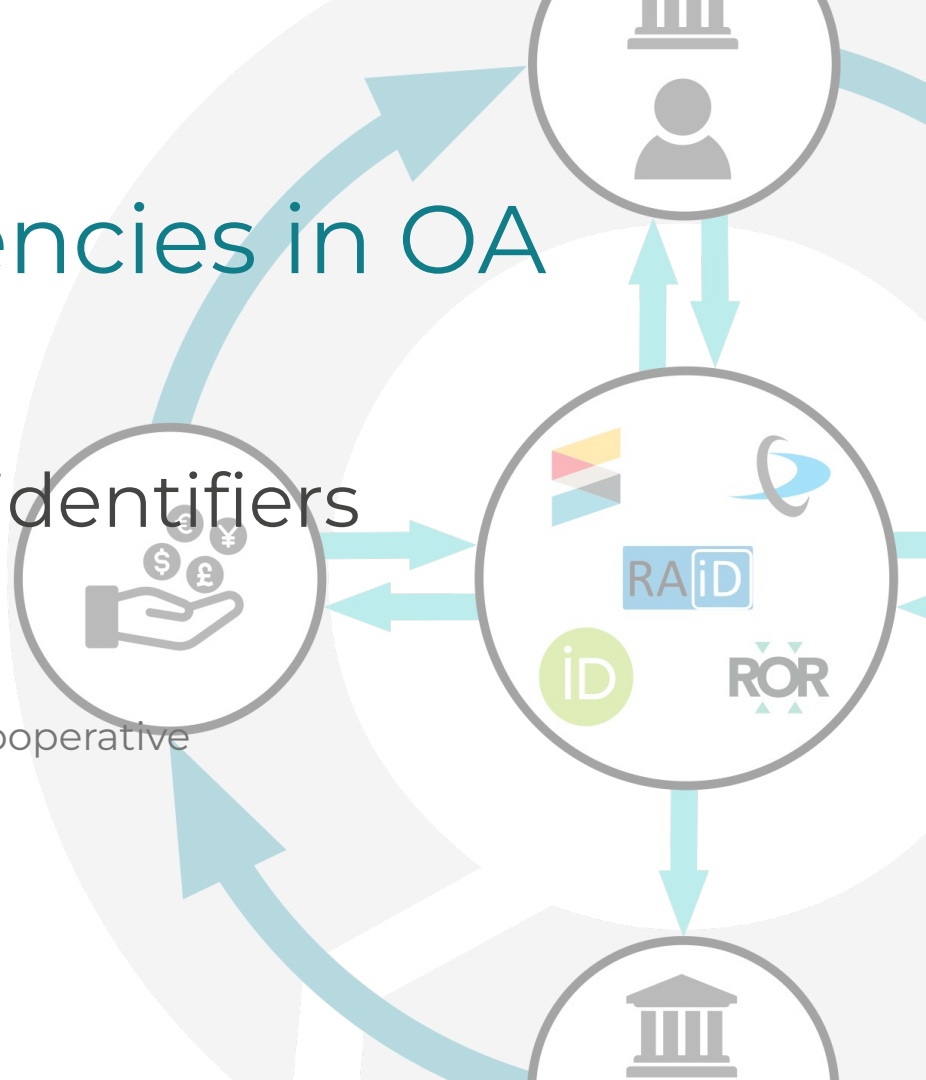
Unserved dependencies in OA workflows

A case study: persistent identifiers

Josh Brown
Co-founder, research & strategy, MoreBrains Cooperative

MOREBRAINS



Acknowledgements

The workflows shown in this presentation were developed as part of a project commissioned by Jisc, and sponsored by Research England.

We would also like to thank the 78 experts from the worlds of research funding, data management, institutional research management and administration, repositories, and publishing who took part in workshops or reviewed these workflows.

Jisc

UKRI Research England

MOREBRAINS

Why use PIDs as a case study?

*"This is all about **shared infrastructure**, standardised messaging protocol, and integration with existing systems, **powered by Persistent Identifiers (PIDs).**"*

*"**PIDs are indispensable**."*

*"The biggest insight was that **PIDs in article-level metadata are critical for solving real problems**. And, that it is not too late to implement PIDs."*

