

Making the case for investment in PIDs



Mary Beth Barilla
Director of Business
Development
NISO



Phill Jones
Co-founder,
Digital & Technology
MoreBrains Cooperative



Co-funded by
the European Union





Agenda

- Introduction
- Making the case at the national level
- Small group discussion
- Panel discussion
- Q&A



Making the case for investment

National cost-benefit analyses

Phill Jones

Co-founder, Digital and Technology
MoreBrains Cooperative

 <https://orcid.org/0000-0003-0525-6323>



Co-funded by
the European Union



Acknowledgements

The other members of the MoreBrains team

Thanks for the data!



Alice Meadows



Josh Brown



Fiona Murphy, DPhil



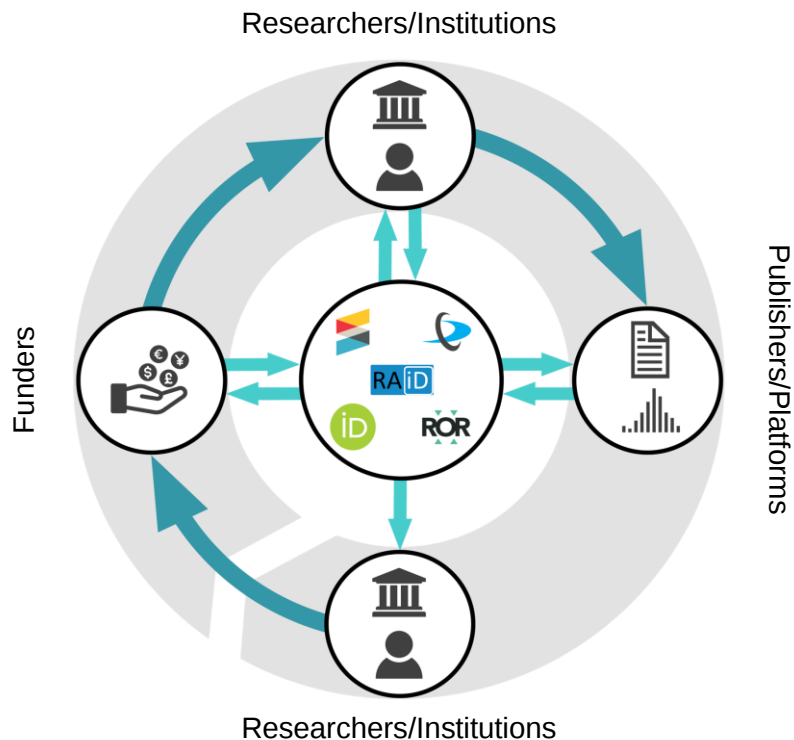
Partners and clients that have contributed to related work



MORE+BRAINS

1) The why and how of driving PID adoption

PID Optimised research cycle



- Open: Information in the form of metadata can move from one system to another
- Efficient: Automated movement of information results in complete and timely records
- Trackable: Links between PIDs enable connections to be analysed
- Requires standards, technical and social integrations between
 - Funders
 - Research Institutions
 - Publishers

<https://resources.morebrains.coop/pidcycle/>
<https://doi.org/10.5281/zenodo.4991733>

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>

If you build it, will they come?

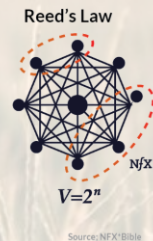
Whenever you build anything, you inevitably make assumptions:

Two important assumptions are sometimes made:

- If you provide a way for people to enter information, they will do so
- There is an **obvious** need to create profiles / records of activity for
 - Researchers
 - Institutions
 - Funders
 - Publishers

...but PIDs have value as a network

- Value grows as number of participants grow



“Open research requires collective action. The transition to a national open research environment is a shared responsibility across the research system, with local efforts required to implement policies, infrastructure and incentives within research performing and research funding organisations.”

From the Irish National Plan for Open Research
2022-30

Why do we need national strategies?

Yes...

Research and technology are global

A common set of core PIDs makes sense

We all have the same end goal

but...

