Pacific Northwest
NATIONAL LABORATORY

PNNL-33396, Rev 0
DVZ-RPT-084, Rev 0

Requirements for Cataloging Hanford Geophysical Datasets

September 2022

Kenneth D Ham

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

PNNL-33396, Rev 0
DVZ-RPT-084, Rev 0

Requirements for Cataloging Hanford Geophysical Datasets

September 2022

Kenneth D Ham

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Summary

Environmental management activities at the Hanford Site produce extensive data about site conditions, contaminants, cleanup, and more. Managing and archiving that data requires a high degree of collaboration among site contractors and a high level of awareness by project managers and staff. Part of that effort is developing a Hanford Environmental Information and Data Index (HEIDI) to organize the data and maximize its value by making it findable and available for reuse. The objective is to catalog the disparate data sets collected to address the evolving needs of planning, executing, and documenting cleanup over several decades up to the present day, including links to active data sources when available. A properly implemented data catalog makes finding environmental datasets related to an area or theme a routine, reliable process, without requiring the searcher to have special knowledge that a data set exists and where it may be stored. In this project, a working group, including the U.S. Department of Energy, the Hanford Site contractors, and Pacific Northwest National Laboratory staff, identified needs and requirements for handling complex site data. Geophysical data was chosen as a test case because it can be large and complex and often involves multiple processing steps to extract the information incorporated into deliverables. The ability to document those steps was one of the requirements identified for the catalog. In addition to developing requirements, other activities included selecting a metadata schema and initial testing with the objective of determining whether the workflow and capabilities of selected data catalog software platforms were sufficient to implement and impose the identified requirements. This initial testing involved running the default catalog instance using the software platform of interest and altering the configuration to achieve each requirement, if possible. Where configuration alone was insufficient, the possibility of modifying the software by changing the code was examined, but not implemented. A follow-on task is planned to reprogram the code as necessary to implement requirements in a prototype catalog.

Acknowledgments

Funding for this work was provided by the U.S. Department of Energy Richland Operations Office under the Deep Vadose Zone – Applied Field Research Initiative. The Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DE-AC05-76RL01830.

Acronyms and Abbreviations

ATOM	an XML language used for web feeds
CPCCo	Central Plateau Cleanup Company
CSDGM	Content Standard for Digital Geospatial Metadata
DOE	U.S. Department of Energy
ECF	Environmental Calculation File
EM	Office of Environmental Management
ESRI	Environmental System Research Institute
FAIR	findable, accessible, interoperable, and reusable
FGDC	Federal Geographic Data Committee
FY	fiscal year
GeoRSS	a specification for encoding location as part of a web feed
GIS	geographic information system
GSSC	Geospatial Sciences Steering Committee
HEIDI	Hanford Environmental Information and Data Index
HLAN	Hanford Local Area Network (the enterprise information system)
HMIS	Hanford Mission Integration Solutions
IDMS	Integrated Document Management System
ISO	International Organization for Standardization
KML	a file format for displaying geographic data in an earth browser
LM	Office of Legacy Management
NAP	North American Profile of ISO 19115:2003
NNLEMS	Network of National Laboratories for Environmental Management and Stewardship
NQAP	Nuclear Quality Assurance Program
OGC	Open Geospatial Consortium (OGC) compliant CS-W 2.0.2 service
PNNL	Pacific Northwest National Laboratory
RBAC	role-based access control
REST API	an application programming interface (API) that conforms to the constraints of the representational state transfer or “REST” architectural style
SOAP	a messaging protocol specification for exchanging structured information in the implementation of web services

Contents

Figures

Tables

1.0 Introduction

Waste cleanup sites create, collect, and use extensive environmental information to characterize the site and guide cleanup decisions. Cataloging that information so that it can be discovered, retrieved, interpreted, and reused requires that each dataset be well documented (Wilkinson et al., 2016). Metadata describes dataset formats and content, the circumstances of data collection, procedures used to manipulate or model data, custodianship, data quality, preservation information, and details on access. By documenting and maintaining the proper context for a dataset, metadata forms a basis for cataloging data. Cataloging data supports data management and analysis in the near term but also supports long-term stewardship of cleanup sites. At the Hanford Site, a catalog of environmental datasets known as the Hanford Environmental Information and Data Index (HEIDI) is under development.