Yvonne Campfens, OA Switchboard.
https://www.oaswitchboard.org/blog8dec2021

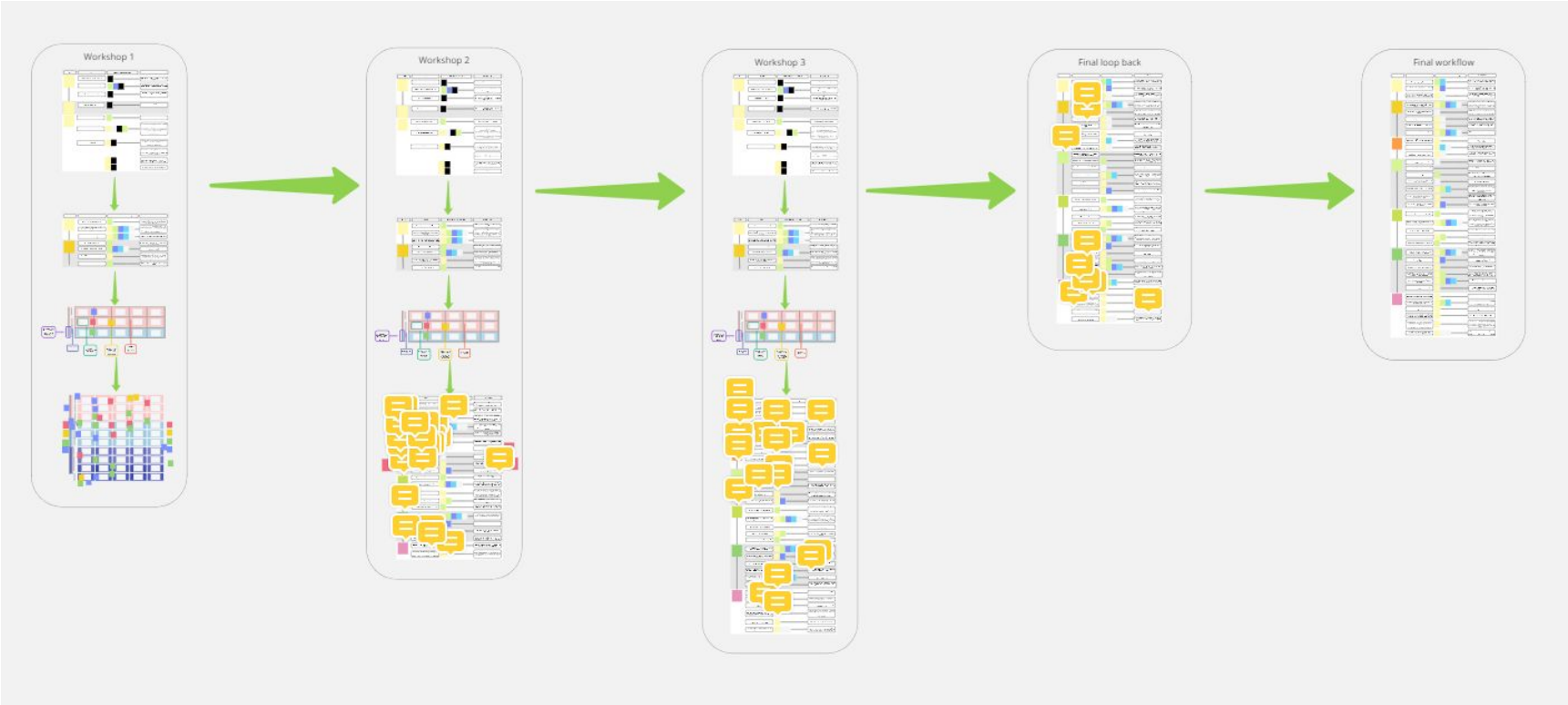
Priority workflow development

We developed four ‘ideal world’ PID-optimised workflows, beginning with an initial model co-created with our UK partners, and then validated internationally:

- Funding
- Content publication
- Research data
- Institutional research management

