

USDA EDMC

Metadata Management Plan

To be launched in April 2024

Due: June 14, 2024



PURPOSE OF THE PLAN



The vision for the **USDA Data Catalog** is to be a centralized location for anyone across USDA to broadly discover data sets that exist across the organization, and to be directed appropriately to other resources as needed for more in-depth details.



Data stewards must collect metadata about the data sets they govern and enter this information in the USDA Data Catalog. **The purpose of documenting metadata is to make our data more usable.** For example, someone searching for a data set can use metadata to understand the appropriate use of the data set in question.



Our objective in documenting ongoing and planned metadata management efforts through the Metadata Management Plan is to help Mission Areas (MAs), Departmental Administration and Staff Offices (DASO), and the Office of the Chief Information Officer (OCIO) chart their path to **better serve stakeholders that have use for USDA data, fulfill compliance with the Evidence Act, and maintain a Departmental catalog with our most valuable data sets.**

METADATA MANAGEMENT PLAN



MAs, DASO, OCIO submit Metadata Management Plans to EDMC towards the end of FY24 Q3 (June 14, 2024)

EDMC reviews Plans, identifies similarities (shared projects, investments, and risks) & shares findings with submitting groups

EDMC provides feedback on the Plans, discusses with each submitting group (ACDOs, ACIOs, DASO / OCIO representatives), then groups finalize their Plans

Update Plans annually



The Metadata Management Plan asks (1) what **metadata management efforts** have you invested in, (2) the maturity of your **compliance with federal guidance**, (3) the criteria used to identify your **high-priority data sets** and your 3-5 top priority data sets, (4) what **metadata governance processes** are being put in place, (5) and **key priorities** for both the **near and long term**.



The Metadata Management Plans can serve as **roadmaps to achieve integration with the USDA Data Catalog** and promote consistent information sharing across Mission Areas & DASO / OCIO. These plans will ultimately help Mission Areas & DASO / OCIO chart their path to better serving the customers who may have a need for agency data, **fulfilling compliance with the Evidence Act** and **maintaining catalogs of their valuable data sets**.

SCOPE OF THE METADATA MANAGEMENT PLAN



Mission Area & DASO / OCIO Metadata Management Plans will include:

- A **big-picture** overview encompassing **all metadata** to be documented, for **all different kinds of data sets, including:**
 - Inherent metadata (names of data tables and columns, data types such as string, integer, etc.)
 - Curated metadata (descriptions and identification of data roles)

Metadata Management Plans will ensure **compliance with the Evidence Act requirements** for Agencies to include in the comprehensive data inventory metadata on each data set of the agency:

- Structural metadata describing data structures (e.g., data format, syntax, and semantics)
- Descriptive metadata describing data contents (e.g., information security labels)

EDMC will use the Metadata Management Plans to gather approaches from Mission Areas & DASO / OCIO for how to collect/enter metadata for all datasets in the **USDA Data Catalog**, the authoritative source of publicly available USDA data and the centralized location for USDA data sets.

TECH OPTIONS FOR MAs: OPT-IN OR PLUG-IN



To encourage flexibility due to unique technological and business requirements across USDA, the following options are available when developing your Metadata Management Plan.

- Centrally-managed feeds for USDA-wide platforms – Any data fed into Cloudera and EDAPT is captured automatically as part of existing scans. This includes inherent and some curated metadata for content within EDAPT (e.g. data lake tables, Tableau Server workbooks) is already fed into the USDA Data Catalog. Planned future additions to be included are selected attributes from AgData Commons, GeoData (details TBD).
- Option 1: Opt-in – Inherent metadata from a list of Mission Area data sets are **scanned** by the Department into the USDA Data Catalog from Mission Area-owned or other data systems, AND **Mission Areas will input curated documentation** about those contents in the USDA Data Catalog.
- Option 2: Plug-in – A Mission Area uses another data catalog tool or method to manage their inherent and **curated** metadata, due to additional Mission Area or technical requirements. The Mission Area must extract content for integration with the USDA Data Catalog, in accordance with requirements.
- Option 3: Both – A Mission Area chooses to use a combination of a plug-in tool and opt-in scans to the USDA Data Catalog, based on the variety of content needing to be documented.



DEFINING METADATA MANAGEMENT EFFORTS



Below are general guidelines for what to include in your Roadmap. You may expand on this list to include other priorities, as needed.

