



March 4, 2026 | Chantilly, VA

From Repositories to Infrastructure: Making DOE's Scientific Knowledge AI-Ready

Presented By: Olga A. Kuchar, ORNL

DOE Data Curation Working Group

Metadata Sub-working Group



ORNL IS MANAGED BY UT-BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY



Authorship and Provenance

DOE Repository Metadata Profile (DRMP)

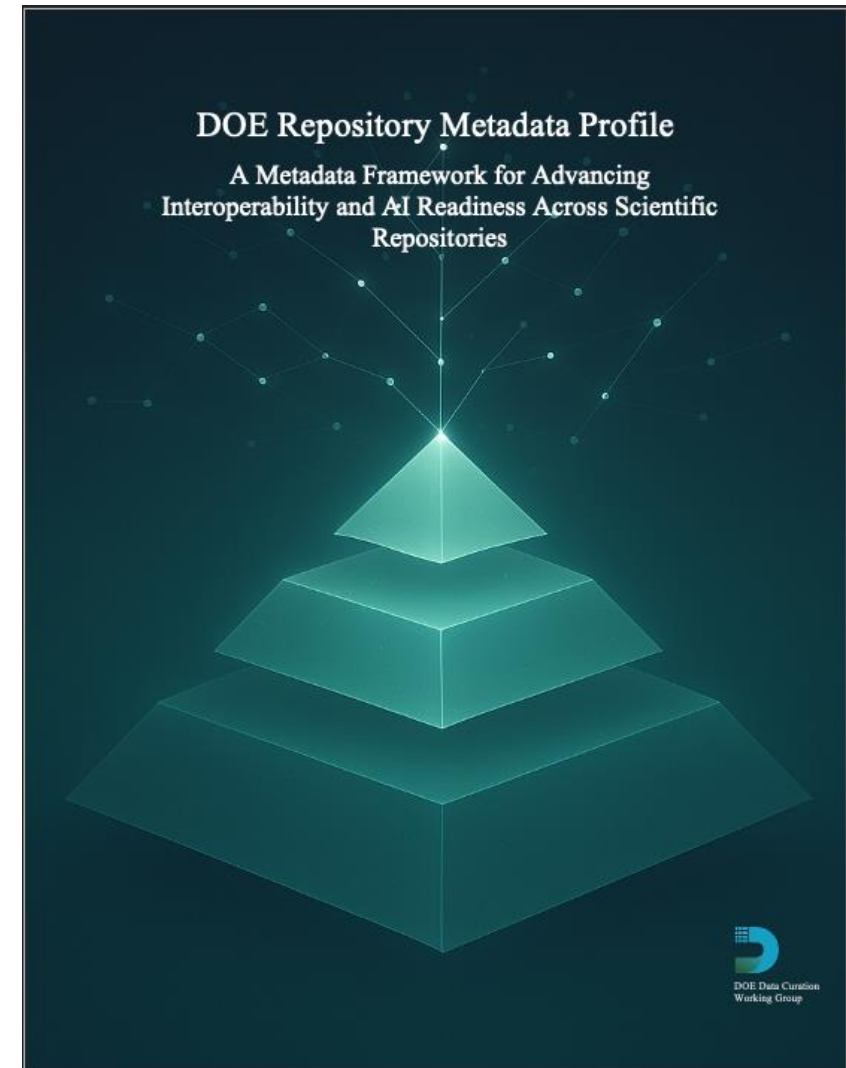
Developed by the DOE Data Curation Working Group
(Metadata Subgroup)

Primary Authors

- Olga Anna Kuchar (ORNL)
- Abigail Martin (NETL)
- Brian J. Cain (LANL)
- Kevin Kuhn (NETL)
- Lauren Spath Luhring (NLR)
- Juliane Schneider (PNNL)

Key Contributors

- Jonathan Huitt, Natalie Barnett, Sara Studwell (OSTI)
- Miriam Blake (PNNL)



*Reviewed by the broader
DOE Data Curation Working Group*

The Problem

We treat repositories as endpoints, not infrastructure.

Repositories are critical scientific infrastructure, but the metadata layer doesn't support the automation we now need.



Fragmented Metadata

Inconsistent fields across platforms



Policies Not Machine-Readable

Governance locked in narrative text



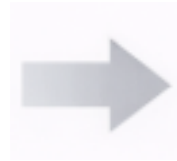
Automation Breaks at Scale

Workflows fail without structured rules

The Structural Gap

Repositories Today

- Described primarily through documentation
- Metadata implemented locally and inconsistently
- Policies expressed in human-readable, non-actionable text
- Integrations built as one-off solutions



Repositories As Infrastructure

- Machine-actionable repository metadata
- Shared governance and capability signals across the ecosystem
- Policies expressed in structured, enforceable form
- Harvestable, interoperable, automation-ready by design

The challenge is not intent – it's the underlying structure.

DRMP: Standards-Aligned Profile (Not a New Schema)

A harmonized profile mapping what repositories already expose, so machines can understand it.



Reuse, Not Reinvention

- Built on existing practices
- Profile, not a new schema



Harmonized Across Standards

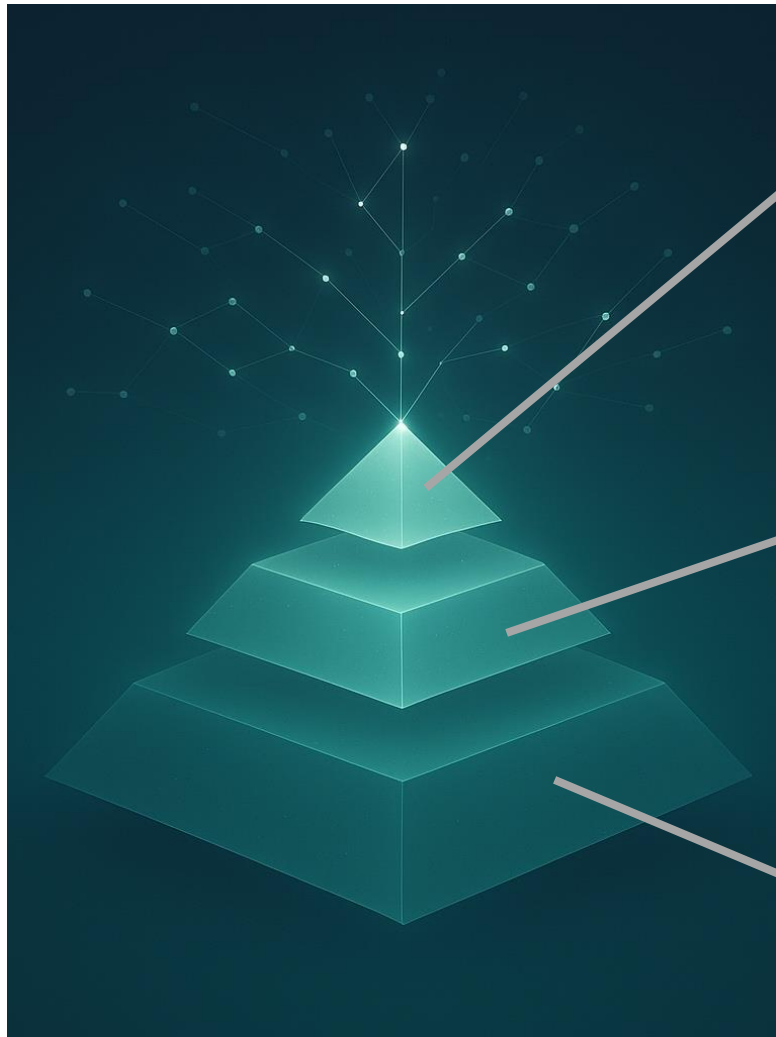
- RE3Data
- DCAT-US v3
- Schema.org
- Dublin Core
- DataCite 4.6
- PREMIS 3.0



Low-Burden Implementation

- Consistent, interoperable layer
- Governance and automation-ready

DRMP Conformance Model



Level 3 – AI-Ready

Semantic context · Provenance & integrity · Access models · AI-use policy

Level 2 – Interoperable

APIs · Metadata formats · License, retention & harvest policies · Capabilities

Level 1 – Minimum Viable Record

Identity · Operator · Contact · Repository type

What AI-Ready Means

Machines understand constraints, provenance, and meaning, not just data.



Explicit AI-use policy



Provenance and integrity signals



Semantic context for reasoning

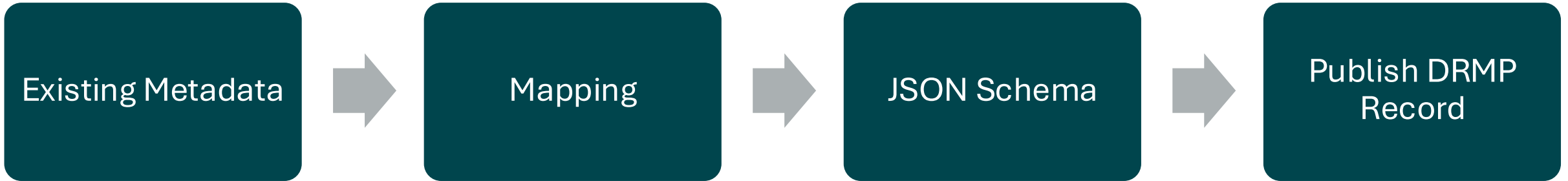


Machine-readable access constraints

AI-Ready \neq Open Season \rightarrow It Means Guardrails

Implementation Reality

Implementation is incremental and low-risk. Start with what you have.



- Minimal developer overhead
 - Most work is metadata mapping, not platform changes
- Start by validating your repository against Level 1 (Minimal Viable Record)
- API and policy patterns included
- Optional MCP templates
- Challenges: Aligning policy language with machine-readable formats may require governance input

Why This Matters to DOE Data Days (D3)

Curation and governance, operationalized for DOE's scientific infrastructure.

- Policies become executable → governance signals move from narrative to machine-readable
- Stewardship becomes visible → repository-level metadata makes curation transparent
- Interoperability becomes repeatable → reduces bespoke integrations across DOE repositories
- AI use becomes governable → enables policy-aware automation and safe AI adoption

DRMP is a foundation for FAIR, compliance, and sustainable automation, aligning with DOE's mission to advance data-driven science

What Is Next

FY26Q1: DRMP white paper released by the
DOE Data Curation Working Group

CY2026: Pilot repositories planned

Ongoing: Community feedback and extensions
welcome

Now: Voluntary, incremental adoption



Get involved: scan the QR code to
download the whitepaper and share
feedback!

Thank you! Q&A Session

What problem does DRMP solve?

- Makes repository governance, policy, and capability machine-readable
- Enables automation without losing human control
- Treats repositories as infrastructure, not just endpoints

How does this relate to automation / AI?

DRMP provides the repository-level context AI agents and workflow systems need to operate safely and correctly in federated environments.

What if I am not involved in the American Science Cloud?

You should still care because DRMP:

- Improves federated discovery
- Makes policies executable
- Reduces bespoke integrations
- Future-proofs repositories for any automation or AI use

DRMP is not an AI mandate – it is a governance and interoperability foundation.

Acknowledgements

Disclaimer. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

This report was prepared as an account of work sponsored by agencies of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.