Preserving the Integrity of the Scholarly Record: Insights from the Crossref Community

Madhura S. Amdekar, Ph.D
Community Engagement Manager, Crossref

Crossref

What is scholarly <metadata>?

→ **Basic:** titles, dates, author names, abstracts, DOI, location URL
→ **Full-text URLs:** e.g. for text-mining and Similarity Check
→ **Crossmark:** updates, retractions, corrections
→ **Relationships:** versions, translations, data, references, citation
→ **Provenance:** publisher/funder/steward information
→ **Subject-specific:** e.g. clinical trial info
→ **Funding information:** Funder Registry ID, award numbers/Grant DOIs
→ **Contributor & Affiliations:** ORCID iDs and ROR IDs preferred

Metadata and integrity of the scholarly record (ISR)

- Scholarly record = research and published outputs + relationships between these outputs, inputs, relationships, and contexts.
- Preserving the integrity of the scholarly record (ISR) is a key component of research integrity: correct, complete, and comprehensive metadata about scholarly outputs can act as signal of their trustworthiness.
- Metadata enables assessment of trustworthiness by providing important context about work
  - Who authored the work
  - Who was it funded by
  - Which other work does it cite
  - Was it updated after publication
- Crossref provides infrastructure that enables the scholarly community to provide and deposit metadata about the content that they produce.
- We have been engaging with our member communities to identify key metadata elements that are trust signals.

What have we been up to..

- We organised in-person roundtable discussions with the scholarly community in 2022 and 2023, and an online discussion in 2024.
- The participants included publishers, research integrity specialists, funders, policy-makers, researchers, institutions, and other organisations such as COPE.

RESEARCH NEXUS

Relationships that connect research organisations, people, things, and actions:

→ HOW is paper A and dataset A connected?
→ WHO contributed to the work?
→ WHICH funding program supported the work?

Here's what the community said:

What metadata is important for signalling trust?

1 Retractions and Corrections

Inform readers whether the work that they are reading or citing has been updated.

Useful to add: reasons for retractions, and submission dates of retracted papers.
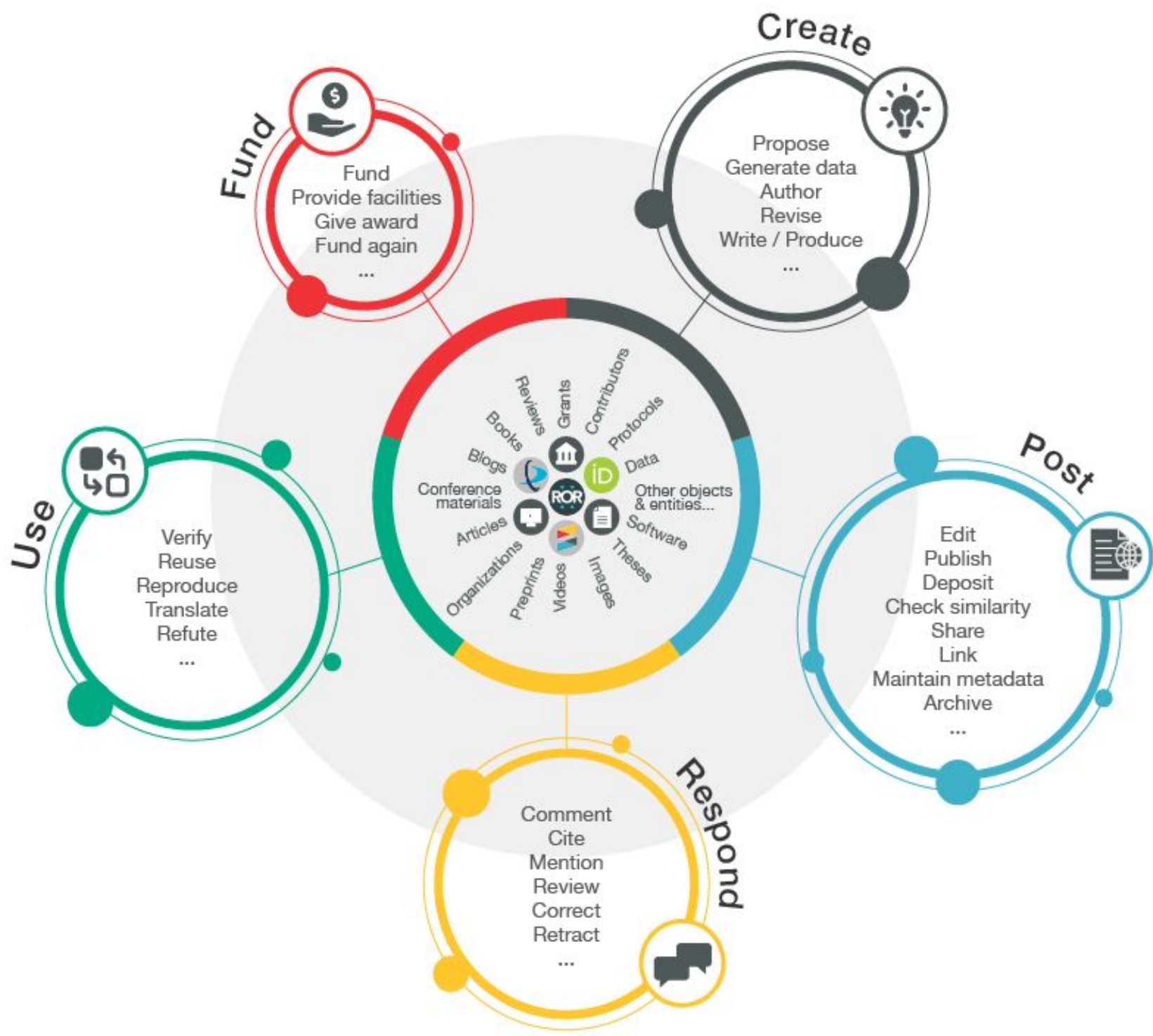
2 Abstracts

Improve discoverability of content.

Enable text mining and systematic reviews for large-scale analyses.

3 References

Important for research assessment.

Provide context to the research and point readers to the sources of your content.

4 Affiliations

Connect authors and research outputs to institutions.

Institutions can measure research produced by researchers.

What metadata is "nice to have" in the scholarly record?

1 Peer review information

Identity of peer reviewers
Identity of handling or decision-making editors
Corresponding author

2 Special Issues

Identity of guest editors and editorial boards
Submission and acceptance dates of articles

3 Adding transparency to research

Ethics approvals
Clinical trials
Conflicts of interest
Better links between preprints and published articles

What can you do:

Researchers
- Include important metadata along with your journal submissions: Crossref grant IDs, ORCID iDs, and ROR IDs.
- Check Crossmark for updates.

Research Institutes
- Use affiliation identifiers- ROR IDs.
- Data can be used to identify organisations with paper mill activity.

Publishers
- Use Crossmark: add retractions and other post-publication updates.
- Use Participation Reports to identify gaps in your metadata.

Editors
- Update journal policies to collect and deposit key metadata.
- Join community.crossref.org for knowledge sharing.

More about ISR at: crossref.org/blog | We want to hear from you: mamdekar@crossref.org, feedback @crossref.org