The growth in data quantity and complexity at the Hanford Site is accelerating. One example is the increasing use of geophysical methods to complement borehole data to improve site characterization and perform real-time monitoring of remedial activities. These complex datasets require thorough documentation of data sources and processing steps to capture the lineage of the data and demonstrate that quality objectives have been met. For those reasons, geophysical data provides a rigorous test case for developing requirements for cataloging data. Metadata standards such as the North American Profile of ISO 19115 (NAP-Metadata Working Group, 2007) include an extensive and flexible array of fields that support a wide variety of use cases. This investigation developed a set of requirements to assess whether that flexible structure is sufficient to support the conventions and restrictions that enable interoperability and integration of a data catalog into Hanford Enterprise Network (Hanford Local Area Network, HLAN) workflows.

The requirements for geophysical data archival and associated metadata described in this document were developed and specified through a Pacific Northwest National Laboratory (PNNL)-led a collaborative working group that included PNNL, The U.S. Department of Energy (DOE) Richland Operations Office, and Hanford Site contractors responsible for data archival: Central Plateau Cleanup Company (CPCCo) and Hanford Mission Integration Solutions (HMIS).

2.0 FY22 Activities

Activities in fiscal year (FY) 2022 built on the findings of FY21 activities, which identified ArcGIS Enterprise Sites as a software platform capable of implementing key catalog requirements for Hanford environmental information and data and with the advantage of currently being available on the HLAN (Ham and Crockett, 2021). The ArcGIS suite of tools includes several applications that manage the creation, import, and export of metadata in standard formats that are being evaluated and tested in FY22. The activities in FY22 included developing a list of Hanford-Site-specific requirements, selecting a metadata standard, and evaluating how those requirements can be implemented in the schema (the structure of how information is represented and stored) of the selected metadata standard and cataloging software. These activities consider how to catalog geophysical data as a test case with considerable complexity because successful approaches will accommodate most, if not all, types of data.

2.1 Defining a Path to Environmental Data Catalog Implementation

Building on the work performed in FY21 that focused on identifying the full range of geophysical data sets and associated metadata standards, requirements were established for specific Hanford systems in collaboration with DOE and the site contractors. The following subsections describe the activities executed to achieve this goal.

2.1.1 Metadata standard identification

Although there is no standard approach for documenting geophysical data, federal agencies are mandated by Executive Order 12906 to use metadata standards endorsed by the Federal Geographic Data Committee (FGDC), including:

- The Content Standard for Digital Geospatial Metadata (CSDGM) or its extensions for biological data (Biological Data Profile) and shoreline data, or
- The International Organization for Standardization (ISO) series of standards (19115, 19115-2, 19139, etc.).

The U.S. is moving to the use of ISO standards for metadata, so those were seen as preferred over a CSDGM-based standard if they provided the necessary functionality. ISO metadata standards have a flexible hierarchical structure that supports many types of data and approaches to cataloging that data. The downside of flexibility is that it can allow ambiguity in how some information is entered into the metadata hierarchy, which may lead to inconsistency. Metadata standards produce records that are meant to be shared widely and easily, but there are additional considerations and different priorities that arise in the management of Hanford Site environmental data. In addition to considering whether a metadata standard can fulfill its role in cataloging the data, it is also necessary to consider whether it can support the implementation and interoperability of that catalog on the HLAN. This document is intended to deliver a list of requirements for implementing a data catalog that fits Hanford Site needs and can integrate with the existing or planned HLAN structure. This work examines how those requirements can be met using existing FGDC-endorsed metadata standards and data management tools that can be implemented on the HLAN.

The North American Profile (NAP) of ISO 19115:2003 is FGDC-endorsed and is both an American National Standard and a Canadian National Standard (NAP-Metadata Working Group, 2007). The NAP has been implemented by various organizations and is supported by many geospatial software tools. This endorsement and industry support make the NAP an attractive option. For application to Hanford Site environmental data, it can be advantageous that ISO metadata standards include extensive and extensible

data fields with a great deal of flexibility in how they are used. Within that flexible structure, it should be possible to impose additional structure by requiring entries to conform to additional requirements. The resulting entries would remain compliant with the NAP schema while maintaining the consistency needed for interoperability across multiple data systems. As requirements were developed in the following activity, the ability to implement them in the NAP was considered. Where possible, those capabilities were tested in the final activity in this task by implementing them in the test suites of software. Implementation was tested first in ArcGIS Hub, the non-enterprise equivalent of ArcGIS Enterprise Sites, which had been identified as a promising software platform for developing the HEIDI capability (Ham and Crockett, 2021)

