

The State of Scholarly Metadata: 2023



In late 2022, CCC and Media Growth Strategies undertook a thorough examination of metadata management across the research lifecycle.

This in-depth review builds on an existing body of work to uncover multiple policy and system complexities and breakages, which – separately and together – create missed opportunities for the communities for whom Open Access (OA) and Open Science models are designed to serve.

CCC is sharing this analysis with the scholarly communications community to spark dialogue and to drive action. Drawn directly from our research interviews, this living infographic depicts the significant economic and social impact that a fragmented metadata supply chain has today on researchers, institutions, funders, and publishers. Researchers, in particular, shoulder a significant administrative burden that ultimately disrupts and delays the process of scientific discovery.

As the scholarly communications community continues its shift to full OA, stakeholders recognize that new strategies, inclusive policies, and a robust network of interoperable data and systems are essential for making critical infrastructure improvements, and much progress is underway. In that environment, a dedication to data stewardship across each stakeholder group, and the service providers supporting them, will lead not only to a smoother OA transition, but also to greater research integrity; data sharing; reliable, trustworthy metrics on research impact; and a responsive, equitable rewards and recognition system.

<p>Research stage Idea Development</p> <p> RESEARCHER Researcher seeks collaborators; meets with colleagues and library / research office staff</p>	<p>CHALLENGES</p> <p>Underutilization of ORCID Some institutions don't require researchers to use ORCID; records can be outdated if authors don't consistently update; ORCID may not be accessible to authors in some geographies.</p>	<p>IMPACT</p> <p>If authors can't be identified with a standard ID, they may not be able to authenticate to content, get credited appropriately for their work, secure OA funding, or complete downstream processes without unnecessary manual effort. Costly manual effort is also required of publishers, institutions, and funders to disambiguate authors retrospectively.</p>
<p>Research stage Proposal Submission</p> <p> RESEARCHER Researcher submits application for funding</p> <p> FUNDER Funder selects reviewers and begins application review</p> <p> FUNDER Funder logs funding terms in grant management system</p>	<p>CHALLENGES</p> <p>Inconsistent Metadata Capture Variability across grant application process/systems results in possible loss of metadata necessary to determine OA funding entitlements at a later stage, e.g., institutional affiliations.</p> <p>CHALLENGES</p> <p>Legacy System Limitations Low adoption of standardized PIDs (FundRef, RAiD, Ringgold, ISNI, ROR) due to limitations of legacy systems and/or lack of awareness.</p> <p>CHALLENGES</p> <p>Low-Quality Data Free text fields are great for gathering feedback; they're not designed to capture granular data like an organizational identifier. Researchers often confuse proposal numbers with grant IDs later in the publication process--they need structure to improve the accuracy of data capture.</p>	<p>IMPACT</p> <p>Without disambiguated grant and funder details, grants may not be effectively utilized in later publication stages, leaving OA funding unclaimed and shifting coverage to research institutions. In an ecosystem that values a sustainable OA shift, this impacts everyone.</p> <p>IMPACT</p> <p>Hindered conflict of interest management among peer reviewers threatens research integrity, and low-quality data results in low accuracy of later-stage funding identification, tracking, and analysis of research output.</p> <p>IMPACT</p> <p>Lack of registered grant DOIs makes it difficult and costly to link funding to particular research outputs, resulting in missed OA opportunities as well as incomplete analysis to inform future funding investments.</p>
<p>Research stage Research & Authoring</p> <p> RESEARCHER Researcher conducts literature review</p> <p> RESEARCHER Researcher posts pre-print / shares early outputs</p> <p> RESEARCHER Researcher selects publication for submission</p>	<p>CHALLENGES</p> <p>Researcher Inequities & Research Barriers</p> <ul style="list-style-type: none"> Valid research coming from under-represented researchers is hard to find due to lack of metadata, including DOIs. Search and discovery are difficult due to inconsistency in identifying the user and enabling appropriate access to research. Authors from under-represented areas may not have equitable access to search and discovery services or equitable opportunities for publication. <p>IMPACT Global inequities hinder scientific progress.</p> <p>CHALLENGES</p> <p>Poor Connections Across Research Outputs Lack of persistent identifiers (PIDs) and inconsistent application of PIDs across research outputs e.g., data sets, equipment, setting(s), samples, software</p> <p>CHALLENGES</p> <p>Risk of OA non-compliance Metadata lost upstream makes managing funding compliance onerous.</p>	<p>IMPACT</p> <p>Inability to easily find, verify, and reuse the data and artifacts underlying research, making it difficult to accurately interpret, cite and reproduce research findings.</p> <p>IMPACT</p> <p>Lack of available information about both corresponding author and all co-authors leads to manual input to identify funder and institutional mandates at best and missed funding requirements at worst.</p>

