

NSF CI Compass FAIR Topical Working Group

Meeting Notes

- Future speakers:
 - Sample Identifiers FAIROS & RDA - Natalie Raia and Andera Thormer
 - Data usage mining - Neil Byers

2026-06-02

- Group Discussion Meeting
- Christine Laney
 - Agent-aided data egress; data license changing from CC0 to CC BY, required authentication, and signed (expiring) urls. Auth only required for data downloads, not access to endpoints that provide metadata.
 - A publisher has said they don't want to accept a manuscript that refers to a dataset that is in a repository that requires a login.
 - Seems strange; repository would not share login information with outside groups.
 - Ted: Ryan from Dryad - having same problem, highly costly data egress, events can be recognized, they are also moving to authentication and using some tools to identify scraping.
 - David: Decentralization may be necessary in the future (bittorrent model?)
 - Don:
 - NSDF - have a caching layer on top of the data sources
 - There is a nature opinion piece from a year ago identifying this AI scraping problem.
<https://www.nature.com/articles/d41586-025-01661-4>
 - Chuck: <https://github.com/google-deepmind/science-skills>
 - NEON may want to create agentic skill that teaches the agent how to explore the data before downloading it.
 - Probably not malicious, may just be dumb agents that need to be taught to be more careful about resources.
 - Captchas not an alternative to authentication. AI can solve them.
 - Frame communications in that we're not trying to control access, just want to help end users.

- Claude opus 4.8 is looking for licensing prior to data use.
 - RSL (really simple licensing) - <https://rslstandard.org/>
 - <https://www.linuxfoundation.org/research/sbom-guide-for-enterprise>
 - Guidance from science on [schema.org](https://github.com/ESIPFed/science-on-schema.org/blob/main/decisions/47-use-spdx-licenses.md) group-
<https://github.com/ESIPFed/science-on-schema.org/blob/main/decisions/47-use-spdx-licenses.md>
- Would be helpful for there to be a list of licenses and auth requirements by all of the NSF-funded repositories - both for the repositories and for the NSF program officers to refer to.
- David -
 - New DMSP form is not really optimized for repositories, can we get feedback to NSF?
 - [NSF guidance](#) on new DMSP
 - Angela - also really hard for people to collaborate on it.
 - "Gold science standard rule" - be on lookout for downstream consequences
 - Suggestion that it might be informative to discuss open weight models for facilities
 - Chuck pointed out
<https://microsoft.ai/news/introducing-mai-thinking-1/>
- Chuck:
 - AI and Co-Data: The conversation covers CODATA's global policy role, Croissant, CDIF, DIDs, verifiable credentials, ODRL, provenance, deterministic question-answer pairs, knowledge graphs, drift, multilingual vocabularies, open compute, and enterprise AI governance. All of these are also represented in the QueryGraph.ai OSS AI community project.

 Chief Scientist: Slava Tikhonov, Head of AI at CODATA, Paris AI4LAM Youtube Talk

 AI4LAM Community Call - May 19, 2026
 - Semantic Croissant and Cross-Domain Integration in AI at NFDI4DataScience: <https://zenodo.org/records/17711820> and Building Data Infrastructure for Agentic AI <https://querygraph.ai/building-data-infrastructure-for-agentic-ai/>
 - Barrage of Nature papers: E. Aygün *et al.*, "An AI system to help scientists write expert-level empirical software," *Nature*, May 2026.
 - A. E. Ghareeb *et al.*, "A multi-agent system for automating scientific discovery," *Nature*, May 2026.
 - J. Gottweis *et al.*, "Accelerating scientific discovery with Co-Scientist," *Nature*, May 2026.

- L. Messeri and M. J. Crockett, "The uncritical adoption of AI in science is alarming — we urgently need guard rails," *Nature*, vol. 653, no. 8115, pp. 675–676, May 2026. <https://www.nature.com/articles/d41586-026-01557-x>
- "Why AI cannot do good science without humans," *Nature*, vol. 653, no. 8115, pp. 650–650, May 2026. <https://www.nature.com/articles/d41586-026-01551-3>
- Agentic Mainstreaming: <https://www.microsoft.com/en-us/microsoft-365/blog/2026/06/02/announcing-the-new-work-iq-apis/>
- Microsoft Build 2026: Building agentic apps with Microsoft Fabric and Microsoft Databases <https://azure.microsoft.com/en-us/blog/microsoft-build-2026-building-agentic-apps-with-microsoft-fabric-and-microsoft-databases/>

2026-05-07

Meeting Cancelled.

2026-04-02

Present: Matthew Barcys, Chuck, Werner, Don, Christina, Nicky

- Group Discussion:
 - Don: Kerchunk? <https://fsspec.github.io/kerchunk/> chunk aggregation tool? anyone heard of or using it?
 - [ESIP presentation](#)
- Chuck:
 - Shopify: <https://shopify.engineering/evolution-product-classification>
 - Moving from One-Shot KE pipeline to Agentic system. <https://youtu.be/bxToahwOVpY>
 - [Agentic Meshes:](#) <https://www.oreilly.com/library/view/agentic-mesh/9798341621633/>
 - Data Mesh: <https://www.oreilly.com/library/view/implementing-data-mesh/9781098156213/>
 - Knowledge Fabrics: <https://agenticmesh.substack.com/p/agentic-process-automation-cbd>

- <https://agenticmesh.substack.com/p/the-agentic-knowledge-fabric-an-approach>
- Observation, movement away from MCP to agentic cli
 - <https://github.com/googleworkspace/cli>

2026-03-05

- Presentation Werner Sun and David Butcher. [FOXDEN](#).
 - Tool to help with metadata collection. Command line interface and centralized server. Metadata is stored side-car, and no changes are needed to the existing data flows.
- Chuck's experiments with agents using metadata:
 - <https://github.com/LA3D/cogitarelink-fabric>

2026-02-05

- Presentation by Chris Erdmann RAID US Pilot ([Agenda](#), with more links) with team Christine Kirkpatrick, Julianna Christopher, Sheila Rabun.
- Slides:
 - https://docs.google.com/presentation/d/1vWuD-GEAtUq_FEsA0vhkODVk5AO-vHsL/edit?slide=id.g37bed3cb092_0_31#slide=id.g37bed3cb092_0_31
- [RAID schema](#)
- contact:
 - form:
 - https://docs.google.com/forms/d/e/1FAIpQLSfadmKYbj_OqfMMVAEUQIGa7GOi65Qej3cZX95rNfz1A_CxZw/viewform
 - email: projectpid@ucsd.edu

2026-01-01

- **No meeting due to Holiday.**