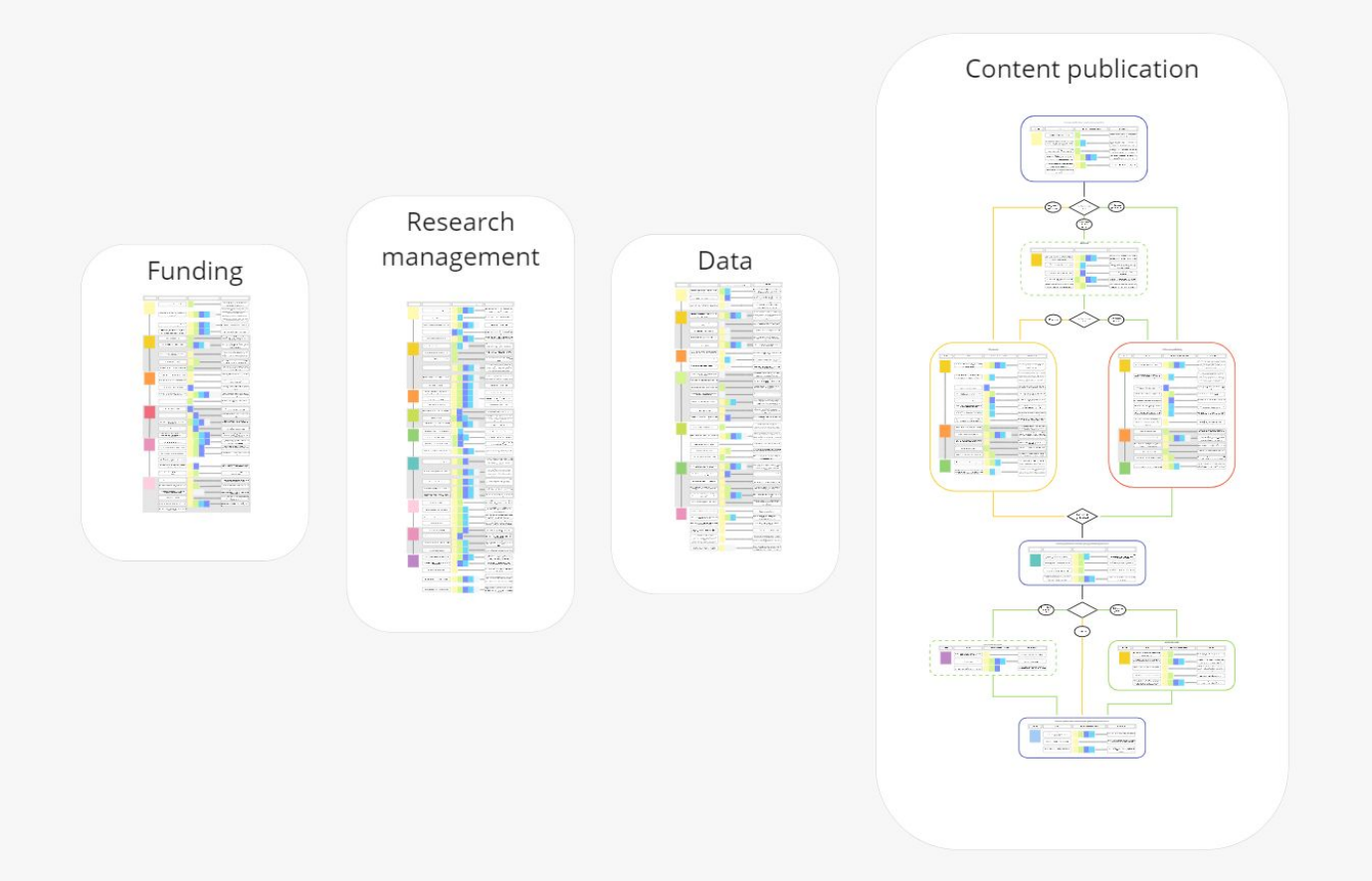
MOREBRAINS

A snapshot of our workflow development



MOREBRAINS

The workflows

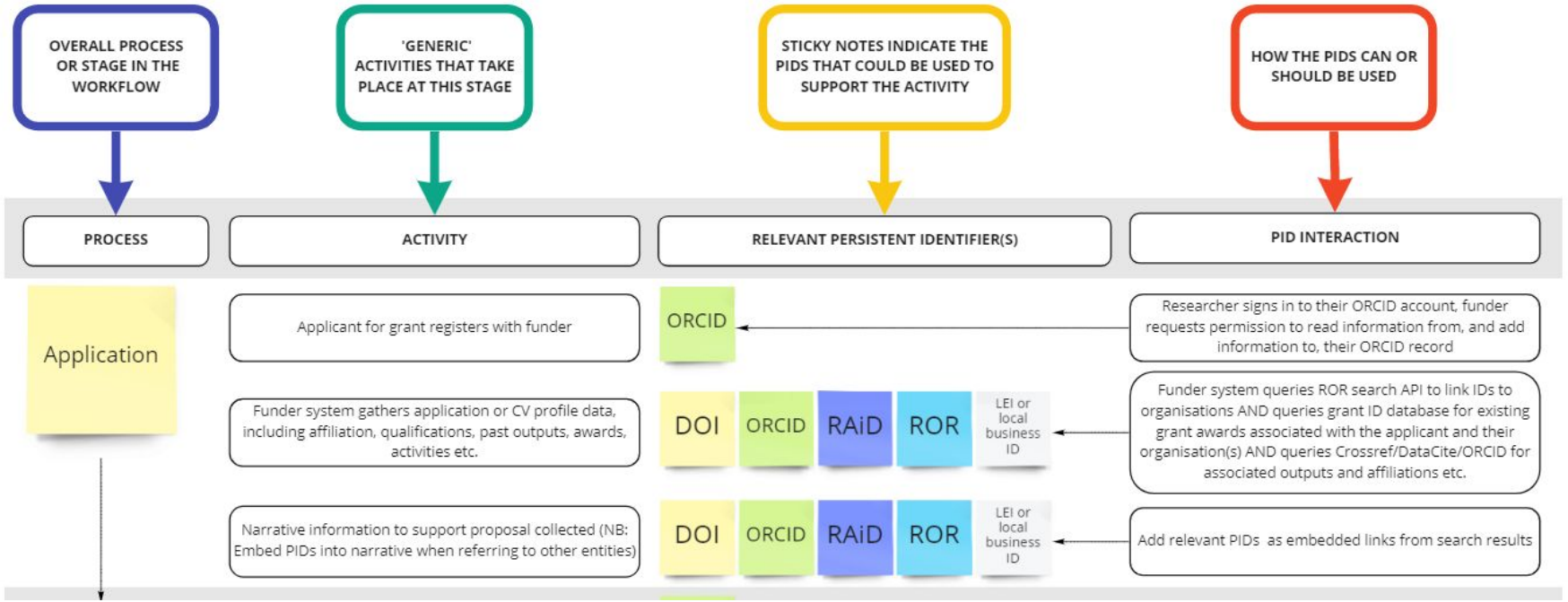


The four workflows cover each stage in the research lifecycle, from developing a hypothesis to project impacts

MOREBRAINS

How the PID-optimised workflows are designed

Our approach is to link activities to specific ‘PID optimisations’ at every step