Categories of metadata management efforts:

- The metadata management method being pursued at your organization (opt-in, plug-in, both, or TBD)
 - *If TBD, please explain why you have not yet chosen a metadata management method (lack of clear guidance, funding limitation) and explain how the Department can help*
- Maturity level with metadata requirements and standards (both in and outside of USDA)
- Identify the criteria that Mission Areas will use to identify their highest priority data sets to be documented (mission-critical data, or for which documentation is otherwise required)
- Provide your 3-5 highest priority data sets (ex. ULO data)
- The scope of data addressed by the plan, including the subject domain areas (financial, geospatial, or other categorization method) and existing systems hosting the data identified above
- Governance processes for ensuring the capture and timely updating of high-quality metadata
- Time-bound efforts related to metadata management (6 months to 1 year and 1-3 years)

Mission Area & DASO / OCIO

Template for Metadata Management Plans

FY24



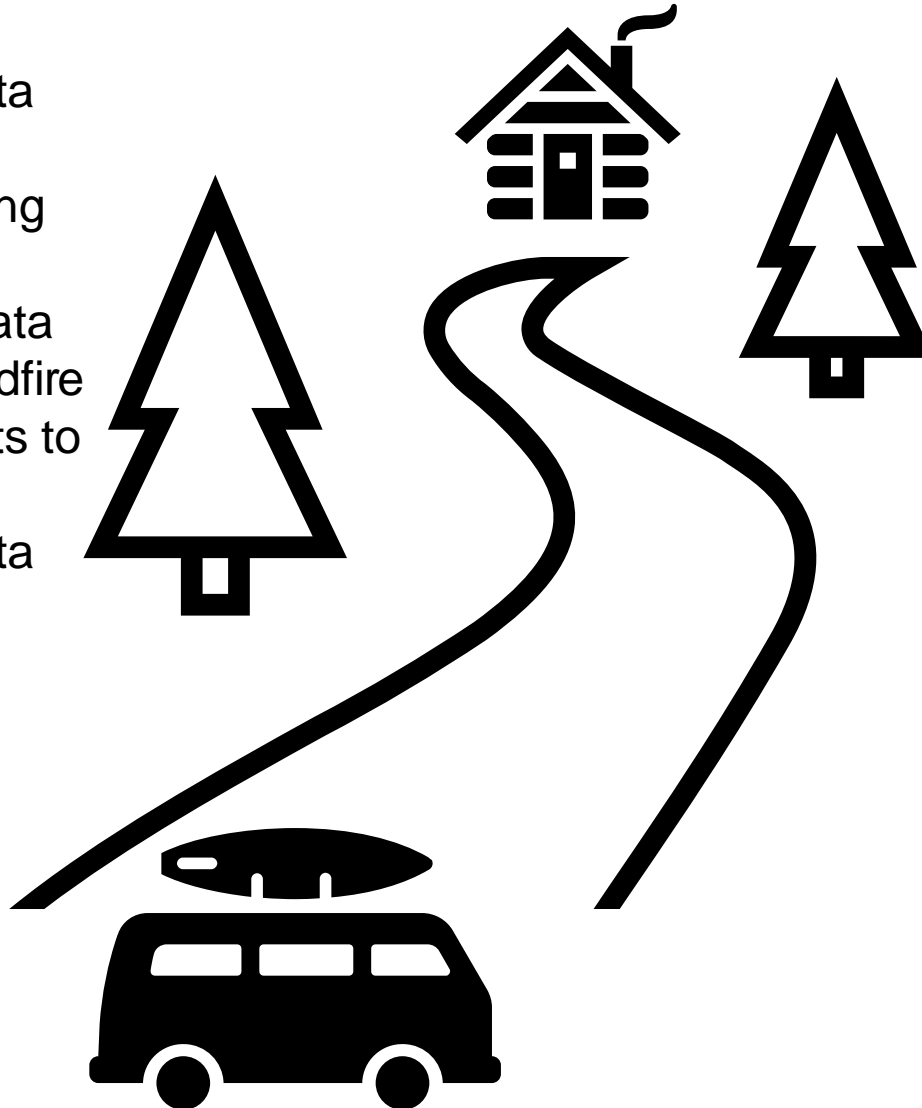
- The following slides outline the categories of information EDMC seeks to collect through the Metadata Management (*see slides 4-6 for details*). Please feel free to use the provided templates or provide your own. Modifications to this template can be requested by emailing Data Driven.
- Your organization's ACDO and/or ACIO may already have materials outlining your organization's active and planned metadata management efforts. If so, we encourage you to reuse those materials provided these materials answer the questions on each slide.
- Please submit this PPT to [this SP FOLDER](#) (MAs can upload, Chris/Fredy can review, MAs can look at each other's submissions) by June 14, 2024.

