifier for organisations for accurate discovery of the activities, outputs and impacts of research institutions

Outputs - DOIs as registered by both Crossref and DataCite are unique and permanent identifiers of publications and other outputs

3. Research

The scale of Australian research activity

Number of researchers

- HERDC latest figures are ~ 108k FTE
  https://app.powerbi.com/view?r=eyJrIjoiZDQxMDZmN2UtNTc3OS00OWU3LThiMWQtYmVlNzgzNDhkYTEyIiwidCI6ImRkMGNmZDE1LTQ1NTgtNGIxMi04YmFkLWVhMjY5ODRmYzQxNyJ9

Number of researchers per publication (average)

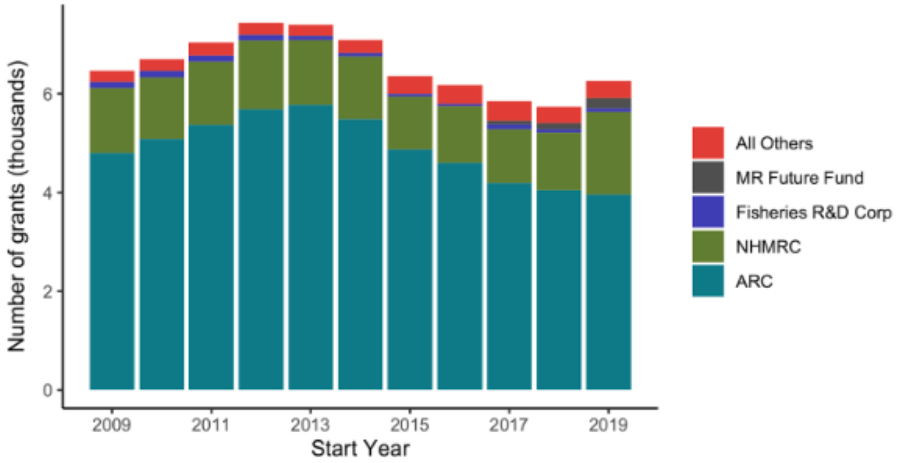
- D. Fanelli and V. Larivière, 'Researchers' Individual Publication Rate Has Not Increased in a Century', *PLoS ONE*, vol. 11, no. 3, p. e0149504, Mar. 2016, doi: 10.1371/journal.pone.0149504.

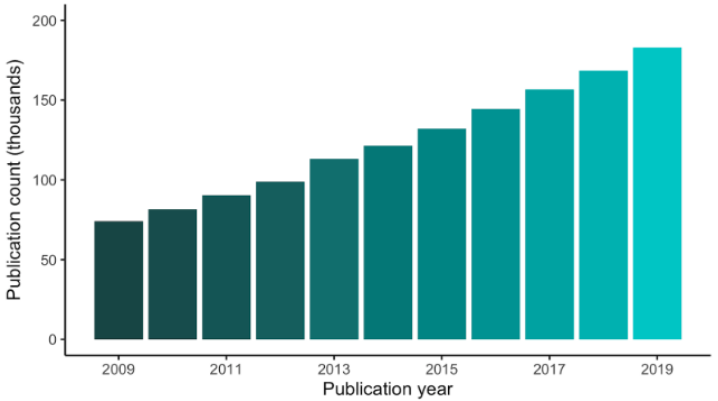
Length of time needed to enter project / grant / publication information

- Research Consulting, 'Counting the Costs of Open Access', London Higher and SPARC Europe, Nov. 2014. Accessed: September 5, 2022. [Online]. Available at: https://www.research-consulting.com/wp-content/uploads/2021/07/Research-Consulting-Counting-the-Costs-of-OA-Final.pdf
- M. H. Klausen, 'Even Minor Integrations Can Deliver Great Value – A Case Study', Procedia Comput. Sci., vol. 106, pp. 153–159, 2017, doi: 10.1016/j.procs.2017.03.011

6 thousand grants and 180 thousand publications

- Funded grants - Digital Science Dimensions & data from ARC and MRFF
- Volume of publications - Digital Science Dimensions





Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

Data that were harder to come by

Number of Australian researchers with an ORCID

- 122k ‘active’ ORCIDs with .au email suffixes

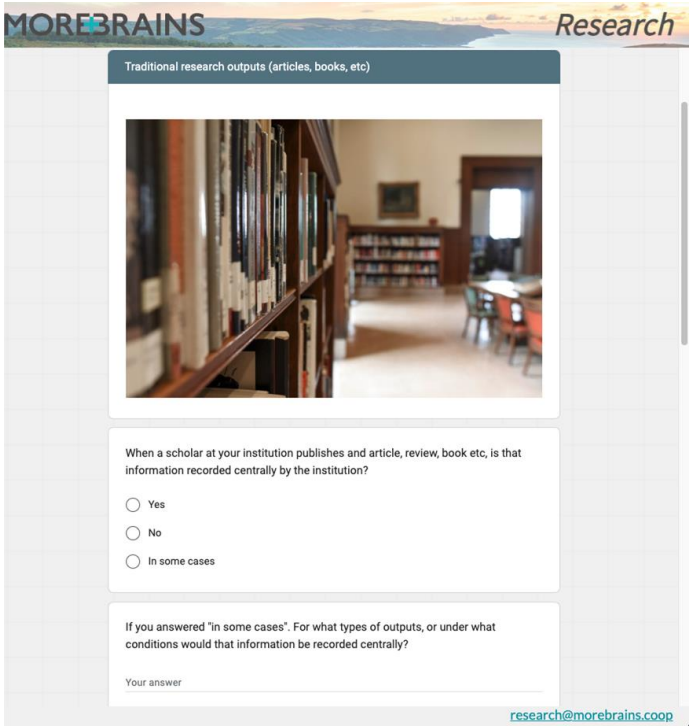
Number of projects in Australia

- Based on a previous estimate of the number of projects in the UK, scaled by the total level of R&D funding based on OECD data (~25,000)

How many times are data about grants and publications rekeyed

- We asked research administrators - see next slide

Metadata rekeying survey



We asked repository and research managers at universities across Australia

How many times is information about grants and publications entered into systems?

- 23 research institutions responded to our survey
- Grants information entered **3.25** times and read 5.9 times
- Publication information was entered **3.1** times and read 7.3 times

Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

Information that proved impossible to find

Some data was just not obtainable or did not exist

- Number of research instruments
  - No identifier or common definition on which to base an estimate
  - There is RDA working group on PIDs for Instruments - potentially a PID to incorporate in future strategy assessments
- Geological samples
  - IGSNs are minted by ARDC and will be adopted by TERN
  - No way of knowing what proportion of samples have IGSNs

Nearly 38 thousand days and $24 million dollars

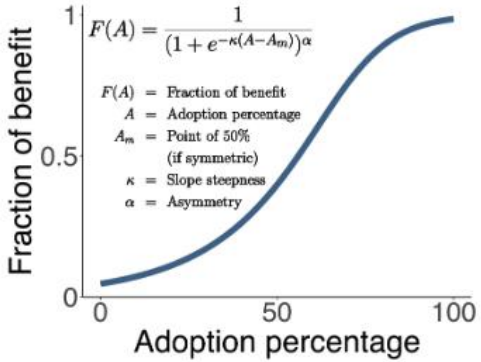
Potential total sector savings								
	Number	*# authors*	*# rekey events*	*# minutes / event*	*cost / author / minute*		*Time savings per year (person days)*	*Financial savings per year*
Publication metadata	180,000	4	3.1	6.73	$1.13	- if Admin	34,532	$17,037,747
					$1.06	- if Junior Researcher		$15,972,888
					$2.13	- if Senior Researcher		$31,945,775
						Average		**$21,652,137**
Grant metadata	6,000	-	3.25	10	$1.13	- if Admin	448	$221,176
					$2.13	- if Senior Researcher		$414,705
						Average		**$317,940**
Project descriptions	21,083	-	6	10	$1.13	- if Admin	2,908	$1,434,817
					$1.06	- if Junior Researcher		$1,345,141
					$2.13	- if Senior Researcher		$2,690,282
						Average		**$1,823,413**
Total predicted annual savings from auto-feed of key metadata via API links:							**37,888**	**$23,793,490**

Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

Network effects are important for benefits



Savings levels based on levels of adoptions						
Institutional adoption levels	**0%**	**20%**	**40%**	**60%**	**80%**	**100%**
Realised benefit	0.0%	1.8%	11.9%	50.0%	88.1%	100.0%
Effective time savings (person days)	0	681	4,516	18,944	33,372	37,888
Effective financial savings ($ millions)	$0.00	$0.43	$2.84	$11.90	$20.96	$23.79

- The more organisations that adopt PID-enabled workflows
  ⇨ the more data is available
    ⇨ the more benefit to adoption
- A strategy is required with a high target of adoption (e.g. 80%) for all five priority PIDs

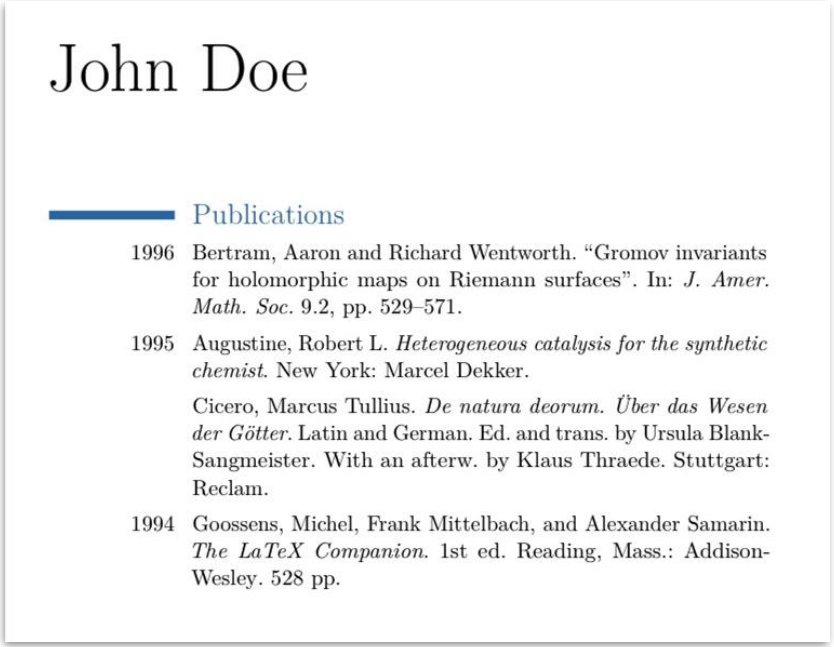
Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

The case studies

1. ORCID integration at ARC - the problem



John Doe

Publications

1996 Bertram, Aaron and Richard Wentworth. "Gromov invariants for holomorphic maps on Riemann surfaces". In: *J. Amer. Math. Soc.* 9.2, pp. 529–571.

1995 Augustine, Robert L. *Heterogeneous catalysis for the synthetic chemist*. New York: Marcel Dekker.

Cicero, Marcus Tullius. *De natura deorum. Über das Wesen der Götter*. Latin and German. Ed. and trans. by Ursula Blank-Sangmeister. With an afterw. by Klaus Thraede. Stuttgart: Reclam.