MOREBRAINS

Which PIDs optimise these workflows

In line with the UK national PID strategy[1], we have focused this analysis on five 'priority PIDs':

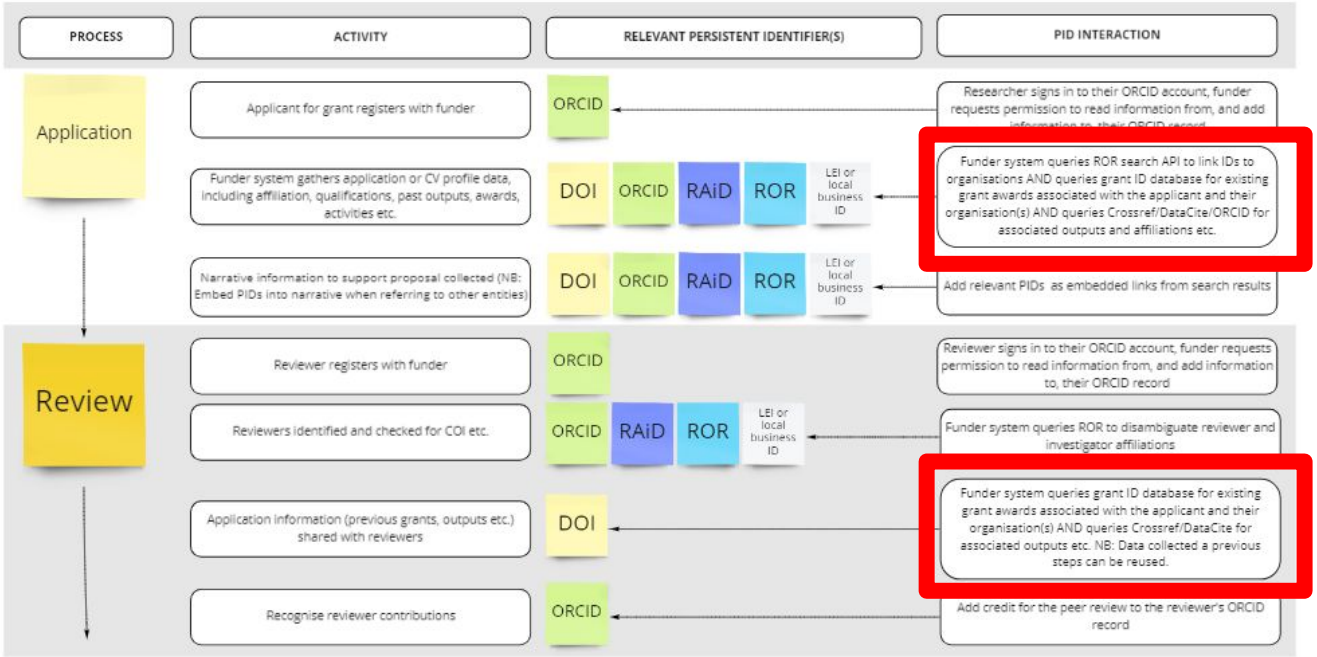
- Grants (Crossref DOIs)
- People (ORCID IDs)
- Projects (Research Activity Identifier - RAiD)
- Organisations (Research Organization Registry - ROR)
- Outputs (Crossref and DataCite DOIs)