2.1.2 Requirement identification

Data is documented through metadata, and Hanford Site procedures result in a rich set of information detailing how and why environmental data was collected as well as how it was processed or analyzed and indicators of its quality. That information is the mainstay of any effort to catalog the data and deliver on the FAIR principles (Wilkinson et al., 2016). Data fulfills the FAIR principles when it is findable, accessible, interoperable, and reusable – all characteristics that facilitate the type of collaborative work required at environmental cleanup sites. A data catalog must capture and organize this metadata and provide methods for finding and accessing entries. Making that catalog interoperable with other systems on the HLAN will require a structured and secure approach. The additional structure would support and enforce the integration of Hanford-specific locations, linkages to supporting documents, and defined metrics for things such as quality. In addition to the consistency of data in the catalog entries, additional information about controls on data access must be incorporated.

Complex data from geophysics sampling was selected as a test case, as it may require documenting additional characteristics that influence how it might be used or how its quality can be interpreted. This task engaged the Hanford environmental data community to compile requirements that support objectives for data documentation, management, and interoperability. The following requirements are not intended to encompass the bulk of metadata considerations for Hanford environmental data, but instead represent those areas where the metadata standard may need to be thoughtfully configured and populated, or supplemented to fit the data types and data systems involved in the management of Hanford environmental data.

2.1.2.1 Requirement implementation

The requirements are developed to serve an information need, but the implementation of those requirements is important if the functionality is to become available to users. The four areas where implementation of a requirement is possible are:

- Metadata schema: Defines the structure of catalog entries and what is available for searching and organizing the catalog.
- Metadata editor: Controls how catalog entries are created, requiring certain fields, limiting field contents, etc.
- Data Catalog Interface: Defines how users interact with the catalog, how searches are built, and what fields are available to be searched.
- HLAN: Controls user authorization and roles, and facilitates interaction among data systems.

The following sections on requirements identify what must be implemented in these areas to achieve the desired results.

2.1.2.2 Theming

Once data is cataloged, there is a need to search the catalog to find data of interest. Keywords are associated with catalog entries (the metadata) to quickly narrow searches. Themes are one type of keyword implemented in metadata schemas derived from ISO 19115. The standard theme categories overlap with some Hanford environmental data themes, but additional granularity is needed to highlight contaminants and cleanup in searches. Figure 2.1 illustrates how Hanford environmental data themes provide a more granular linkage to the objectives that have driven the data to be collected. This Hanford-specific list of themes differs from standard keyword lists (https://www.fgdc.gov/nap/metadata/register/codelists.html), but NAP supports custom sets of keywords that would allow a Hanford Environmental Data Theme list to be used for organizing and finding datasets.

ISO Topic Keywords	Hanford Environmental Data Theme
Farming	Atmospheric
Biota	Biota
Boundaries	External Radiation
Climatology	Groundwater
Meteorology/ atmosphere	Miscellaneous Material
Economy	Pore Water
Elevation	Sediment
Environment	Soil Gas
Geoscientific information	Soil
Health	Surface Water
Imagery/ basemap/ earth cover	Waste Solid
Intelligence	Waste Water
Military	
Inland waters	
Location	
Oceans	
Planning/ cadaster	
Society	
Structure	
Transportation	
Utilities/ communication	

Figure 2.1. Hanford Environmental Data Themes Contrasted with ISO 19115 Topic Keywords

To incorporate keywords consistent with Hanford environmental data themes into catalog entries, two capabilities are required. First, a standardized list of those terms must be created and agreed to for site-wide use and, if broader availability is desired, made available in a public registry. Second, the metadata editor must restrict entries for theme keywords to those terms found in the Hanford-specific standardized list. Table 2.1 breaks down those requirements by the tool where they must be implemented.