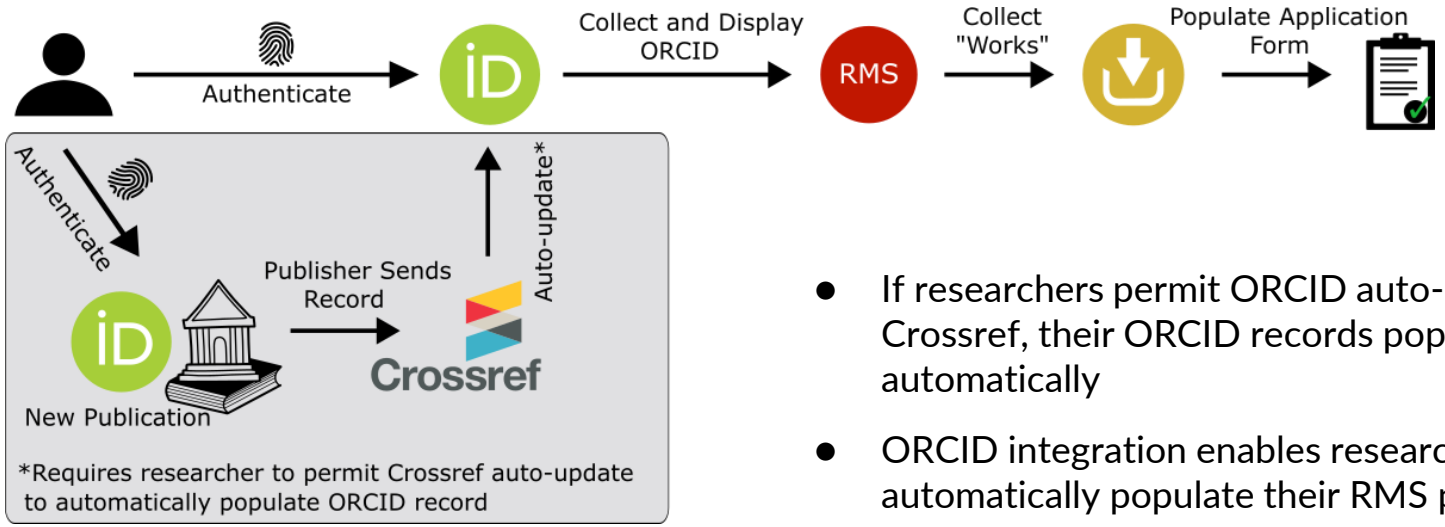
1994 Goossens, Michel, Frank Mittelbach, and Alexander Samarin. *The LaTeX Companion*. 1st ed. Reading, Mass.: Addison-Wesley. 528 pp.

Prior to 2018, applications for grants from ARC required a hand-formatted list of publications

Before using ORCID, an average application used to take a few weeks; "formatting took time, getting the publications right took many days of work". This took time away from the actual grant process.

\- Joe Shapter
Pro-vice-Chancellor, University of Queensland

1. ORCID integration at ARC - the solution



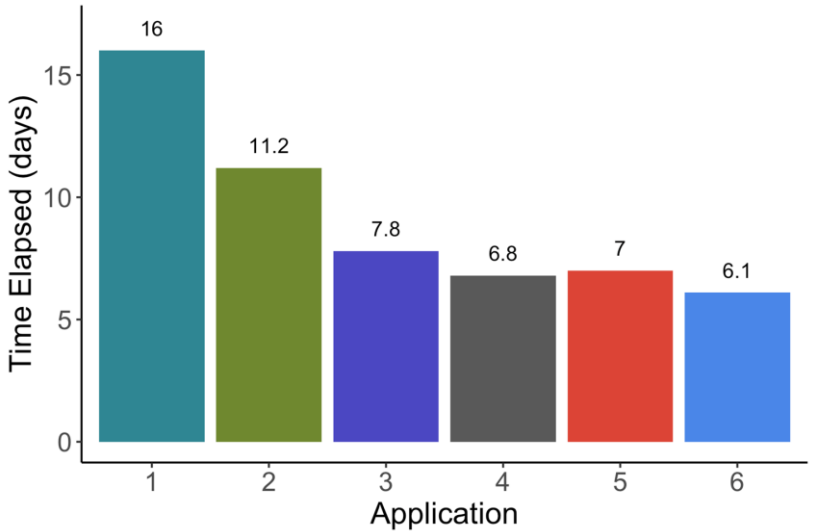
- If researchers permit ORCID auto-update via Crossref, their ORCID records populate automatically
- ORCID integration enables researchers to automatically populate their RMS profiles
- The RMS profile is then used to auto-populate the grant application

This saved me 3-4 days per grant application - the difference in workload was staggering!"