Note that these are all **open infrastructures** providing **FAIR metadata**

1) https://repository.jisc.ac.uk/7840/

MOREBRAINS

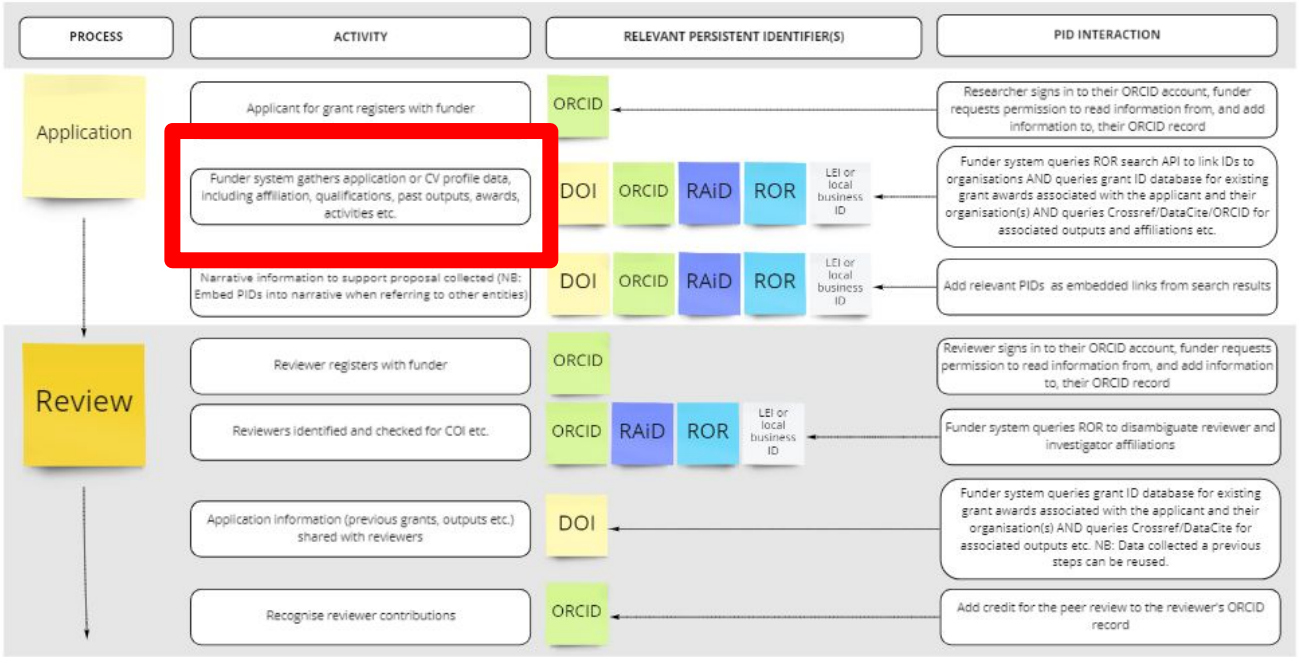
Learning from the PID-optimised workflows



One principle we identified is that collecting and sharing information at the earliest possible stage maximises efficiency AND transparency

MOREBRAINS

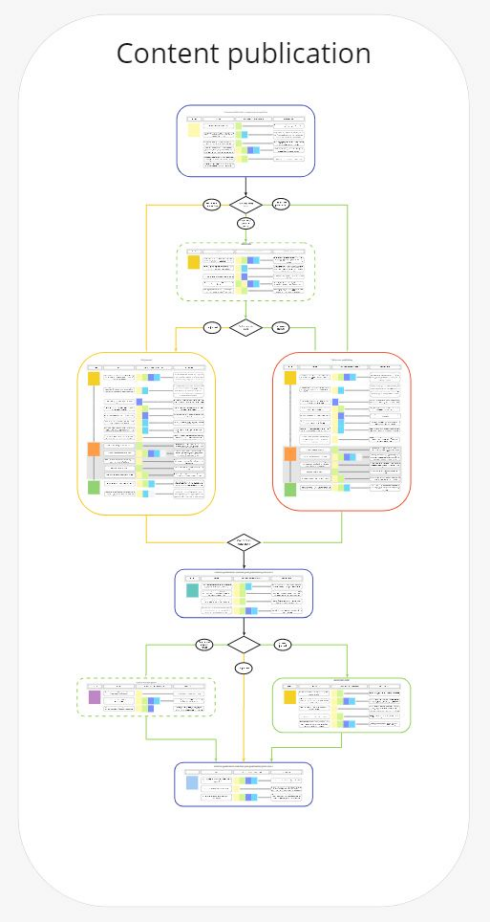
Learning from the PID-optimised workflows