Table 2.1. Hanford-Specific Theming Implementation

Implementation Area	Required Functionality
Metadata Schema	Support custom lists for keywords
Metadata Editor	Enforce custom lists for keywords
Data Catalog	Present Hanford-specific keywords as choices in search tools
HLAN	Maintain a site-wide list of theme keywords for use in all HLAN environmental data information systems

2.1.2.3 Limitations on Access

Datasets containing information that is sensitive or not yet approved for sharing require limitations on access. Other Hanford environmental data systems implement role-based access control (RBAC). Using RBAC, limitations can be associated with a user through their enterprise login, which will be assigned to a role that determines what permissions are granted. In practice, this can take two forms: (1) blocking access to the catalog entry describing the resource (the dataset, graphic, or whatever item the catalog entry represents) or (2) blocking access to the resource, while allowing the catalog entry to be viewed. Blocking access to the catalog entry hides the fact that the data exists from those without access permission while blocking access to only the resource allows the user to discover the information, though they must request access to view or download it. Table 2.2 breaks down the capabilities needed to restrict catalog entry access and Table 2.3 covers restricting access to the resource itself.

Table 2.2. Catalog Entry Access Restriction Implementation

Implementation Area	Required Functionality
Metadata Schema	Support differentiating classes of records by access limitation
Metadata Editor	Require access limitation entry
Data Catalog	Enforce limitation on access to catalog entry based on HLAN identity or role (RBAC)
HLAN	Maintain lists of identities with similar roles and access

Table 2.3. Resource Access Restriction Implementation

Implementation Area	Required Functionality
Metadata Schema	Support identifying access limitations for the resource
Metadata Editor	Support or require access limitation entry
Data Catalog	Enforce limitation on access to a resource based on identity or role (RBAC)
HLAN	Maintain lists of identities with similar roles and access. File systems may also be involved in limiting access to resources.

2.1.2.4 Limitations on use

Once access to a dataset or other resource is granted, it is useful to specify any limitations on its use. It is also helpful to specify when no restrictions are imposed. Many limitations on use are formalized through a variety of open data licenses or standard specifications. Examples of some of these licenses can be found at https://docs.data.world/en/99118-common-licenses-in-order-of-most-open-to-most-restrictive-.html, with some discussion of what is restricted by each license type. For example, one type of license

might restrict commercial use, while another might prevent derivative works from being shared due to the need to maintain proper source documentation. These licenses can be specified, but there is little ability to enforce them in the catalog software. Compliance relies on users to follow accepted norms about licensing, attribution, sharing, etc. Table 2.4 details how use limitations can be documented and shared with users.

Table 2.4. Use Limitation Implementation

Implementation Area	Required Functionality
Metadata Schema	Support specifying constraints on use
Metadata Editor	Support or require use limitation entry
Data Catalog	Prominently display constraints on use
HLAN	Enforcement not possible. Once access is allowed (see above), limiting how data is used elsewhere involves voluntary compliance with terms of use.

2.1.2.5 Link to deliverable

When Hanford Site deliverables rely on environmental data to support their findings, it is necessary to document that link. The source data is important for documenting and justifying the need for environmental cleanup activities as well as for documenting progress. Metadata schemas typically have well-established methods for identifying sources used to generate the resource. In contrast to that more common approach, linking to works derived from a resource creates an additional burden of awareness and response because the links in the metadata entry must be updated each time a derivative work arises. Further development and testing will be needed to identify what approach to linking will be required.

Hanford deliverables are archived in the Integrated Document Management System (IDMS), rather than in the environmental data catalog. A way to cross reference deliverables and data sources is needed. There are two approaches to this that appear viable: (1) create aggregate entries in the catalog that cross reference the deliverable and the datasets or data sources used, or (2) create catalog entries representing the deliverables themselves, with the resource being their official archive location with data sources represented in the lineage fields of the metadata. The second approach is simpler, but the first may provide additional flexibility and provide a way to represent Environmental Calculation Files (ECF; a formal approach to documenting the preparation, processing, and analysis of data for use in Hanford Site deliverables) as aggregates. Table 2.5 details the two alternative approaches for linking datasets to a document or deliverable.

