

## OSMI WG4 – July 2025

### Participation and Representation Challenges

- Some participants felt excluded from the survey because their role is project-based and not tied to a single institution with formal data-sharing practices.
- Uncertainty over “which hat to wear” when filling out the survey was a barrier.
- There is currently low representation from the Global South in survey responses and in OSMI generally.
- Only about 50% of the current group has responded so far (21 of ~40 members).

### Survey Design and Scope Issues

- The survey didn't capture participants' involvement in *other open infrastructure initiatives*—an omission flagged as important (e.g. Comet, other projects).
- Difficulty for project-based or cross-institutional people to respond meaningfully suggests that the survey may need revision or clearer guidance on intended perspective.

### Potential Value of Broader Outreach

- Several participants suggested sharing the survey more widely:
  - Across other OSMI working groups.
  - Possibly outside of OSMI (e.g. Force11 suggested in chat).
- Concerns were raised about timeline impact if broader distribution delayed next steps.
- Agreement that even if we can't get “perfect” representation, more diverse input (especially from Global South) is still valuable for understanding different contexts and needs.

### Next Steps for Survey Analysis

- A. suggested comparing Global South vs. European responses, even if numbers from the South are small—these differences could inform benchmarking and tool

- There was debate about whether to proceed in parallel (continue analysis now) vs. waiting for more responses.
- Consensus seemed to lean toward not delaying work indefinitely for new responses, but remaining open to adding more data.

## **Importance of Metadata Standards in Publishing**

- S. emphasized how existing publishing standards (e.g. from NISO) affect data mining:
  - Some metadata standards are stronger than others, impacting ease of mining.
  - Not all publishers have implemented even relatively new requirements (e.g. standard Data Availability Statements).
- S. suggested this group could consider advocacy or action items around pushing adoption of such standards.
- She offered to share resources (e.g. AGU report, video on “leaky pipes” of data citation).

## **Technical Approaches to Metadata Improvement**

- T. shared work from Comet using NLP and tools like Selenium to infer or “heal” missing data citations.
- Highlighted the complementarity of fixing publisher metadata vs. repairing citations downstream using language models.
- Emphasized the value of mapping where metadata breaks and exploring technical solutions.

## **Coordination with Other Groups**

- There’s recognition that other OSMI working groups may already be collecting related information.
- Suggestion that cross-group coordination is needed to avoid duplication and ensure alignment.

- A. suggested that the question of expanding outside OSMI may need to be raised in broader group leadership meetings.

## **Immediate Action Ideas and Suggestions**

- Consider revising the survey or giving clearer guidance for people in cross-institutional or project-based roles.
- Explore further dissemination of the survey within and possibly beyond OSMI, balancing benefits with timeline impacts.
- Analyze existing responses now, while staying open to new inputs later.
- Coordinate with other groups to share data and reduce overlap.
- Leverage external standards (e.g. NISO) and ongoing projects (e.g. Comet) to inform recommendations or actions.