The workflows make it very easy to identify dependencies on PID systems and actions taken by other players in the ecosystem

MOREBRAINS

PIDs in various open access publication workflows

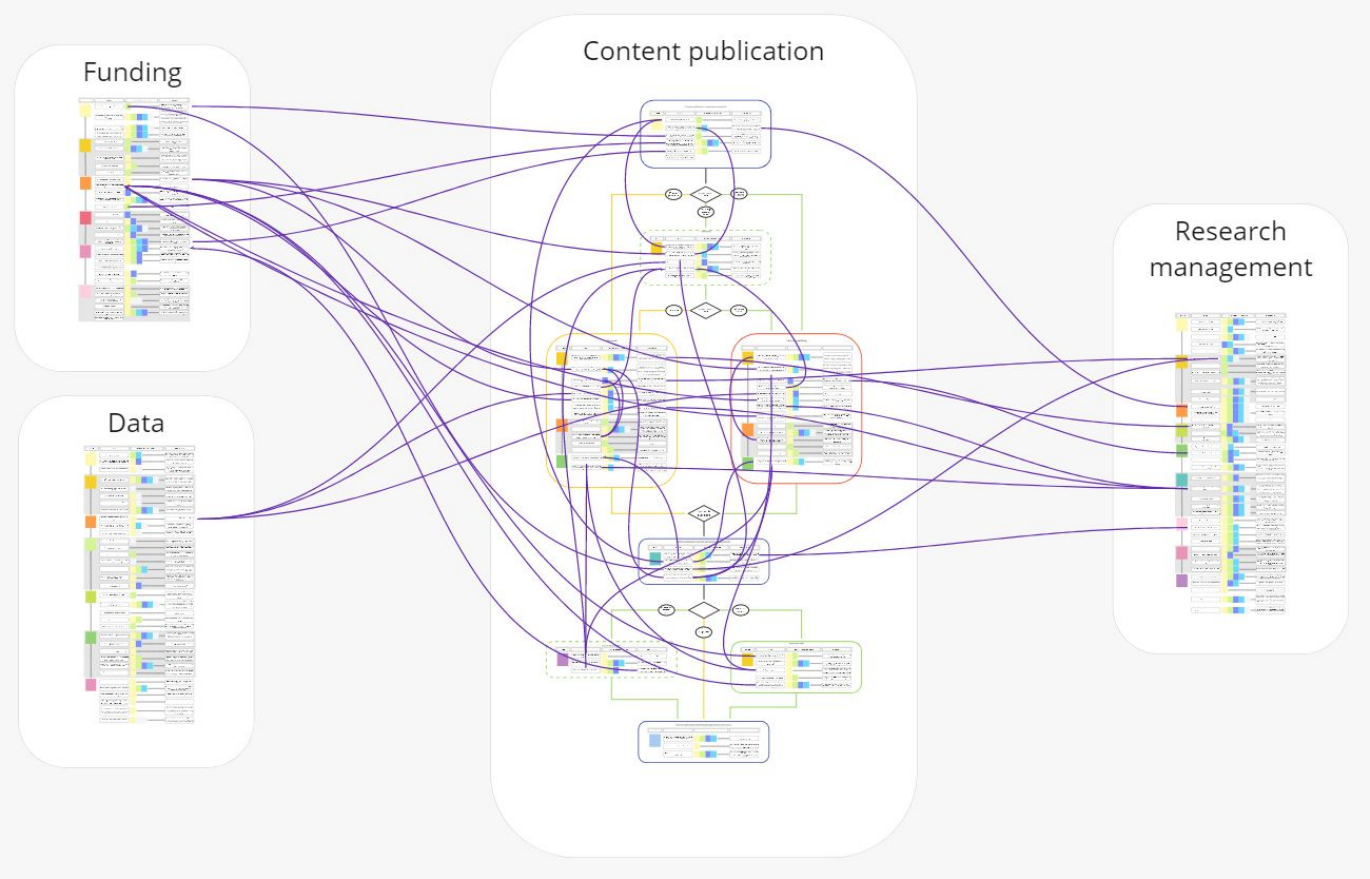


This is our content publication workflow.