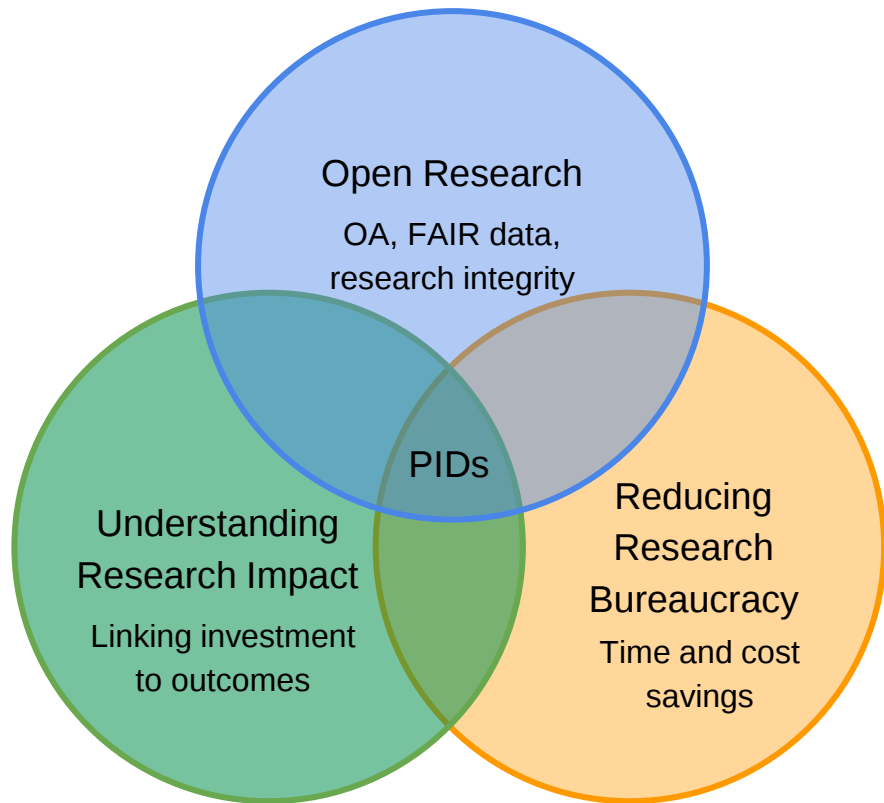
Policies and funding are national / regional

There are national priorities, sensitivities and requirements

We're all starting in different places

2) Policy drivers

| Key policy drivers



Landscape review of national PID strategy development work, including RDA PID IG

Countries/regions we have looked at:

- United Kingdom
- European Union
- USA
- Australia
- Canada
- Ireland
- Netherlands
- Finland
- Colombia
- South Korea

Open Research

In the UK, the PID strategy began as a way to support the movement to open research



“Jisc to lead on selecting and promoting a range of unique identifiers, including ORCID, in collaboration with sector leaders with relevant partner organisations.”

Professor Adam Tickell, [Open access to research: independent advice](#) (2020)



<https://doi.org/10.5281/zenodo.5609266>



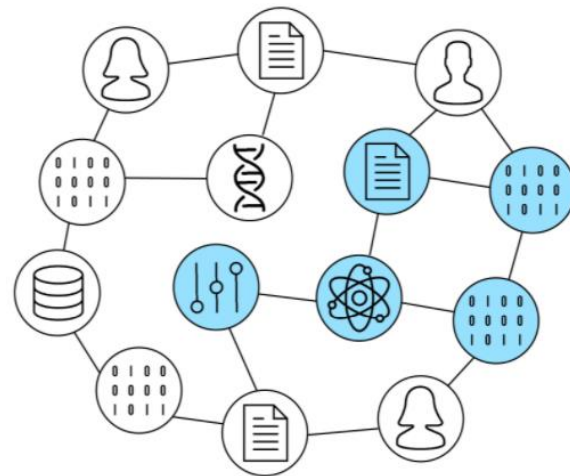
PIDs are considered fundamental to making research data Findable, Accessible, Interoperable and Reusable

FAIRsFAIR “Fostering FAIR Data Practices in Europe”,
EU Horizon 2020 project: INFRAEOSC-2018-2020