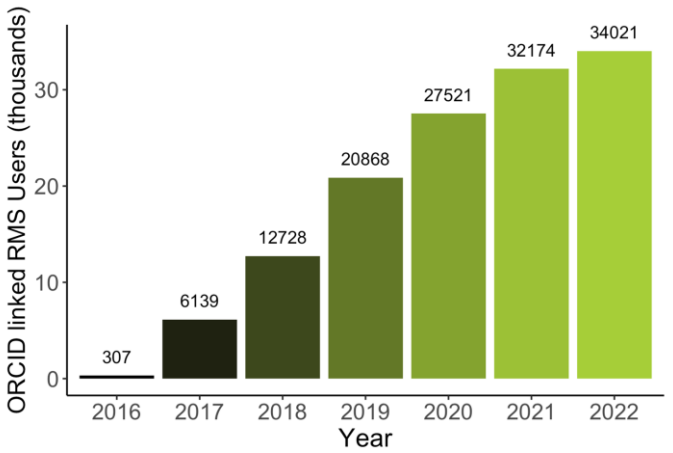
- Joe Shapter

1. The savings from a single integration

Median time between a the application first save to final submission shows the more people use the new system, the less time the spend populating it



- **78%** of publications submitted to support ARC grant applications are via ORCID
- Total cost savings of nearly **$850 k**



Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

2. The use of PIDs at TERN

The Terrestrial Ecosystem Research Network (TERN) is Australia's terrestrial ecosystem observatory.

Research conducted through TERN is necessarily highly collaborative and extremely diverse.

- Time-series remote sensing
- In-situ sensors data sets
- Vocabularies
- Traits
- instrument records
- Methods
- Organisations
- People.
- Tabulated data
- Images
- Acoustic files
- Geospatial vector and raster data
- Among many other types…

We spoke to a number of researchers at TERN and learned how **PIDs create a ground truth for datasets, vocabularies, and samples that is critical to collaborative work in ecology and environmental science**

Photo: Warra Tall Eucalypt

Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

3. Central services provided by ARDC and AAF

Central services reduce risk and barriers to entry for individual stakeholders

Australian Research Data Commons

DataCite

RAiD

Handle.Net®

IGSN

DOIs - DataCite registration agency

RAiD - Research Activity Identifier

Handles - General purpose identifier

IGSN - Physical samples and specimens

PURLs - for Research grants

AUSTRALIAN ACCESS FEDERATION

ORCID

National ORCID Consortium

Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

3. ORCID Consortium saves about $4.6 million

	via Consortium (AUD)	Via ORCID (AUD)	Savings / Member (AUD)
2016	$11,454.85	$27,894.00	$16,439.15
2017	$10,912.19	$26,109.66	$15,197.47
2018	$12,270.74	$26,546.39	$14,275.65
2019	$13,621.12	$29,096.05	$15,474.93
2020	$13,886.22	$29,137.20	$15,250.98
2021	$13,511.60	$29,137.20	$15,625.60
2022	$14,310.18	$30,011.32	$15,701.14
TOTAL	$89,966.90	$197,931.82	$107,964.91

- Forty-two institutional members from across Australia
- Lower barrier to entry of adopting ORCID
  - Universities do not have to pay for their own membership until they reach a certain level of usage
- Costs to members of ORCID membership reduced by over 50%

Incentives to invest in identifiers

We estimate the total wasted time and effort due to rekeying of information for grants, projects and publications

38 Thousand person days - $24 million per year

PIDs are unique, permanent identifiers for people, places and things that allow information to be automatically moved between computer systems - thereby radically reducing tedious reentry of information

- A national PID strategy for Australia building on:
  - Success of the AAF-led Australian ORCID consortium
  - Leadership provided by ARDC
- Funders should build on the success of ARCs integration with ORCID
  - Implement similar integrations and expand out to other PIDs
- Whole sector approach focused on integrations of priority, open PIDs:
  - ORCIDs for people
  - ROR for institutions
  - RAiDs for projects
  - DOIs for research outputs
  - DOIs for grants
- Set a target of 80% adoption within five years

Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS

Thank you!

Any Questions?

Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION

MOREBRAINS