It takes in pre-print servers, Gold OA journals, toll-access journals, and institutional/subject repositories.

MOREBRAINS

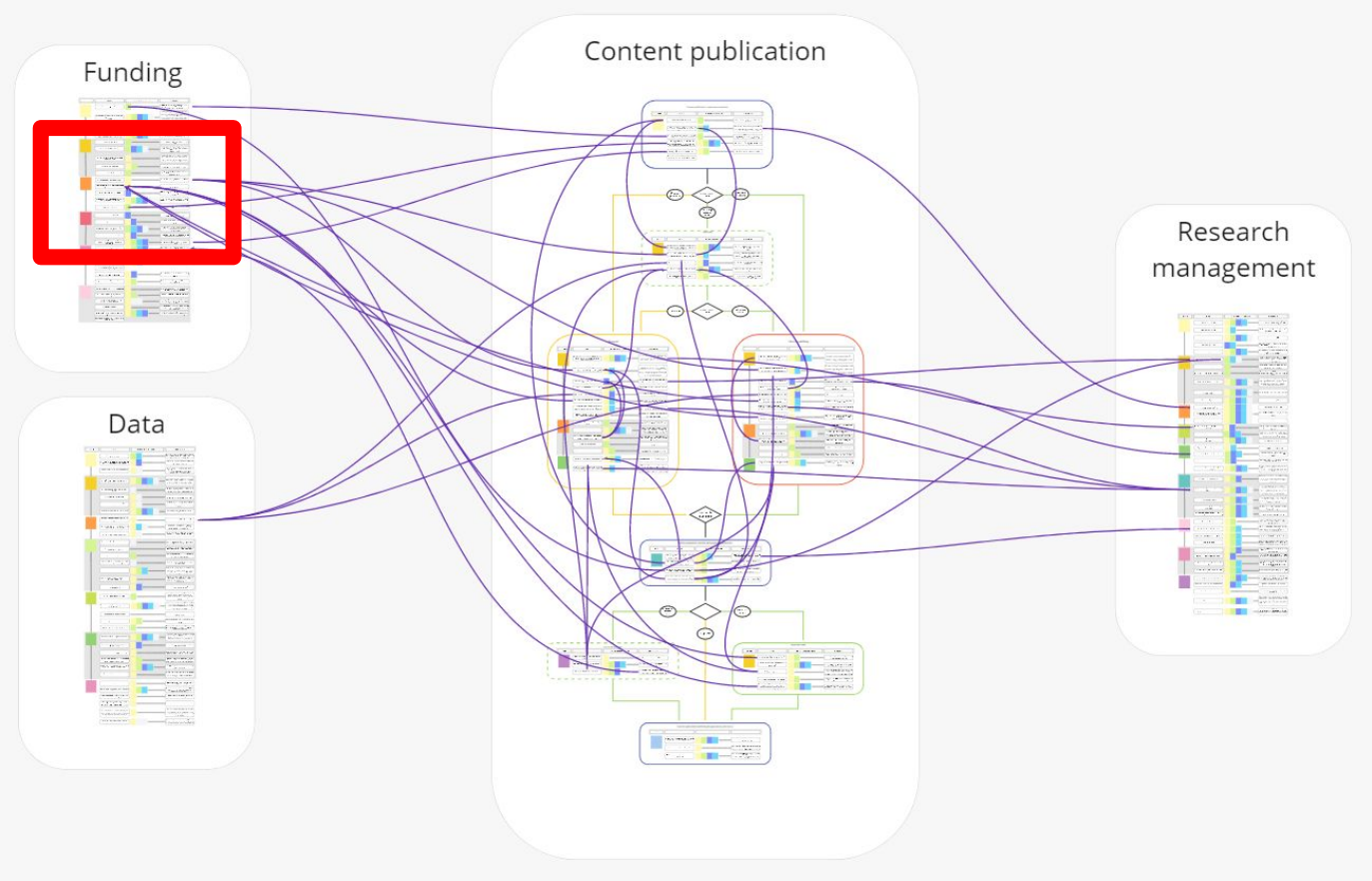
Mapping dependencies across workflows



Every purple arrow represents a dependency on an action taken earlier in the publication process, or on a funder or institutional action.