Research impact assessment

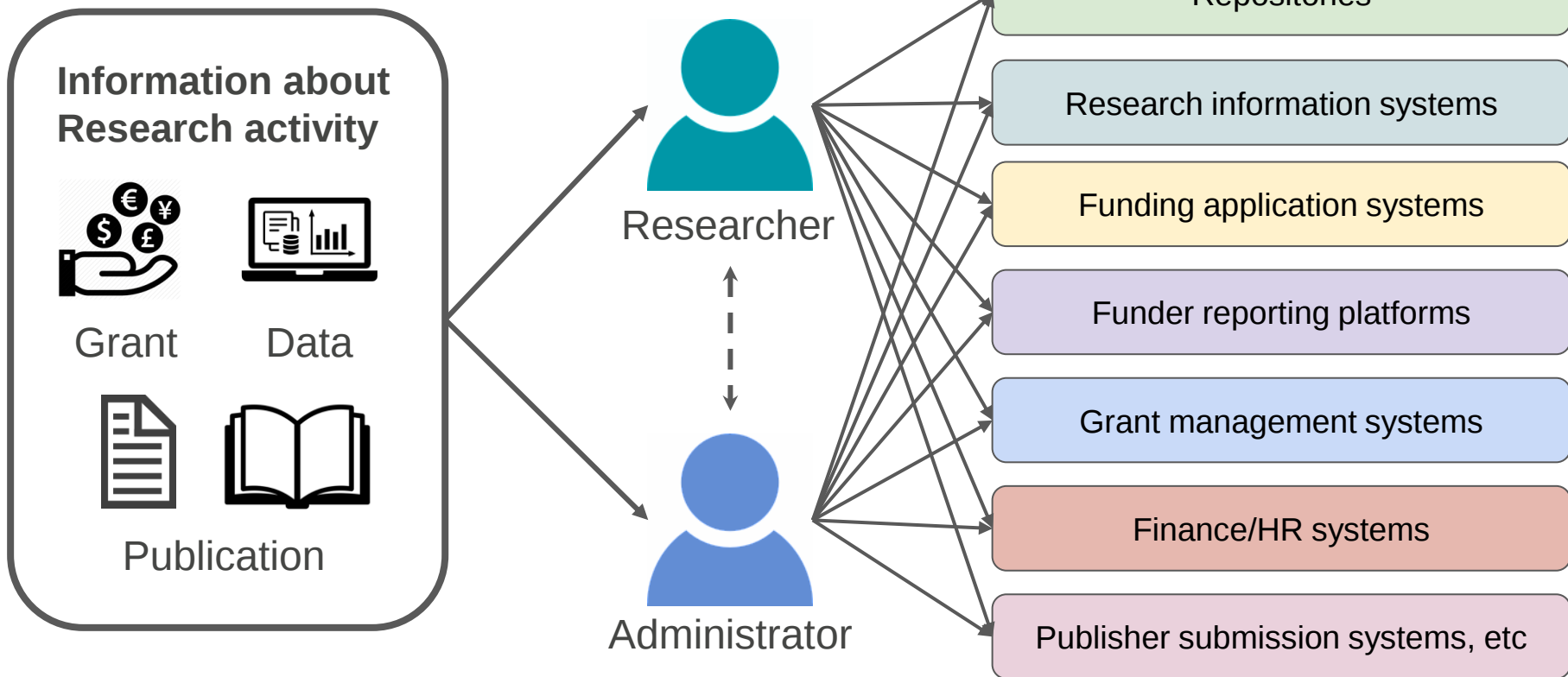
PID metadata are the edges of the research graph

- Governments need to understand the impact of their investments in R&D
- Funders need to understand the research outputs associated with grants
- Institutions need to understand who they're collaborating with, where they're publishing and where their funding is coming from
- Researchers need to connect research outputs, collaborators and find funding opportunities
- Publishers need to understand which institutions their authors are from, and connect research outputs to make them discoverable



Introducing the PID Graph, Fenner and Aryani (2020)
<https://www.project-freya.eu/en/blogs/blogs/the-pid-graph>

Wasted time and effort duplication



Lack of integrations cause massive inefficiencies

United Kingdom

- 55,000 person days / year
- £19M / year total cost
- £45M net savings (5 years) after investment

DOI: [10.5281/zenodo.7356219](https://doi.org/10.5281/zenodo.7356219)

