

## Solution Session

# How FAIR-R Is Your Data? Enhancing Legal and Technical Readiness for Open and AI-Enabled Reuse

*Katharina Miller, Vanessa Guzek*

Organization(s): Miller International Knowledge

### Abstract

The FAIR data principles (Findable, Accessible, Interoperable, and Reusable) are foundational to Open Science and data-driven research.

But are they enough in the age of AI? Increasingly, researchers and data stewards must also ensure their datasets are FAIR-R: not only technically reusable but also Responsibly licensed and legally open for machine learning and downstream applications.

This solution session invites participants to evaluate real-world datasets through the lens of FAIR-R compliance, with a particular focus on intellectual property, licensing, and AI-readiness.

Participants will work in small teams to conduct a quick “data readiness audit” of openly shared datasets. They will assess:

- Metadata completeness and clarity
- Licensing visibility and compatibility
- Reusability by humans and machines
- Legal clarity around reuse in AI workflows

By the end of the session, each team will produce a public-facing “Data Readiness Snapshot” that includes improvement suggestions and a simple legal checklist for researchers.

Challenge:

While FAIR guidelines address many technical aspects, legal ambiguity around data licensing—especially in AI training contexts—remains a major barrier to reuse. This session addresses common questions:

Can this dataset be reused for AI?

Is the license machine-readable and legally sound?

Are restrictions (e.g., NC, ND) clear or problematic?

Is the metadata sufficient to ensure traceability and attribution?

Initial Proposed Solution:

The session offers a practical method to:

- Rapidly assess open datasets using a lightweight FAIR-R rubric.
- Identify legal and technical gaps that limit reuse and AI integration.
- Produce actionable “audit snapshots” and collective recommendations.

Selected audit examples can be published on Zenodo or institutional repositories under CC-BY 4.0, with attribution to participants and dataset creators.

### Session Plan

- 10 min – Introduction: Quick overview of FAIR vs. FAIR-R, licensing red flags, and AI-specific reuse barriers.
- 35 min – Team Activity: Teams audit sample datasets (or bring their own), using a provided checklist covering:
  - Metadata quality
  - Licensing clarity
  - Accessibility and format
  - Reusability in AI workflows
  - Legal interoperability
- 15 min – Report & Wrap-Up: Teams briefly share results and suggestions. A joint summary of best practices and frequent issues will be compiled live and shared post-conference.

Participants will receive a copy of the FAIR-R checklist and can reuse it in their institutions or projects.