MOREBRAINS

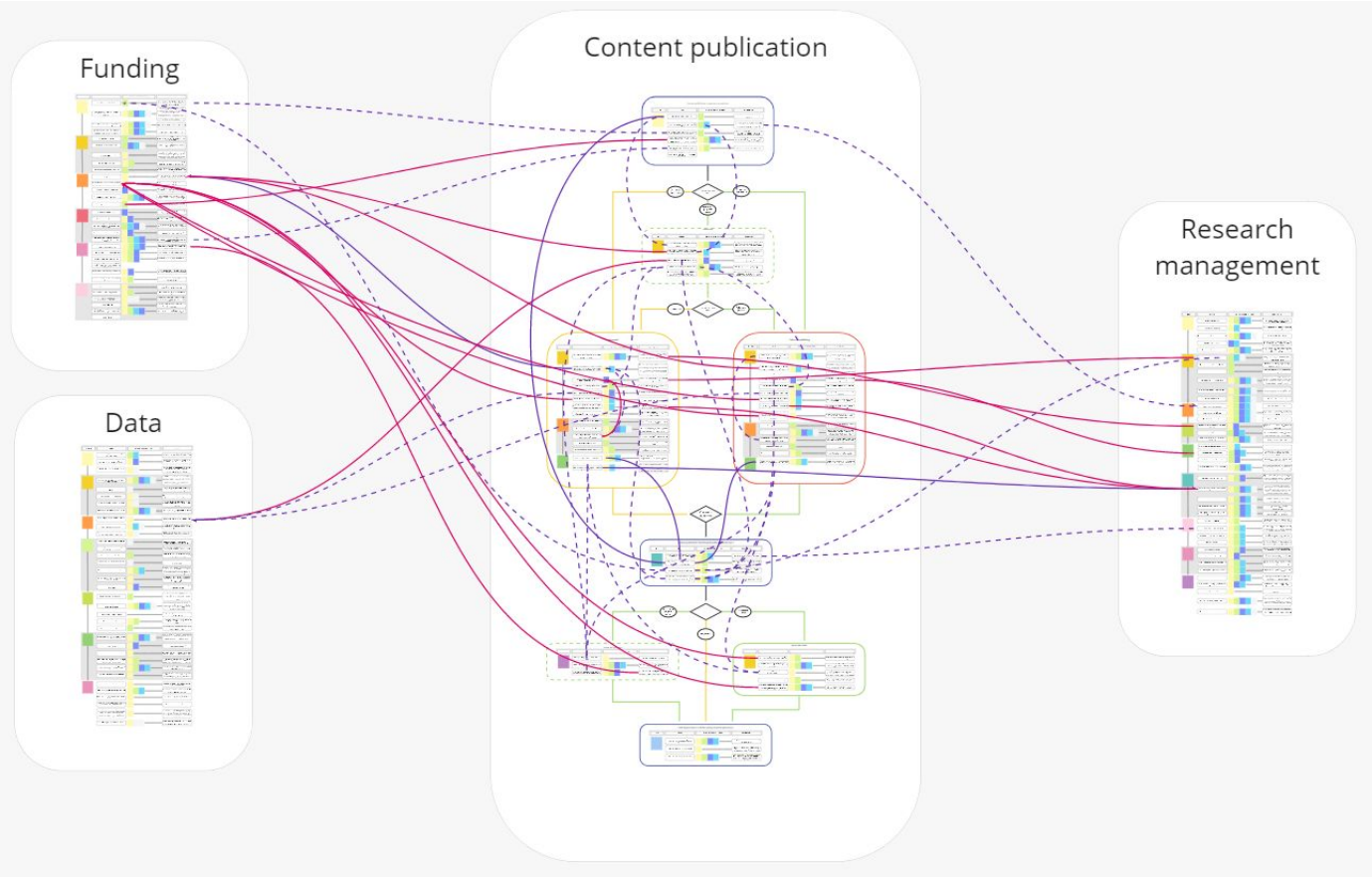
Mapping dependencies across workflows



Here is a critical dependency: funders registering a DOI for each grant, and linking it to actionable information about their OA policies etc.

MOREBRAINS

Mapping dependencies across workflows



Dotted lines show dependencies which are *sometimes* served.

Red lines are missing.

MOREBRAINS

The negative

Every red or dotted line represents manual data entry, multiple emails back and forth to request and share information, and a lot of frustration

The negative

Every red or dotted line represents a waste of time, money, and expertise

The positive

These dependencies could all be served using PIDs, services, and metadata that ***already exist***

The take home messages

1) Everyone benefits IF AND ONLY IF everyone else does their bit

2) We need better standards for metadata exchange between PID registries - because the benefits come from reusing that metadata

3) We need to work collectively to lower the cost and difficulty of implementing these systems - large and small players need to be able to leverage these critical infrastructures (see point one above!)

MOREBRAINS