Australia

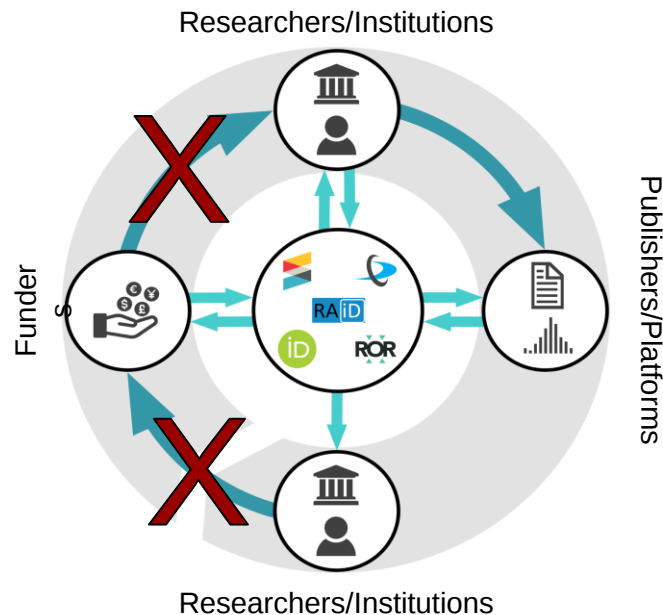
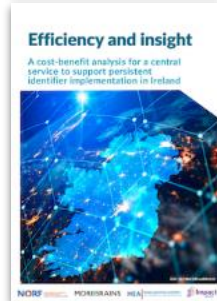
- 38,000 person days / year
- \$24M AUD / year total cost

DOI: [10.5281/zenodo.7100578](https://doi.org/10.5281/zenodo.7100578)

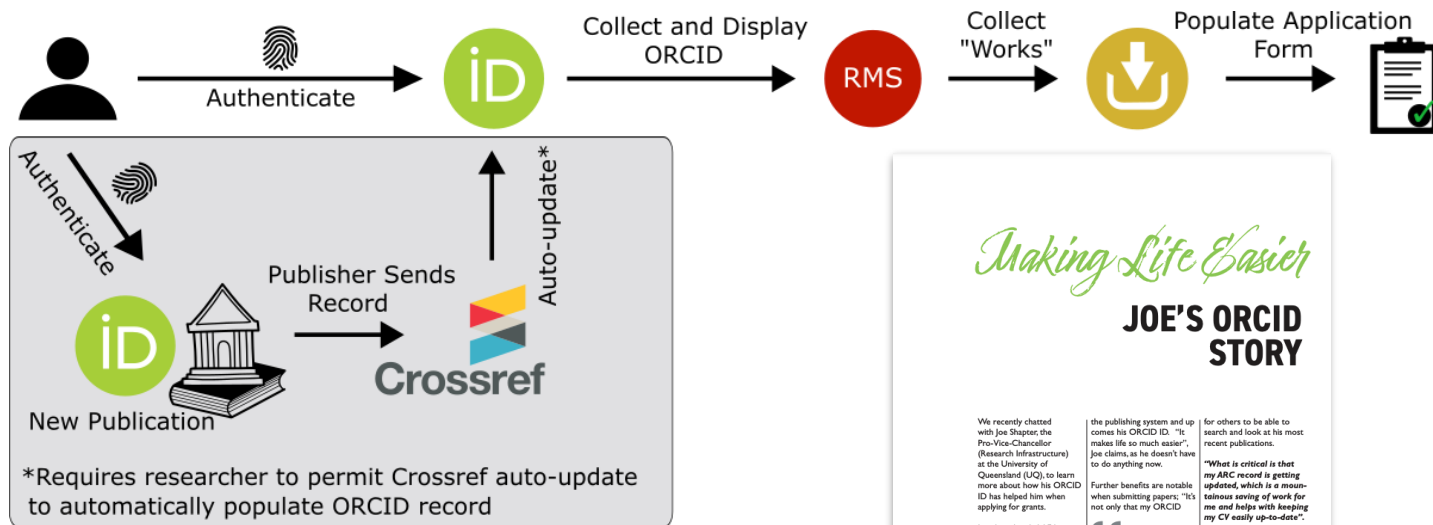
Ireland

- 4,054 person days / year
- €1.8M / year total cost
- € 1.6M Net savings (5 years) after investment

DOI: [10.7486/DRI.nz80kt123](https://doi.org/10.7486/DRI.nz80kt123)