Research stage
Publication

RESEARCHER
Researcher submits article

INSTITUTION
Institution funds OA publication

PUBLISHER
Publisher indexes metadata to enable search & discovery

CHALLENGES

Missed Funding Opportunities

- Under-utilization of metadata validation services
- If the researcher has submitted before, outdated information from their existing profile can be pulled into the submission
- Inconsistency between journal policies and metadata procedures
- Lack of funding information captured at submission and validated at acceptance
- Demand for increased interoperability between IDs

CHALLENGES

Missed Funding Opportunities & Costly Billing Complications

If funder/institution information manually input by the author does not use a standardized name or PID (e.g., abbreviations, nicknames), this can interfere with matching to the correct OA funding source.

CHALLENGES

Unnecessary Manual Intervention

Publishers are sometimes manually entering PIDs prior to registering DOIs for a more complete publication record.

IMPACT

Without granular, accurate organizational affiliation identifiers for a manuscript, coupled with incomplete funding details, authors may miss the opportunity to get OA funding or miss the chance to opt into OA due to affordability concerns. OA initiatives driven by institutions and funders may lack uptake as a result. Publishers are also unable to automate processes that reduce the cost of business model transformation. Manual effort is required to retrospectively cover the publication with proper funding sources, driving up the cost of publishing. No one benefits in this scenario.

IMPACT

Publishers and institutions take on the time and expense of manually finding the papers that should have matched to an agreement and collaborating on a resolution. Funding decisions cannot be based on abbreviations or free-form data.

IMPACT

This is a laborious practice with high economic and opportunity costs that could be reduced with earlier, automated PID assertion and/or validation.

Research stage
Reuse & Measurement

RESEARCHER
Researcher evaluates research impact

INSTITUTION
Institution assesses historical subscription & publication data to inform institutional deals

FUNDER
Funder evaluates research impact

PUBLISHER
Publisher assesses historical subscription and publication to inform institutional deals

CHALLENGES

Problematic Research Impact Measurement

Difficult to track research/researcher impact due to lack of adoption of metadata standards.

CHALLENGES

Problematic Deal Modeling

- Lack of consistent affiliation and funding data makes modeling future agreements hard for institutions.
- Data is not standardized across publisher platforms, creating unnecessary manual work to gather and normalize data for analysis.

CHALLENGES

Problematic Research Impact Measurement

Difficult to track funder impact due to lack of adoption of metadata standards.

CHALLENGES

Problematic Deal Modeling

Lack of consistent affiliation and funding data makes modelling future agreements difficult for publishers and institutions.

IMPACT

Researcher rewards and recognition decisions, or future opportunities for funding, may be based on incomplete or inaccurate data, affecting reputation and career advancement.

IMPACT

The transition to modern models of OA publication is delayed, prolonging a mixed-model landscape and the availability of open outputs to advance science.

IMPACT

Incomplete analysis to support future funding investments and to report activities to the public.

IMPACT

The transition to OA is delayed, putting some publishers at risk of losing authors to funding mandates and losing revenue that is necessary to sustain operations.

To view the interactive map, visit stateofmetadata.com

About CCC

A pioneer in voluntary collective licensing, CCC (Copyright Clearance Center) helps organizations integrate, access, and share information through licensing, content, software, and professional services. With expertise in copyright, information management, artificial intelligence, and machine learning, CCC and its subsidiary RightsDirect collaborate with stakeholders to design and deliver innovative information solutions that power decision-making by harnessing information from a wide variety of data sources and content assets.