WHAT IS YOUR VISION FOR METADATA MANAGEMENT?



Near-term (next 6-12 months)

- Explore options for internal data catalog
- Explore options for public-facing data catalog
- Begin implementing the two data catalogs, including adding Wildfire Crisis Strategy-related datasets to them
- Simplify generation of metadata for Data.gov and GeoPlatform



Long-term (next 1-3 years)

- Develop additional service(s) to link FS data catalogs with USDA counterparts
- Expand internal and public-facing data collections in respective catalogs
- Develop dashboards showing success of public-facing data catalog
- Fully integrate existing FS data catalogs into enterprise infrastructure
- Leverage synergies across land management and scientific data catalogs
- Explore how an AI-enabled catalog can support a modern data architecture

WHAT IS YOUR METHOD FOR METADATA MANAGEMENT?



Opt-in, Plug-in, Ingest to centrally-managed feeds, Both, or TBD?

- In the Data and Analytics Modernization Plans, we asked:
 - Are you considering any pilots or proofs of concepts for other data catalog tools (e.g., pilots from NASS, REE, National Ag Library; Purview pilots)?

We are seeking confirmation that your organization's solution meets minimum metadata requirements (matching metadata elements to the corresponding required USDA Data Catalog elements) and that there is a **plan for meeting formatting and transfer requirements**.

- NRE plan is to primarily plug-in to the USDA catalog
- NRE Mission Area currently has two public-facing catalogs:
 - the FS Enterprise Data Warehouse catalog uses Content Standard for Digital Geospatial Metadata (CSDGM) as its metadata standard and plugs into the USDA structure for feeding Data.gov
 - the FS Research Data Archive catalog uses CSDGM-Biological Data Profile as its metadata standard and plugs into Ag Data Commons
 - Both catalogs are currently able to generate compliant JSON for Data.gov
- As noted on the Vision slide, NRE is working on pilots for NRE-wide data catalog tools

METADATA MANAGEMENT MATURITY



What is your metadata management maturity level in the context of the Evidence Act requirements summarized in the ACDO Letter of Delegation? If you are tracking any other requirements and standards for your metadata, please indicate below (ex: DR 3465, Geospatial Data Management).

The following requirements for metadata management included in the ACDO Letter of Delegation are drawn from the Evidence Act. Please summarize the activities at your Mission Area related to each requirement and indicate where you need support from the Department to increase the maturity of your metadata management.

- Ensure compliance with USDA Data Catalog requirements that all assets should be catalogued, especially priority datasets;
 - *Priority datasets are substantially cataloged now in the FS Enterprise Data Warehouse (estimated >99% completion); new priority datasets continue to be developed, and there is a defined and governed process for adding these datasets to the catalog*
 - *We project incorporating Mission Area data into the USDA Data Catalog in the 2nd half of FY25. We anticipate working with the Department in FY25 on the appropriate structure of web services to interlink the catalogs.*
- Govern administrative and programmatic data and manage data assets of the Mission Area (including the standardization of data format, sharing of data assets, coordinating with third parties that USDA shares data with, and the publication of data assets in accordance with applicable law);
 - *Consistent with the MAPLand Act, NRE is coordinating with Department of the Interior and the U.S. Army Corps of Engineers on standardizing data formats for various recreation data types*
 - *NRE will continue to operate its Enterprise Data Warehouse (EDW); the Content Governance Board for the NRE EDW was re-chartered in FY24 by the ACDO, ACIO, and NRE Geospatial Information Officer, and will continue to be the primary governance structure for NRE land management data assets*
 - *The governance structure for the Forest Service Research Data Archive creates data publications of a quality that classifies them as 'refereed' for Research Grade Evaluation Guide purposes; i.e., the quality classifier used for scientific journal articles*

METADATA MANAGEMENT MATURITY



What is your metadata management maturity level in the context of the Evidence Act requirements summarized in the ACDO Letter of Delegation? If you are tracking any other requirements and standards for your metadata, please indicate below (ex: DR 3465, Geospatial Data Management).

- Coordinate with Mission Area Associate Chief Information Officer (ACIO) to establish the necessary infrastructure for data sharing, leveraging enterprise solutions where appropriate
 - *NRE works closely with our ACIO regarding data sharing infrastructure. The FS Enterprise Data Warehouse infrastructure (internal and public-facing) is operated by NRE ACIO. The Forest Service Research Data Archive's public-facing servers are acquired from USDA DISC by NRE ACIO; the servers are managed jointly by Archive and ACIO staff.*
- Oversee and have decisional authority on Open Data release plans and processes: Do you currently oversee / make decisions on your Mission Area's Open Data release plans and processes? Have you worked with your Mission Area leadership to set priorities for Open Data? Any additional support or guidance required here?
 - *The ACDO is a co-sponsor of the Enterprise Data Warehouse's Content Governance Board (along with the Geospatial Information Officer and the ACIO); the board provides / coordinates data governance for the Enterprise Data Warehouse*
 - *The ACDO is a co-sponsor of the Mission Area's data catalog project (along with the FS Research and Development's Research Data Services program); this project is also referenced in the MA's Geospatial Data Act Action Plan approved in 2023*
 - *The ACDO must sign off on public-facing Tableau dashboards and is involved in priority-setting for new dashboards*
 - *ArcGIS Online (AGOL) "governance" falls outside of ACDO*
 - *Publication of scientific datasets falls outside of ACDO (this area has been governed by Research Data Services for >14 years; its repository conformed to FAIR standards before those were established)*

HOW DO YOU PRIORITIZE DATA SETS?



Prioritization

What is your criteria for determining what a high-priority data set is? *High-priority data sets should be documented in the USDA Data Catalog, it is helpful to understand prioritization criteria across the Department.*

- National Geospatial Data Assets are high-priority
- Scope – datasets that are national in scope are high-priority
- Importance – datasets relevant to national priorities are high-priority, even if the dataset is not national in scope
- Legislatively mandated – datasets that are legislatively mandated, such as datasets associated with Resources Planning Act (RPA) assessments, are high-priority

HIGH PRIORITY DATA SETS (3-5) AND SCOPE OF DATA



Which 3-5 data sets does your organization identify as high priority to be documented (mission-critical data, or for which documentation is otherwise required)? What is the scope of data addressed by your organization's plan?

| Data Set | Domain (i.e. geospatial, financial, etc...) | What is the name and type of the data system technology currently hosting this data (i.e. Redshift, Oracle, etc...)? | Is this data currently documented in the USDA Data Catalog? Yes/No | If not yet in USDA Data Catalog, timeline for documentation If already in catalog, planned actions | Blockers or Potential Support Requests from EDMC/Other MAs |
|--|---|--|---|---|---|
| FIA database (NGDA) | Tabular with some spatial information | SQLite database | No (it is in the NAL Geodata catalog and the Ag Data Commons catalog) | August 2025 | No known blockers beyond establishing the NRE data catalog and links to USDA Data Catalog |
| Monitoring Trends in Burn Severity (NGDA) | Geospatial | shapefile | No (it is documented in the NAL Geodata catalog) | August 2025 | See FIA database |
| National Forests proclaimed boundaries (NGDA) | Geospatial | Esri GDB | No (it is documented in the NAL Geodata catalog) | August 2025 | See FIA database |
| Resources Planning Act 2022 Assessment (multiple datasets) | Mix of tabular and geospatial | CSV, GDB | No (though they are documented in the USDA Ag Data Commons catalog) | August 2025 | See FIA database |

GOVERNANCE PROCESSES



What governance processes are in place to ensure the capture and timely updating of high-quality metadata?

| Governance Process | Timeline of Adoption | Current State | Future State | Blockers or Potential Support Requests from EDMC/Other MAs |
|---|----------------------------------|------------------------|---------------------------|--|
| Metadata for high-priority datasets are written by SMEs and reviewed by metadata specialists as part of the Enterprise Data Warehouse submission process. | Has been in place for >10 years. | Operating successfully | Continue existing process | No blockers |
| Metadata for formally published scientific datasets are written by the authoring science team with formal peer review by Forest Service Research Data Archive metadata specialists as part of the FS-RDA's submission and publishing process. | Has been in place for >10 years. | Operating successfully | Continue existing process | No blockers |
| Check with dashboard team for that piece! (Chris S) | | | | |
| | | | | |