Table 2.5. Link to Deliverable Implementation

Implementation Area	Required Functionality	Alternate Functionality
Metadata Schema	Support aggregate datasets (possibly including ECF)	Support a deliverable as a resource
Metadata Editor	Link aggregate datasets to a deliverable	Use lineage fields to identify data sources
Data Catalog	Make aggregate datasets findable based on a deliverable	Handle deliverables as a resource in a catalog entry
HLAN	Enable linking to deliverables in IDMS	Enable linking to deliverables in IDMS

2.1.2.6 Place-naming consistent with other Hanford systems

The number of areas and activities at the Hanford Site require a careful, systematic approach to describe a place unambiguously. Place names may be intertwined with the context of what created that place or the current or past activities at that place. An area might, therefore, be referred to by different names, numbers, or codes to represent different activities or eras. The challenge of naming places on large and active cleanup sites requires coordination of contractors, especially those leading the management of infrastructure and spatial information about the site. The site must adopt a standard for place naming that adequately limits ambiguity without introducing unnecessary or burdensome complexity. Such a standard would limit the possibility of places being associated with the wrong name, which would strengthen the ability of catalog searches to narrow results to the desired locations.

The location of a dataset is identified in two primary ways in metadata: place keywords or geospatial extents. Geospatial extents often specify geolocation information, but it is still good practice to be consistent in naming or describing those extents to address any contextual meaning of the location. Place keywords must stand alone, and so must encapsulate any such context and be differentiated from any places with similar names, functions, or context. Given that a place may serve different roles across activities, there is likely a need to assign more than one type of place name to a resource. To avoid incorrect or ambiguous place names, it will be necessary to incorporate a site-specific list of place names that can act as place keywords. Table 2.6 details how lists of Hanford-specific place keywords can be implemented.

Table 2.6. Consistent Place-Naming Implementation

Implementation Area	Required Functionality
Metadata Schema	Support custom lists for place keywords, including multiple types of place keywords to reflect differing “uses” of a site
Metadata Editor	Enforce custom lists for place keywords
Data Catalog	Make custom place keywords available for searching the catalog
HLAN	Consistent use of place nomenclature across environmental data information systems to allow the formation of custom lists for place keywords

The discussion above demonstrates how requirements are implemented in the metadata schema, the metadata editor, the data catalog software, or the HLAN enterprise. The following sections describe testing of the data catalog software, which in some cases involved other components of the system, including the harvester, the web interface, and the metadata editor.

2.1.3 Geophysical data demonstration and testing

Previous efforts identified ArcGIS Enterprise Sites as a promising software platform for implementing a data catalog for the Hanford Site (Ham and Crockett, 2021). As specific requirements were developed (see Section 2.1), it was possible to test whether and how the requirement could be met using this platform.

2.1.3.1 Testing: ArcGIS Enterprise Sites

ArcGIS Enterprise Sites is not used at PNNL, but capabilities were tested in the non-enterprise equivalent, ArcGIS Hub, which is available at PNNL. Using a non-enterprise version had minimal impact on testing, except that roles used to limit access could not be integrated into the user identity and access

controls on the enterprise network domain. Cataloging datasets and managing their metadata through ArcGIS Hub was straightforward and not difficult to accomplish. Some Hanford requirements, however, proved difficult to implement. The need to search based on Hanford-specific keywords was not easily accommodated. Searches for datasets are based on a limited subset of the keywords available within the metadata, and that does not appear to be configurable to achieve what is needed. This and other limitations encountered during testing of ArcGIS Hub led to the exploration of alternatives.

Tools within the Environmental System Research Institute (ESRI) community of practice were prioritized to maximize interoperability with the ESRI systems implemented on HLAN. The approach of abstracting metadata into a more universal format, allowing easy conversion to the desired format, still held appeal for Hanford applications. Fortunately, ESRI provides more than one tool for cataloging data and managing metadata. Their Geoportal Server (https://www.esri.com/en-us/arcgis/products/geoportal-server) is an open-source tool that appears to provide both broad functionality and the potential for customization. Testing efforts were redirected from ArcGIS Enterprise Sites to the Geoportal Server.

2.1.3.2 Testing: Geoportal Server

This free, open-source software tool is built and maintained by ESRI, with full source code provided for those who wish to modify it to their own requirements. The software is provided as a GitHub repository (catalog: https://github.com/Esri/geoportal-server-catalog and harvester: https://github.com/Esri/geoportal-server-harvester) under the Apache 2.0 license (https://raw.githubusercontent.com/Esri/geoportal-server-catalog/master/LICENSE.txt). In addition to those two major components, it is worth mentioning that the tool includes a metadata editor that would play an important role in achieving the requirements of a Hanford environmental data catalog.