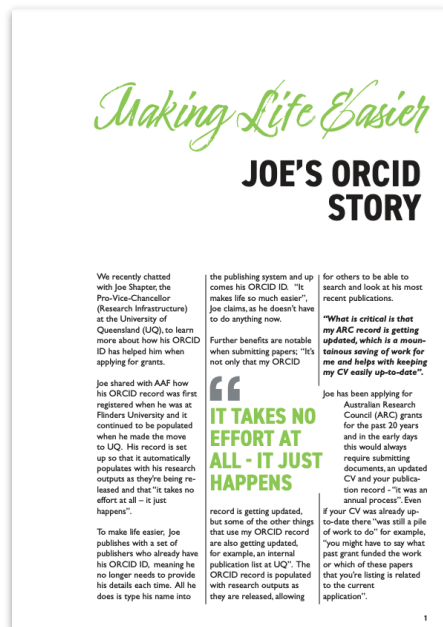


Case Study: ORCID & Australian Research Council



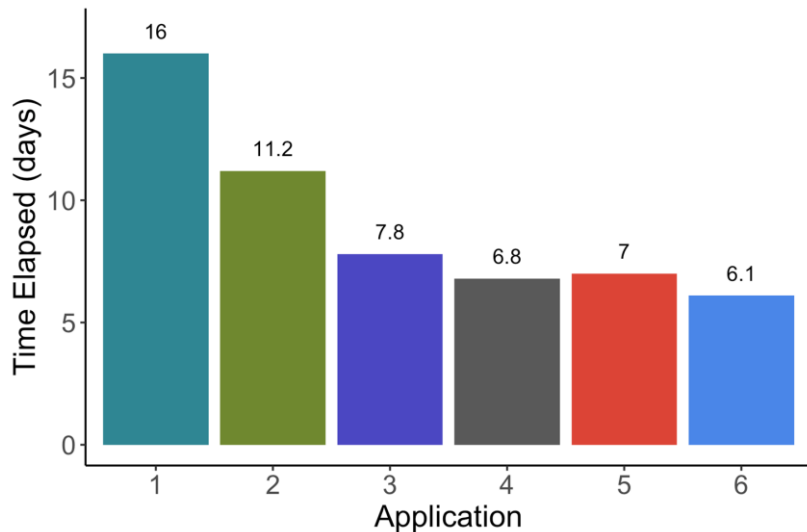
*This saved me 3-4 days per grant application
- the difference in workload was staggering!"*

- Joe Shapter, PVC (Research Infrastructure)
University of Queensland



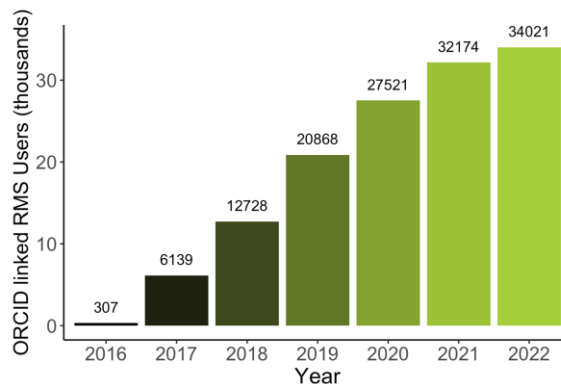
ORCID & ARC - Cost savings

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



Post-award savings:

- **78%** of publications submitted to final reports via Crossref integration
- Cost savings of nearly **\$850 k** post award alone



1/3 of ORCID's for
Australians now
linked to RMS

4. Calculating the benefits of PID adoption

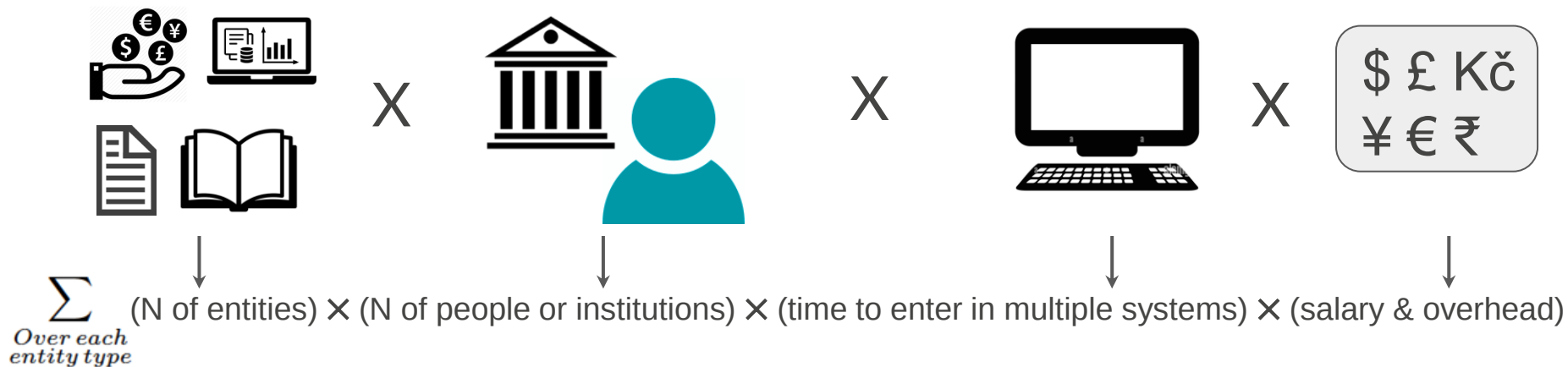
| The mechanism for direct cost savings

Researchers and administrators spend significant amounts of time keying metadata into various systems:

- Funding application systems
- Grant management systems
- Funder progress reporting
- CRISs
- Repositories
- Finance
- Publisher submission systems
- Etc. etc.

Question: How much time (and therefore money) would be saved if this information about research activity moved from system to system automatically?

The formula for calculating costs of rekeying



- 1 - Calculate the number of entities
- 2 - Multiply by the number of people / institutions
- 3 - Multiply by the number of systems and time taken
- 4 - Multiply by salary data

| Estimating the scale of research activity

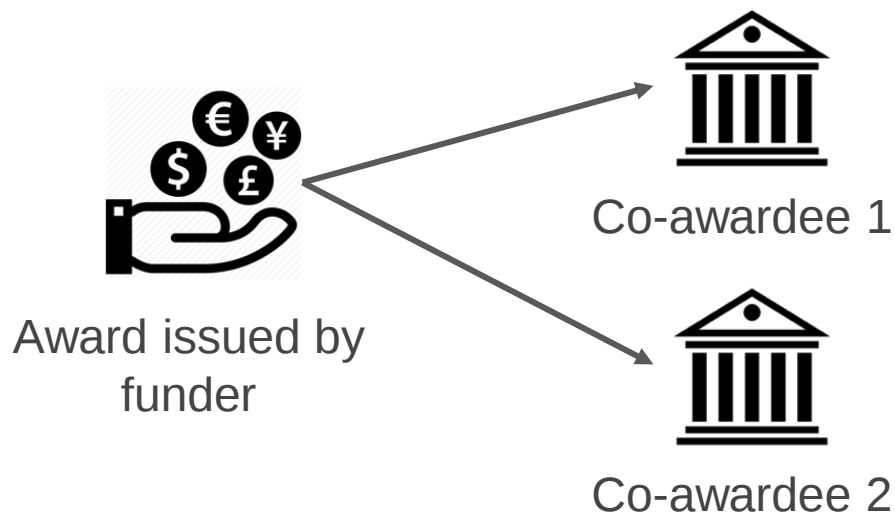
We look at three major quantifiable components of research activity



| Number of grant awardees

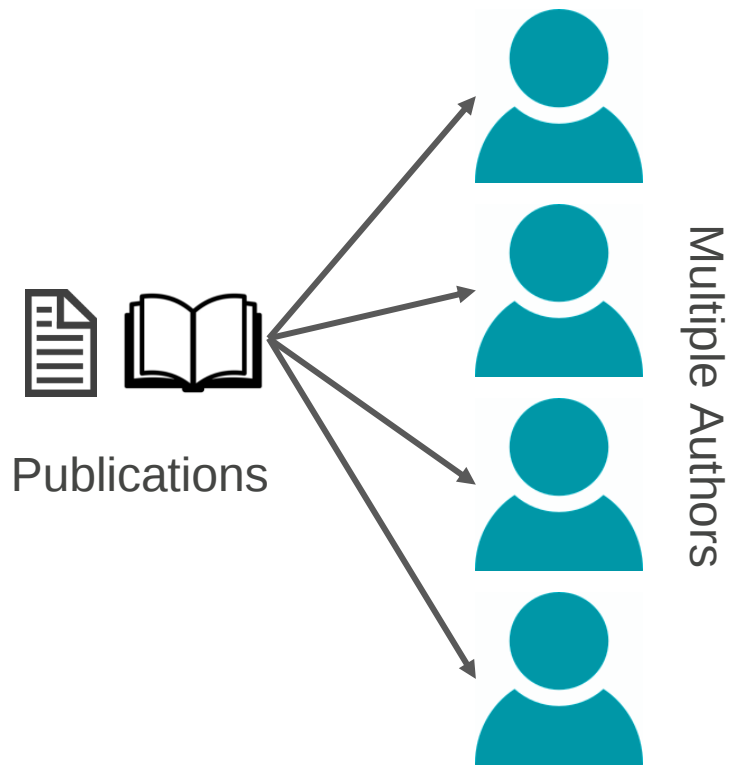
We need to know how many grants were awarded to each institution

E.G.: A grant that is awarded to two institutions is counted twice



- Each institution must manage its own grant information
- We therefore count once for each institution supported by each grant

Publication data



- Each author has reporting responsibilities, resulting in duplication of effort
- We therefore count once for each author at each institution

Estimating the number of projects

- No global or national databases exist, so we estimate
- Assumption: Number of projects is proportional to the size of the research environment

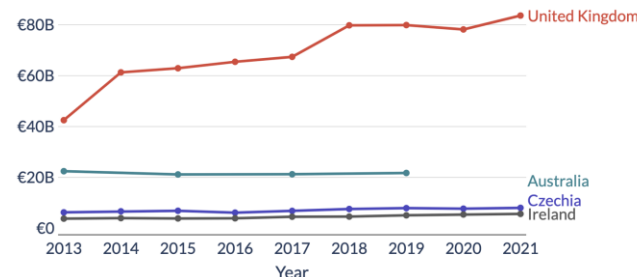


Dr Simon Kerridge, DProf
Director of Research Services, University of Kent
Chair, Association of Research Managers and
Administrators (UK)

How do we estimate?

- Initial estimate of 50,000 projects in the UK from Simon Kerridge
- Scaled by total R&D spend according to OECD*

$$(N \text{ projects in UK}) \times \frac{(\text{R\&D spend in country of interest})}{(\text{R\&D spend in UK})}$$



How many times does data get rekeyed?

MOREBRAIN *Research*

Traditional research outputs (articles, books, etc)



When a scholar at your institution publishes an article, review, book etc, is that information recorded centrally by the institution?

☐ Yes

☐ No

☐ In some cases

If you answered "in some cases". For what types of outputs, or under what conditions would that information be recorded centrally?

Your answer

research@morebrains.coop

- Research shows that project data is rekeyed six times*

We asked repository managers at universities across Australia about grants and publications

- 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

| How much time and money does all this take?

Time to input grant and project metadata:

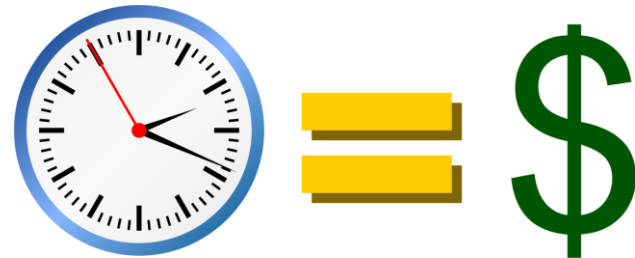
- 10m, based on previous third-party research *1

Time to input publication metadata:

- 6m 43s, based on previous third-party research *1

We work with national data on payscales, tax, and overheads to estimate full costs for:

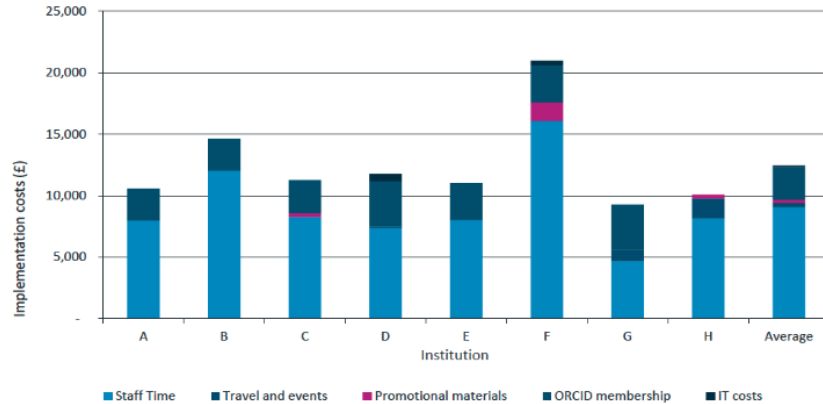
- Research administrators
- Senior researchers (professors)
- Junior researcher (post-docs)



*1 Klausen (2017) <http://dx.doi.org/10.1016/j.procs.2017.03.011>

*2 R. Johnson, H. Henderson, and H. Woodward, 'Institutional ORCID Implementation and Cost- Benefit Analysis Report', Jisc-ARMA, Jul. 2015. [Online]. Available: <https://doi.org/10.5281/zenodo.1445290>

Savings offset against implementation costs



Jisc/ARMA ORCID pilot estimated*

1 ORCID integration = 80 person-days

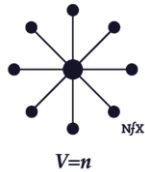
However, much of the costs are in outreach, education, technical development

The creation of national-level central support services create economies of scale

*R. Johnson, H. Henderson, and H. Woodward, 'Institutional ORCID Implementation and Cost-Benefit Analysis Report', Jisc -ARMA, Jul. 2015. [Online]. Available: <https://doi.org/10.5281/zenodo.1445290>

Benefits depend on the proportion of adoption

Sarnoff's Law



Metcalf's Law

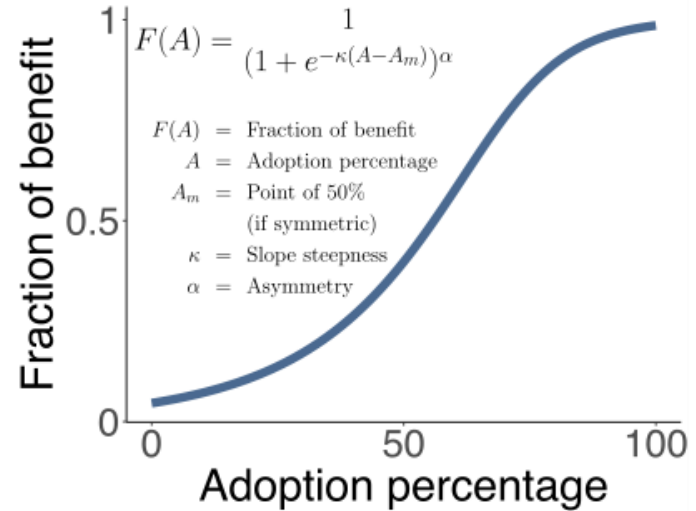


Reed's Law



Source: NFX*Bible

- PIDs are network entities, with network effects
- The value of adopting PIDs increases when other organisations also do so
- We use logistic (lazy S) function to model the relationship between adoption level and value*



* D. Kucharavy and R. De Guio, 'Application of S-shaped curves', Procedia Eng., vol. 9, pp. 559–572, 2011, doi: 10.1016/j.proeng.2011.03.142

Calculating the final benefits

	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Total
Potential savings by not rekeying publication, grant, and project metadata	£18,976,585	£18,976,585	£18,976,585	£18,976,585	£18,976,585	£18,976,585	£94,882,924
Adoption target (estimated current level of 18%)	18%	20%	40%	67%	76%	85%	
Percentage of benefits based on 'S-curve'	10.2%	11.1%	26%	71%	85%	93%	
Adjusted savings by not rekeying publication, grant, and project metadata	£2,099,238	£4,996,402	£13,390,246	£16,063,525	£17,688,878		£54,238,289
Costs for the PSN	(£809,894)	(£890,883)	(£979,971)	(£1,077,968)	(£1,185,765)		(£4,944,481)
Sector-wide implementation costs for institutions when supported by the PSN	(£1,541,523)	(£1,622,513)	(£1,711,601)	(£1,809,598)	(£1,917,395)		(£8,602,631)
Sector-wide net savings	£1,930,179	£557,715	£3,373,890	£11,678,644	£14,253,927	£15,771,483	£45,635,658

**Taken from the UK PID CBA*

<https://doi.org/10.5281/zenodo.4772627>

- The more organisations that adopt PID-enabled workflows
 - ⇒ the more data is available
 - ⇒ the more benefit to adoption

Summary

- We can make the case for PIDs and national / regional interventions by looking at policy priorities
- National policy priorities often include open research, impact tracking and reducing bureaucracy
- Cost-Benefit analyses show that interventions to support PID adoptions result in positive returns on investment
- National PID strategies are valuable ways to drive collective action



Small group discussion

- What PIDs or PID use cases are you most interested in?
- Who are the stakeholders?
- What are the obstacles?
- What are the winning approaches?



Co-funded by
the European Union



Panel discussion



John Aspler
Manager, Canadian
PID Community
Canadian Research
Knowledge Network



Gaëlle Béquet
Director, ISSN
International Centre



Phill Jones
Co-founder,
Digital & Technology
MoreBrains
Cooperative



Natasha Simons
Director (National
Coordination)
Australian Research
Data Commons

11–13 June, Prague, Czechia



Q&A