Catalog

The catalog module inventories the metadata of geospatial resources. Most of the interaction with metadata and resources occurs within this module. Browsing, searching, viewing, mapping, and loading new resources are some of the ways to interact with the catalog. The open-source nature of this tool allows, with some effort, the possibility to reconfigure or reprogram those functions to meet the requirements for Hanford environmental data management. Many interactions will require no reprogramming to achieve. The module serves as a geoportal catalog service, which is an Open Geospatial Consortium (OGC) compliant CS-W 2.0.2 service. Other tools can access data through various interfaces, including REST, GeoRSS, KML, SOAP, and ATOM, providing a variety of options for integrating the catalog capabilities into HLAN environmental data workflows. The catalog includes a web application for interacting with the catalog, but it is also an option to build a custom web interface that complies with HLAN authentication requirements, as one example.

An important feature of the catalog is the metadata editor. This editor allows the creation and editing of metadata in many formats, including FGDC and FGDC-endorsed ISO 19000-series metadata standards. These standards can be supplemented with custom XML formats, which may be necessary to achieve the requirements set out above.

Configurable search is a feature that allows the user to narrow a selection based on a variety of metadata fields. This capability helps bring Hanford-specific themes or keywords to bear on searches for content.

Harvester

The harvester module delivers a powerful, flexible tool for incorporating entries from existing data catalogs that comply with any of a variety of standards into the local catalog. This capability includes an extensible API that supports the creation of custom protocol adaptors, filters, processors, analyzers, and other elements that allow the workflow to be configured to meet requirements. It is possible to run the harvester as a web application or as a stand-alone program from the command line. The benefit of this flexibility is that it should be possible to configure the harvester to convert incoming metadata records or flag them as needing entries to comply with Hanford requirements.

2.1.3.3 Testing against requirements

Initial testing of the ESRI Geoportal Server was completed using a local installation. As a web application, integration with other HLAN data systems should be straightforward. Catalog entries were loaded to allow requirements to be evaluated. Table 2.7 through Table 2.10 indicate what requirements have been evaluated within the Geoportal Server and the pass/fail outcomes of those tests. Additional testing will be completed in the next FY, including an evaluation of metadata editor capabilities that require modifications to program code and those requirements with a greater interaction with HLAN. Greater interaction with HLAN requires additional information about HLAN vocabularies and includes an evaluation of the consistency with HLAN naming conventions.

Table 2.7. Geoportal Server Testing Results: Hanford-Specific Theming Implementation

Implementation Area	Required Functionality	Pass/Fail
Metadata Schema	Support custom lists for keywords	Pass
Metadata Editor	Enforce custom lists for keywords	Deferred to next FY
Data Catalog	Make custom keywords accessible in search criteria	Pass
HLAN	Consistent use of theming nomenclature across environmental data information systems	Deferred to next FY

Table 2.8. Geoportal Server Testing Results: Catalog Entry Access Restriction Implementation

Implementation Area	Required Functionality	Pass/Fail
Metadata Schema	Support differentiating classes of records by access limitation	Pass
Metadata Editor	Require access limitation entry	Deferred to next FY
Data Catalog	Enforce limitation on access to catalog entry based on HLAN identity or role	Roles implemented, access restriction testing still underway
HLAN	Maintain lists of identities with similar roles and access	Pass

Table 2.9. Geoportal Server Testing Results: Resource Access Restriction Implementation

Implementation Area	Required Functionality	Pass/Fail
Metadata Schema	Support identifying access limitations for the resource	Pass
Metadata editor	Support or require access limitation entry	Support for access limitation entry confirmed. Testing whether entry can be required is deferred to next FY.
Data Catalog	Enforce limitation on access to resource based on identity or role	Roles implemented, access restriction testing still underway
HLAN	Maintain lists of identities with similar roles and access. File systems may also be involved in limiting access to resources.	Pass

Table 2.10. Geoportal Server Testing Results: Use Limitation Implementation

Implementation Area	Required Functionality	Pass/Fail
Metadata Schema	Support specifying constraints on use	Pass
Metadata editor	Support or require use limitation entry	Support for access limitation entry confirmed. Testing whether entry can be required is deferred to next FY.
Data Catalog	Prominently display constraints on use	Pass
HLAN	None	N/A

Testing of ArcGIS Enterprise Sites revealed limitations in configuration options that prevent it from providing the required functionality to catalog Hanford environmental data. Testing has not revealed any major impediments to implementing such a catalog using Geoportal Server. Testing involving software modifications and HLAN interactions will continue in the next FY, as described in Section 3.0, *Future Work*.

2.2 Technical Support to DOE and Site Contractor

The working group for this project includes DOE, PNNL, CPCCo, and HMIS. Discussions range beyond project objectives to include data use, access, workflows, security, and more. These discussions help place the objectives in context while facilitating interactions among parties with interest in Hanford environmental data.

In addition, the task lead is participating in two groups that share lessons on data management for Office of Environmental Management (EM) and Office of Legacy Management (LM) activities at DOE sites. Learnings from the project inform the needs of other facilities and vice versa. Those groups are:

- Network of National Laboratories for Environmental Management and Stewardship (NNLEMS)
  - Establishes a partnership of Federally Funded Research and Development Centers to advance the effectiveness of the scientific and technical expertise in the DOE national laboratory complex (national labs) toward meeting the objectives of both EM's legacy nuclear waste cleanup mission and LM's long-term surveillance and maintenance mission. The NNLEMS functions as an interface or focal point for information exchange, consensus building, and coordination to facilitate EM's and LM's ability to access and leverage the capabilities of the national labs.

- DOE Geospatial Sciences Steering Committee (GSSC)
  - Functions in an advisory role to the DOE national labs, major facilities, and headquarters and field office elements to actively promote effective use of GIS science and technology in the DOE complex. The principal goals of the GSSC are to foster technical excellence and communication, to identify and advocate best business practices, to provide sound recommendations on policy and standards, and to promote coordination between DOE and other federal and state agencies in areas of mutual interest.

2.3 Summary of Activities Completed

Table 2.11. FY22 Activity Status

Activity	Status
Develop requirements for managing and integrating a catalog of environmental data resources into HLAN workflows	Complete
Evaluate ArcGIS Enterprise Sites	Complete (failed/rejected)
New activity: ESRI Geoportal Server evaluation	Initial phase completed. Additional phases planned for next FY.

3.0 Future Work

FY23 work will build on the foundational work in FY21 and FY22 on identifying data management needs, promising software tools, and requirements for integration with other data systems. That foundation makes it possible to prototype a data catalog to test the essential functions and confirm that the requirements are sufficient to meet Hanford needs. This work will develop a prototype Hanford environmental data catalog based on the ESRI Geoportal Server software in a PNNL network setting. By configuring and programming both the metadata editor component as well as the catalog component, it will be possible to test requirements in a functional prototype environmental data catalog website. Software design specifications will be developed to capture the lessons learned during prototyping and facilitate the path to HEIDI implementation. Once implemented, HEIDI will improve awareness of and access to existing environmental data that may be brought to bear on new decisions. Access to this data will improve confidence in decisions without incurring the additional costs or time required to reproduce data that might otherwise be overlooked.

4.0 Quality Assurance

This work was performed in accordance with the PNNL Nuclear Quality Assurance Program (NQAP). The NQAP complies with the DOE Order 414.1D, *Quality Assurance*. The NQAP uses NQA-1-2012, *Quality Assurance Requirements for Nuclear Facility Application,* as its consensus standard and NQA-1-2012 Subpart 4.2.1 as the basis for its graded approach to quality.

Any data presented in this document are preliminary, for information only, and subject to revision. The information associated with this report should not be used as design input or operating parameters without additional qualification.

5.0 References

DOE Order 414.1D, "Quality Assurance." Washington, D.C.: U.S. Department of Energy.

Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure." *Federal Register* 59(71):17671-17674, April 13, 1994.

Ham K.D., and D.L. Crockett. 2021. Evaluation of Data Catalog Software for Hanford Site Environmental Datasets. PNNL-31960. Richland, WA: Pacific Northwest National Laboratory.

NAP-Metadata Working Group. 2007. North American Profile of ISO19115:2003 - Geographic information – Metadata (NAP – Metadata, version 1.1).

NQA-1-2012, "Quality Assurance Requirements for Nuclear Facility Application." New York, NY: American Society of Mechanical Engineers.

Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, et al. 2016. "The Fair Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1). https://doi.org/10.1038/sdata.2016.18.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov