



Minnesota Geospatial Advisory Council - Archiving Pilot Workgroup Final Report

Introduction

This report covers the exploratory work carried out by the Minnesota Geospatial Advisory Council's Archiving Pilot Workgroup, active from June 2021 to January 2022. This workgroup was formed to evaluate and test a range of potential archive technologies, create a proof of concept with sample sets of data, and continue to perform community outreach in order to build on the work of the [Archiving Implementation Workgroup and the Archiving Workgroup](#).

Summary

The Archiving Pilot Workgroup was organized into two subgroups: one to focus on performing outreach and education and another to pilot software applications. The Outreach Subgroup gathered user stories and created educational materials. The Pilot Subgroup tested four software applications that could serve as discovery platforms for items in a geospatial archive (CKAN, DSpace, GeoBlacklight, and Preservica). While this report includes many details about each platform's functionality, workflows, and usability, it does not go as far as recommending one software application over another. All of the tools that were tested would be suitable for a geospatial archive. Therefore, the ultimate choice will come down to which entity hosts the archive, available expertise, and how much staff time can be dedicated to it. We hope that the information provided in this report helps to advance the conversation about how and where an archive can be hosted.

Next Steps

Although this pilot focused on testing data pulled from the Minnesota Geospatial Commons ("Commons"), we are aware that aerial imagery is one of the highest priority types of data requested for the archive. As a next step, we propose to convene another short-term workgroup to pilot the archiving of imagery. This would be helpful to determine any major differences in archiving imagery data, such as space considerations due to the size of the files. The next workgroup would also continue to compile stories and use cases for the archive.

Workgroup Members

Outreach Subgroup

- Andra Mathews - Minnesota Department of Transportation
- Ryan Mattke - University of Minnesota Libraries (Subgroup Lead)
- Nancy Rader - MNIT / MnGeo

Pilot Subgroup

- Sarah Barsness - Minnesota State Archives
- Melinda Kernik - University of Minnesota Libraries
- Carol Kussmann - University of Minnesota Libraries
- Karen Majewicz - University of Minnesota Libraries (Subgroup Lead)
- Zeb Thomas - MNIT / Minnesota Department of Natural Resources

Resource People (consulted as needed)

- Jon Hoekenga - Met Council
- Randy Knippel - Dakota County
- Brent Lund - MNIT / MnGeo
- Jesse Reinhardt - Hennepin County
- Cory Richter - City of Blaine

Table of Contents

Introduction	1
Summary	1
Next Steps	1
Workgroup Members	2
Outreach and Education	4
Software Application Pilot	4
Repository Functions	5
Content Organization	6
Discovery Metadata	6
Overview of software applications	6
1. CKAN	7
2. DSpace	7
3. GeoBlacklight	7
4. Preservica	8
Combining Applications	9
Appendices: Technical Details	10
Appendix 1. CKAN Application Details	10
Appendix 2. DSpace Application Details	14
Appendix 3. GeoBlacklight Application Details	21
Appendix 4. Preservica Application Details	30
Appendix 5. Metadata Crosswalk	38

Outreach and Education

The goal of the Outreach Subgroup was to develop appropriate communications in order to continue to build support in stakeholder communities, including MnGeo, government agencies at all levels, academic researchers, U-Spatial affiliates, students/teachers/historians through the Minnesota Historical Society Library, all users of historical geospatial data, non-profit organizations, private sector, and tribal nations. To that end, between June and September 2021, we contacted select respondents to the Archiving Historical Geospatial Data survey from 2020 and invited them to elaborate on how they use historical geospatial data in their work. We used this material to create a [web page highlighting these stories](#). As of December 2021, there were three stories on the page, with more in development. The work of the subgroup will continue informally as more stories are added.

This work was also shared with the GIS community by presenting the material in various ways at the Minnesota GIS/LIS Conference (October 2021):

- Lightning Talk: [Archiving Geospatial Data in Minnesota: Recommendations and Future Directions](#)
- Poster: [The Future of Historical Data in Minnesota](#)
- Panel Discussion: [What is geospatial data archiving and why is it important for Minnesota? \(additional Wisconsin slides\)](#)

Software Application Pilot

The Pilot Subgroup tested the workflows proposed in the [Archiving Implementation Workgroup Report](#). We evaluated four different software applications (in alphabetical order): CKAN, DSpace, GeoBlacklight, and Preservica. During our evaluation, we assessed the software for which **repository functions** it can perform and how it handles **content organization**. We also constructed a preliminary **metadata crosswalk**. Finally, we built or repurposed multiple **pilot installations** of the applications and populated them with datasets pulled from the Minnesota Geospatial Commons (“Commons”).

All of the tools we tested would be suitable for the archive. The final choice will depend on which entity hosts the archive, available expertise, and how much staff time can be dedicated to it.

Repository Functions

An archive that supports both public access to spatial data resources and long-term preservation requires an infrastructure with three related but distinct functions: discovery, access, and preservation.

- **Discovery** refers to the way users find resources and know what exists. For geospatial data, this most commonly takes the form of a front-end search portal that users browse and search.¹
- **Access** describes the infrastructure that provides active file storage and delivers copies of data and metadata to users. This function might be integrated into a repository application or set up separately with a file server. The server may be hosted locally or provided by a third-party using cloud storage.
- **Preservation** includes the technology that manages multiple long-term storage copies of data and the processes needed for continual archival maintenance.²

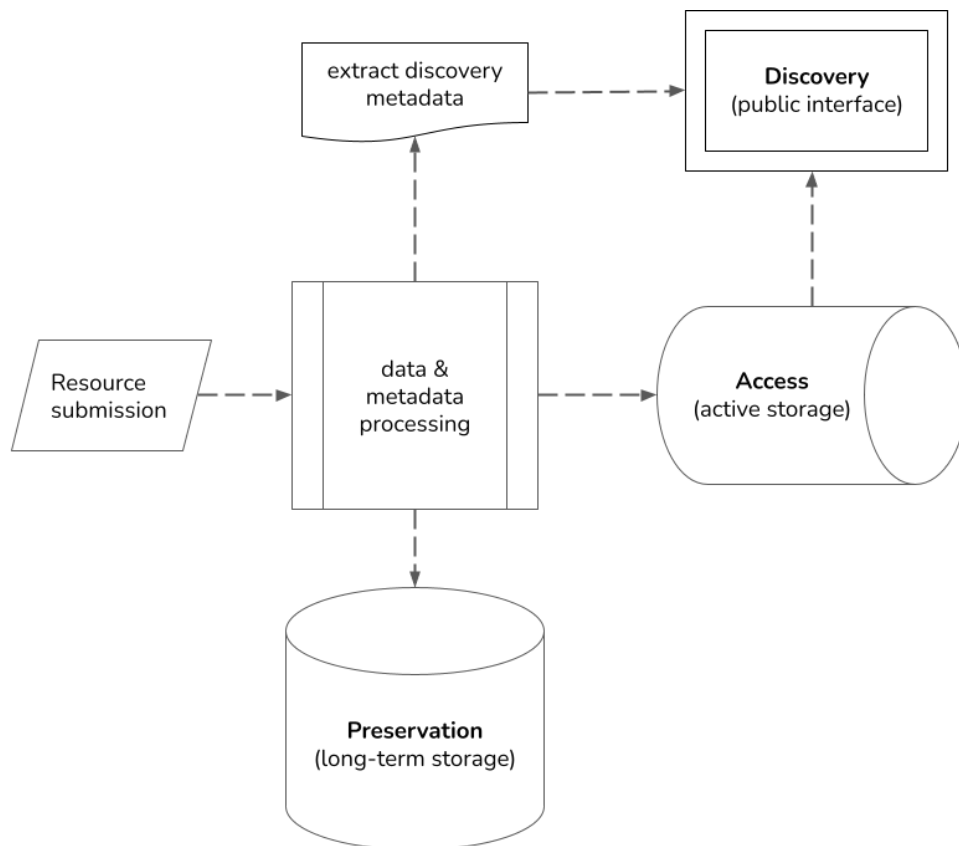


Chart 1. A diagram showing how the functions of Discovery, Access, and Preservation fit together in a workflow

¹ See [“Discovery platform” on page 13 of the Archiving Implementation Workgroup report](#) for the desired functionality for an archival discovery platform.

² See [“Internal Preservation System Activities” on page 17 of the Archiving Implementation Workgroup report](#) for required preservation activities.

Content Organization

Each application we reviewed provided a slightly different way of organizing and presenting data resources. These models are an important factor in choosing an application because they drive the resulting user interface, which may privilege specific browsing and search patterns.

One aspect of content organization that we considered is a hierarchy. If users know what organization created the data they are looking for, it can be grouped under the organization’s name. For a second level, a data series could be grouped together as a “collection” or a “series.” This is how we assembled the hierarchy for our test installations.

Another aspect of content organization is how items can be related to one another. For example, an interface that only provides the most recent version of a given resource can present a simple list of items to the user. In contrast, a spatial data infrastructure that includes archived resources requires a more intricate model that includes multiple relationships, dates, and versions of a dataset. These defined relationships help users to discern between very similar datasets. See [Appendices 1-4](#) for charts of each application’s content organization model.

Discovery Metadata

Front-end data discovery platforms use metadata to enable searching and browsing by various categories, such as place names, topic categories, or dates. Each application under consideration uses a metadata schema that is based upon [Dublin Core](#), a general-purpose schema that is widely used on the web. This discovery metadata does not need to be created from scratch; it can be extracted from the descriptive metadata in the Minnesota Geographic Metadata Guidelines (MGMG).³ The model that defines this conversion is known as a **metadata crosswalk**. See [Appendix 5](#) for a preliminary metadata crosswalk showing how MGMG terms could be mapped to Dublin Core.

Overview of software applications

We tested four different software applications. All of the subgroup participants had experience with one or more of the options, and we discussed the pros and cons of each tool. In the case of CKAN, the data was already present in the Commons, and we used that for evaluation. For the other tools, we stood up or utilized development versions of the software and loaded test datasets from the Minnesota Department of Agriculture into it. We took note of the type of data the application was designed for, which sector typically uses it, which repository functions it offers, and our primary observations. In the appendices, we have included each application’s metadata schema, content organization model, screenshots showing the public interface views, search capabilities, and workflows for ingesting resources and metadata.

³ The majority of the resources currently proposed for inclusion in an archive all have MGMG metadata. [See pg 11 of the Archiving Implementation Workgroup report for a more detailed discussion.](#)

1. CKAN

Description: open-source search portal for general data (<https://ckan.org>)

Typical sector: government

Functions: discovery, access

Comments: The Commons already uses CKAN, so an archive platform built with the same tool could be customized with an interface that looks similar. While this public familiarity would be a meaningful benefit, the interface would need considerable developer time to customize it to accommodate the needs of an archive, including setting up item relations, temporal coverage, and for clarifying dataset versions.

More details: [Appendix 1](#)

2. DSpace

Description: open-source digital repository software for all types of digital content (<https://duraspace.org/dspace>)

Typical sector: academic and non-profit institutions

Functions: discovery, access

Comments: DSpace is the most general-purpose solution we tested. While not explicitly designed for geodata, it can handle many file formats. It does not include built-in preservation, but there is a model to follow that does: the UMN Libraries provides a custom underlying preservation system that accompanies the [University Digital Conservancy](#), which is built with DSpace.

More Details: [Appendix 2](#)

3. GeoBlacklight

Description: open-source geospatial search portal (<https://geoblacklight.org>)

Typical sector: academic libraries

Functions: discovery

Comments: GeoBlacklight was designed specifically for discovery of geospatial resources. Its benefits include searching by location using a map, multiple item relationship options, and custom geospatial metadata fields. This option does not include asset storage for the dataset files themselves. Consequently, it would require setting up a file delivery system or repository at the outset.

More Details: [Appendix 3](#)

4. Preservica

Description: commercial service, primarily a preservation system with an optional discovery platform (<https://preservica.com>)

Typical sector: archives and museums

Functions: discovery, access, preservation

Comments: Available as a hosted or on-premise solution, this licensed product would have annual fees that may or may not include the required amount of storage out of the box. File storage can be purchased or provided onsite. Preservica can manage multiple copies of files, create and verify checksum values, and convert/migrate common file formats (currently not geodata files). As far as we could determine, this option offers the most comprehensive functions, but the least customizable front-end interface for geospatial searching.

More Details: [Appendix 4](#)

Chart 2. A comparison of selected attributes for each application

	Geoblacklight	CKAN	Dspace	Preservica
Example installation	BTAA Geoportal	MN Geospatial Commons	UMN Digital Conservancy	MNHS Collections Online
Free, open-source tool	x	x	x	
Hosted version managed as a paid service				x
Designed for geospatial data	x			
Designed for any type of data		x	x	
Requires in-house developers	x	x	x	
High costs associated with large storage size				x
Easily customizable interface search options	x	x	x	

Combining Applications

Since most of the reviewed applications do not provide all of the necessary repository functions, an archive may need to use more than one to develop a repository that fits their needs. It should be noted that the more systems and tools used to build a repository, the more staff and resources are needed to manage it over time.

Of the four tools we tested, only one (Preservica) has built-in preservation functions. The other tools would require adding on a preservation system to manage and preserve the materials in perpetuity. If, for example, the project was hosted by the University of Minnesota Libraries, the digital preservation department could provide preservation management using their existing tools. If hosted elsewhere, a preservation solution would need to be evaluated and selected.⁴

Likewise, one tool (GeoBlacklight) does not offer built-in access functions and would require a separate file server. However, the benefit gained from using a dedicated geospatial platform may outweigh the cost of the labor required to set up a file access service.

We had hoped to determine which combination of tools would be the most efficient and practical to combine. However, without knowing where the repository would be hosted, these determinations could not definitively be made.

Chart 3. A comparison of repository functions for each application

Application	Discovery	Access	Preservation
CKAN	X	X	
DSpace	X	X	
Geoblacklight	X		
Preservica	X	X	X

⁴ Preservation [strategies](#) are utilized to develop systems that [implement](#) activities addressing preservation needs. These links document the activities that the University of Minnesota Libraries are undertaking in their digital preservation program.

Appendices: Technical Details

Appendix 1. CKAN Application Details

Evaluation Instance: We examined the Commons and discussed its architecture to see how it may work as a geospatial data archive. Note that the Commons does not use the CKAN file storage options. Instead, that function is provided by a custom tool, the [Geospatial Data Resource Site \(GDRS\)](#).

Metadata schema: CKAN metadata or [Data Catalog Vocabulary \(DCAT\)](#)

Content Organization: CKAN uses the [CKAN Domain Model](#). This application structures all of its content by Groups, Datasets, and Resources. In the Commons, Groups are known as Organizations. Examples of Organizations in the Commons are the Minnesota Agriculture Department, Dakota County, or the University of Minnesota. Each Organization holds Datasets, which are single-page entries in the Commons. Datasets may contain Resources, which can be data files for download, metadata files, thumbnail images, or external links to web services and interactive applications. Datasets can be linked via Package Relationships, but the Commons does not utilize this option.

CKAN Content Organization Model

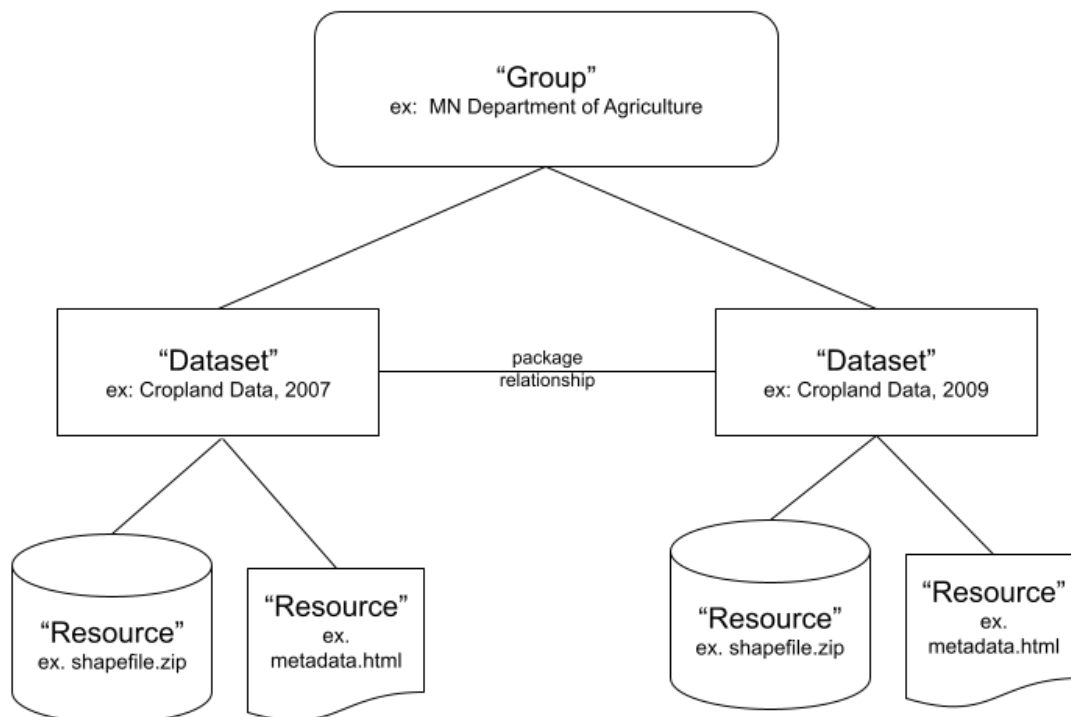


Figure 1.1: CKAN Content Organization Model

Screenshots

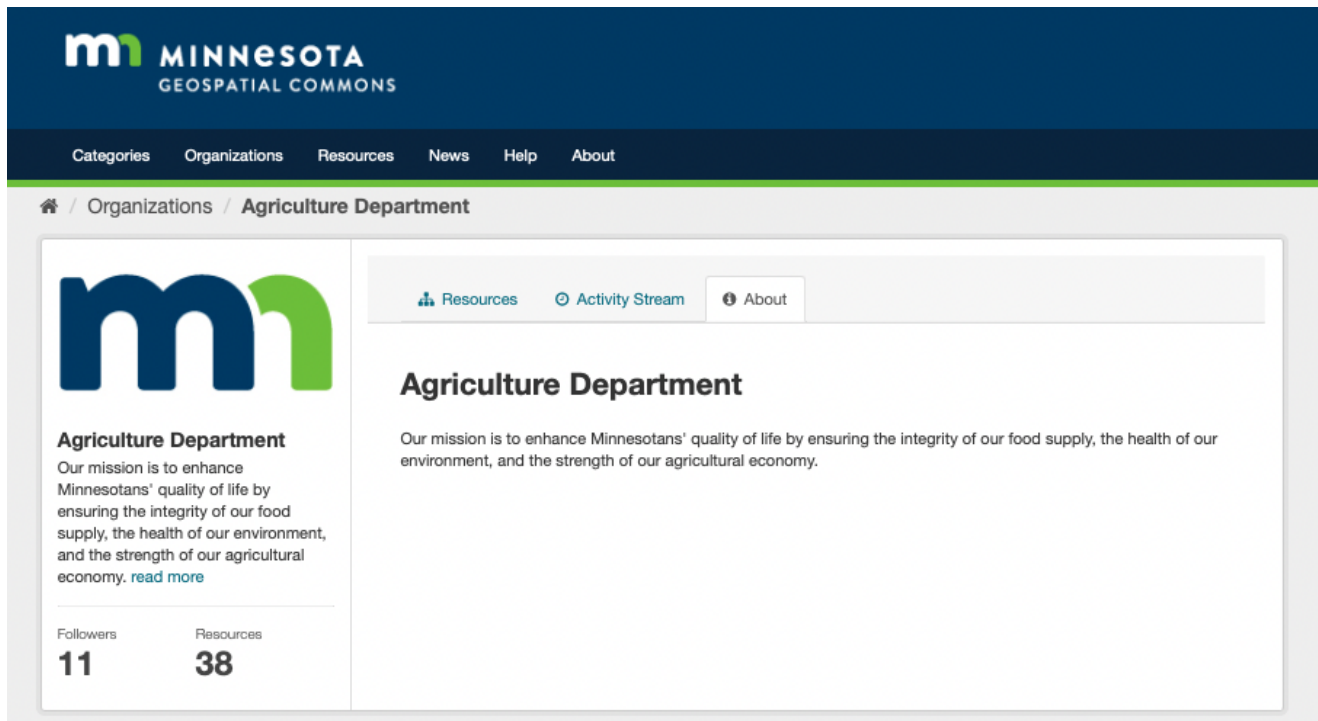



Figure 1.2: Group (Organization) page in the CKAN public interface. There are multiple tabs at the top and the user can toggle to “Resources” to view all of the items from the Agriculture Department.

Cropland Data Layer 2015, Minnesota

Followers
0

Organization



Agriculture Department
Our mission is to enhance Minnesotans' quality of life by ensuring the integrity of our food supply, the health of our environment, and the strength of our agricultural economy. [read more](#)

Social

Twitter

Facebook

License

License not specified
[Legal disclaimer](#)

Resource Categories Activity Stream

Cropland Data Layer 2015, Minnesota

The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land cover data layer.

This is the 2015 NASS USDA CDL dataset for Minnesota. The dataset was clipped and downloaded using the USDA NASS CropScape online web application: <http://nassgeodata.gmu.edu/CropScape/>

The Minnesota Department of Agriculture (MDA) did minimal processing on this dataset to make more useful for Minnesota-specific work. The lineage section describes these steps taken, but the full description of the Cropland Data Layer can be found in the [NASS USDA Metadata](#).

Static Preview - Sample Image

View

ESRI File Geodatabase

Download

Shapefile

Download

Full Metadata Record

View

agriculture

crop

crop cover

crop estimates

cropland

crops

cropland

cropland


cropland

land cover

Additional Info

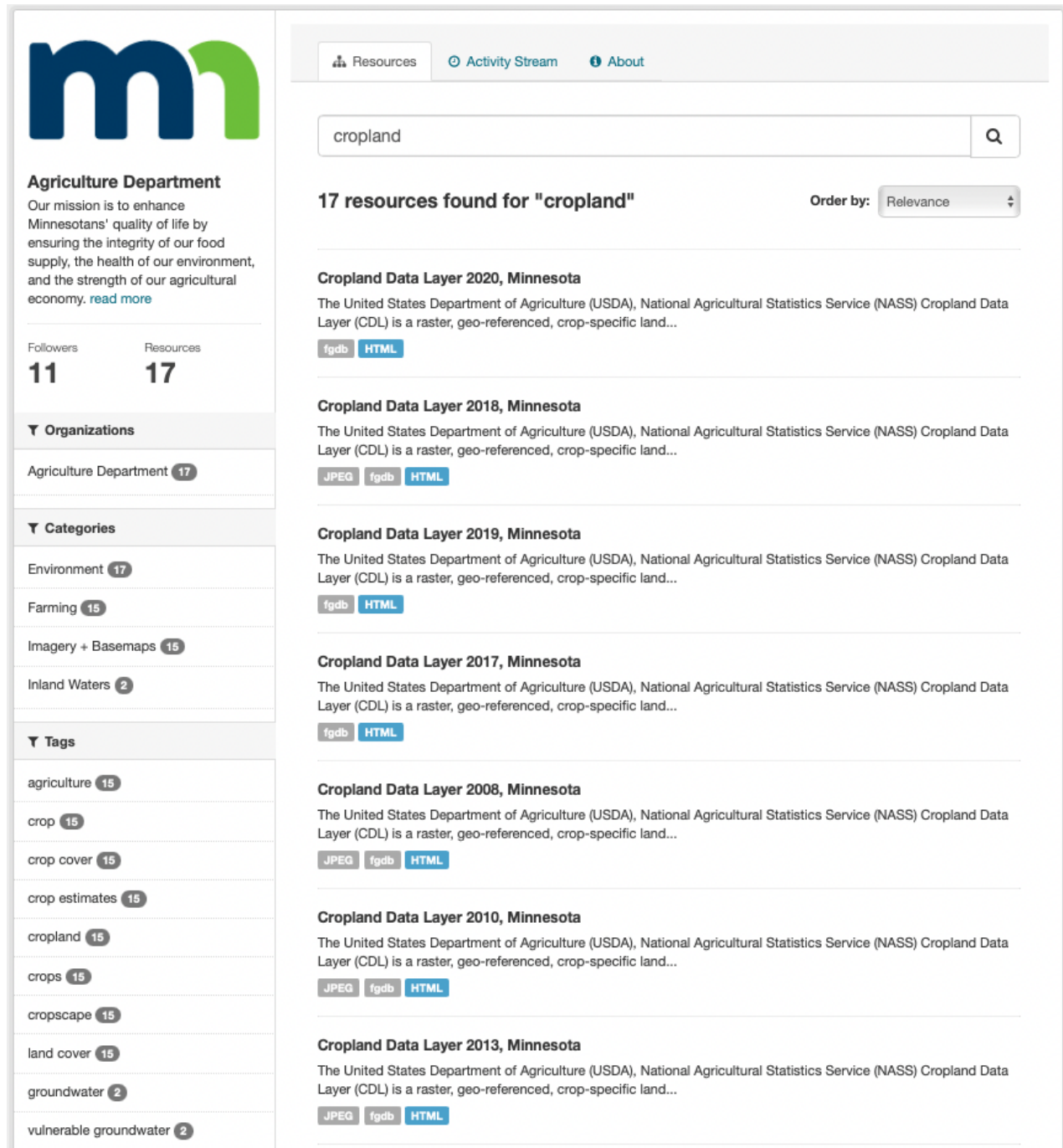
Access constraints	None.
Date details	Data is current as of time period of content.
Purpose	The purpose of the Cropland Data Layer Program is to use satellite imagery to (1) provide acreage estimates to the Agricultural Statistics Board for the state's major commodities and (2) produce digital, crop-specific, categorized geo-referenced output products.

Dataset extent



Map data © OpenStreetMap contributors
Tiles by Stamen Design (CC BY 3.0)

Figure 1.3: Dataset and Resources page in the CKAN public interface. This page includes a predefined set of information, including title, description, keywords, and links to external files. The Commons has also been configured to display "Additional Info", including Access constraints, Date details, and Purpose. It also shows a map inset displaying the dataset extent.



The screenshot shows the CKAN public interface search results for the term "cropland". The left sidebar features the Minnesota Department of Agriculture logo and mission statement, along with statistics (11 followers, 17 resources) and lists of organizations, categories, and tags. The main content area displays 17 resources found, ordered by relevance. Each resource entry includes a title (e.g., "Cropland Data Layer 2020, Minnesota"), a description, and available download formats (e.g., fgdb, HTML, JPEG).

Resources | **Activity Stream** | **About**

cropland

17 resources found for "cropland" **Order by:** Relevance

Cropland Data Layer 2020, Minnesota
The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land...
[fgdb](#) [HTML](#)

Cropland Data Layer 2018, Minnesota
The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land...
[JPEG](#) [fgdb](#) [HTML](#)

Cropland Data Layer 2019, Minnesota
The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land...
[fgdb](#) [HTML](#)

Cropland Data Layer 2017, Minnesota
The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land...
[fgdb](#) [HTML](#)

Cropland Data Layer 2008, Minnesota
The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land...
[JPEG](#) [fgdb](#) [HTML](#)

Cropland Data Layer 2010, Minnesota
The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land...
[JPEG](#) [fgdb](#) [HTML](#)

Cropland Data Layer 2013, Minnesota
The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land...
[JPEG](#) [fgdb](#) [HTML](#)

Figure 1.4: Search page in the CKAN public interface. There are only a few available facets or filters in the Commons to filter results, including Organizations, Categories, Tags, Formats, and Licenses. It also offers a basic spatial search on some pages.

Appendix 2. DSpace Application Details

Evaluation Instance: We worked with the University of Minnesota Libraries to stand up an out-of-the-box instance of DSpace and add geospatial data from the Commons and Minnesota Historical Society into it for testing. Customizations, while possible, were not done or explored.

Metadata schema: Dublin Core or [DC-Lib](#)

Data Model: DSpace can be set up with a data model featuring up to four hierarchy levels. Its terminology is Communities (level 1), Sub-Communities (level 2), Collections (level 3), and Items (level 4).

We used the structure of The Commons as an example of how resources could be organized within DSpace. We utilized the Organization name as a Community with each of the dataset series as a Collection. We decided that using Sub-Communities would make things harder to understand for the user as Sub-Communities are listed first on a page, followed by the Collections. Placing everything under the same Collection was more practical.

DSpace Content Organization Model

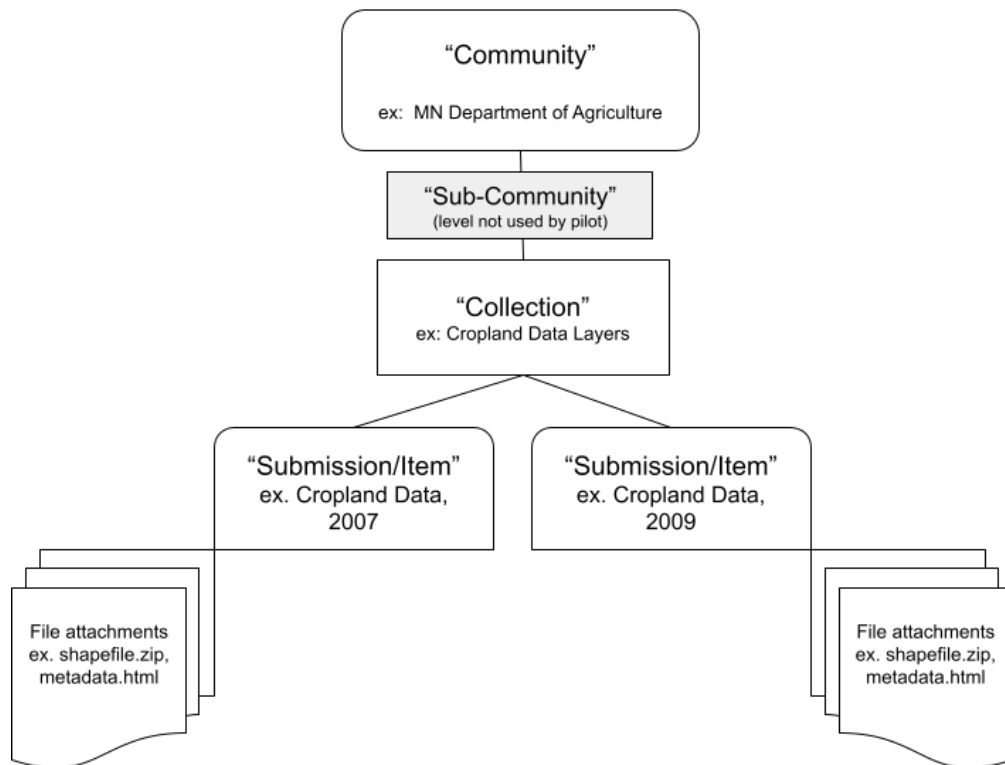


Figure 2.1: DSpace Content Organization Model

Workflow

To add content to DSpace, the administrator or content editor must log into the administration backend.

To create a new community, the user selects “Create Community” on the right sidebar and walks through a few screens. These screens ask for information about the Community, such as a name, short description, introductory text, copyright text, and a logo. After the Community is created, roles can be assigned to help control edit access to the Community. It is not expected that roles would be used for adding Communities, as this would be done by an administrator of the system.

Next, an administrator can add Collections to a Community. Collections could contain one or more datasets. For example, the cropland data from the Agriculture Department has multiple years of data available. Adding a Collection to a Community is a similar process as creating a Community. The first screen asks for the Collection name, short description, introductory text, copyright text, license information, provenance information, and logo. The next page allows for adding roles for access to the Collection.

After a Collection is created, administrators must go to the landing page and click “submit a new item to this collection.” When adding an item, the following information is requested: author/s, title, other titles, date of issue, publisher, citation, identifiers, type, language. Another screen asks for keywords, an abstract, sponsors, description. The next screen is where file/s can be uploaded. Each file that is uploaded should also include a file description. After all the files are uploaded, a review screen is displayed prior to being completed. The next screen is a distribution license (which can be edited) that is an acknowledgment of having the rights to share and make the content available publicly. Clicking the submission button finalizes the process and adds the content to the collection.

Screenshots

The screenshot displays the DSpace public interface for the 'Agriculture Department' community. The main content area on the left features the community title, a 'BROWSE BY' section with buttons for 'By Issue Date', 'Authors', 'Titles', and 'Subjects', and a search bar with a 'Go' button. Below this is the Minnesota Department of Agriculture logo and a mission statement. The right sidebar contains a search bar, radio buttons for 'Search DSpace' and 'This Community', and a 'BROWSE' section with a list of categories: 'All of DSpace', 'Communities & Collections', 'By Issue Date', 'Authors', 'Titles', 'Subjects', 'This Community' (highlighted), 'By Issue Date', 'Authors', 'Titles', and 'Subjects'. Further down the sidebar are 'MY ACCOUNT' links for 'Login' and 'Register', and a 'DISCOVER' section with an 'Author' button.


Agriculture Department

BROWSE BY

By Issue Date Authors Titles Subjects

Search within this community and its collections:

Go



The mission of Agriculture Department is to enhance Minnesotans' quality of life by ensuring the integrity of our food supply, the health of our environment, and the strength of our agricultural economy. These datasets help provide data that supports our mission.

Copyright text

Sub-communities within this community

Cropland Data Layer Minnesota

The purpose of the Cropland Data Layer Program is to use satellite imagery to (1) provide acreage estimates to the Agricultural Statistics Board for the state's major commodities and (2) produce digital, crop-specific, ...

Collections in this community

Agriculture Chemical Incidents

(Short Des) The Minnesota Department of Agriculture is the lead agency for response to, and cleanup of agricultural chemical contamination in Minnesota. Because of this role the Minnesota Department of Agriculture has ...

Cropland Data Layer Minnesota

Multiple years of the United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land cover data layer.

Emerald Ash Borer Detection, Minnesota

This suite of data is a collection of layers that communicate the introduction risk, detection, bioControl, and response to Emerald Ash Borer (EAB) in Minnesota, including quarantined counties.

Search

☒ Search DSpace
☐ This Community

BROWSE

All of DSpace

Communities & Collections

By Issue Date

Authors

Titles

Subjects

This Community

By Issue Date

Authors

Titles

Subjects

MY ACCOUNT

Login

Register

DISCOVER

Author

Figure 2.2: Community page in the DSpace public interface. This screenshot shows how Sub-communities and Collections are displayed under a Community

Multiple years of the United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land cover data layer.

[Submit a new item to this collection](#)

Recent Submissions

2009 Cropland Data Layer, Minnesota

Unknown author (Minnesota Department of Agriculture, 2015-10-23)

The purpose of the Cropland Data Layer Program is to use satellite imagery to (1) provide acreage estimates to the Agricultural Statistics Board for the state's major commodities and (2) produce digital, crop-specific, ...

2007 Cropland Data Layer, Minnesota

Unknown author (Minnesota Department of Agriculture, 2015-10-23)





The purpose of the Cropland Data Layer Program is to use satellite imagery to (1) provide acreage estimates to the Agricultural Statistics Board for the state's major commodities and (2) produce digital, crop-specific, ...

Figure 2.3: Collection page in the DSpace public interface. This shows the Cropland Data Layer Collection in the Agriculture Department Community. This collection contains multiple years of data, each listed individually.

Public Drainage Systems Hydrography, Buffalo-Red River Watershed District, Minnesota

No Thumbnail

View/Open

-  [HTML Metadata for Public Drainage Systems Hydrography, Buffalo-Red River Watershed District, Minnesota \(12.64Kb\)](#)
-  [XML Metadata for Public Drainage Systems Hydrography, Buffalo-Red River Watershed District, Minnesota \(5.774Kb\)](#)
-  [Shapefile for the Public Drainage Systems Hydrography, Buffalo-Red River Watershed District, Minnesota \(317.1Kb\)](#)
-  [ESRI File Geodatabase for Public Drainage Systems Hydrography, Buffalo-Red River Watershed District, Minnesota \(232.2Kb\)](#)

Date

2020-04-28

Metadata

[Show full item record](#)

This Feature Dataset contains those public drainage system (as defined by Minnesota Statute 103E) layers in Buffalo-Red River Watershed District, Minnesota that are considered hydrographic. Currently those layer are: The ditch/tile centerlines, drainage structures, and watersheds for ditch/tile systems (ditchsheds).

URI

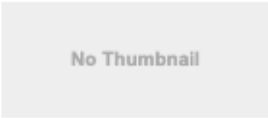
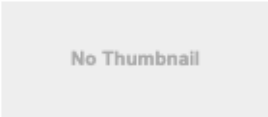
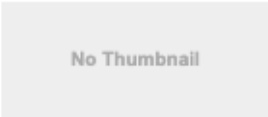
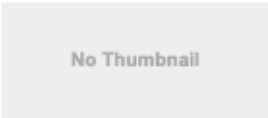
<http://hdl.handle.net/16161/26>

Collections

Public Drainage Systems Hydrography, Buffalo-Red River Watershed District, Minnesota

Figure 2.4: Submission/Item page in the DSpace public interface. This view shows a few basic metadata fields along with links to open or download files.

Files in this item

	Name: metadata.html Size: 12.64Kb Format: HTML Description: HTML Metadata for Public Drainage ...	View/Open
	Name: metadata.xml Size: 5.774Kb Format: XML Description: XML Metadata for Public Drainage ...	View/Open
	Name: shp_water_public_drainage_syst ... Size: 317.1Kb Format: Unknown Description: Shapefile for the Public Drainage ...	View/Open
	Name: fgdb_water_public_drainage_sys ... Size: 232.2Kb Format: Unknown Description: ESRI File Geodatabase for Public ...	View/Open

This item appears in the following Collection(s)

- Public Drainage Systems Hydrography, Buffalo-Red River Watershed District, Minnesota
This Feature Dataset contains those public drainage system (as defined by Minnesota Statute 103E) layers in Buffalo-Red River Watershed District, Minnesota that are considered hydrographic. Currently those layer are: The ditch/tile centerlines, drainage structures, and watersheds for ditch/tile systems (ditchsheds).

[Show simple item record](#)

Figure 2.5: File attachments as shown in the DSpace public interface. When viewing the files in the repository, thumbnails are displayed if provided, and the View/Open buttons either show a preview of the file or a download link. Currently, the view/Open button for GIS files will only download the files as there is no GIS data viewer in DSpace.

DSpace Repository

DSpace is a digital service that collects, preserves, and distributes digital material. Repositories are important tools for preserving an organization's legacy; they facilitate digital preservation and scholarly communication.

Communities in DSpace

Select a community to browse its collections.

Agriculture Department
Our mission is to enhance Minnesotans' quality of life by ensuring the integrity of our food supply, the health of our environment, and the strength of our agricultural economy.

Aitkin County, Minnesota
Resources related to Aitkin County, Minnesota

Buffalo - Red River Watershed District
The Buffalo-Red River Watershed District, a local government unit and legal ditch authority located in NW MN, covers an area of 1,785 square miles in portions of Clay, Becker, Wilkin and Otter Tail Counties.

Freeborn County, Minnesota
Resources related to Freeborn County, Minnesota

Governor's Citizen Advisory Commission on Redistricting
Records documenting the activities of this citizen advisory commission formed to review proposed redistricting plans developed by the Legislature (for both congressional and legislative districts) and to provide advice to the Governor's Office about those plans.

Search

BROWSE

- All of DSpace
- Communities & Collections
- By Issue Date
- Authors
- Titles
- Subjects

MY ACCOUNT

- Login
- Register

DISCOVER

Author

- Borchardt, Erich (1)
- Elsner, Gary (1)

Subject

- Agriculture (5)
- crop (4)

Figure 2.6: Search page in the DSpace public interface. A few facets are part of the default interface, but more could be added via customization.

Appendix 3. GeoBlacklight Application Details

Evaluation Instance: We tested this option by using a development instance of GeoBlacklight in use by the Big Ten Academic Alliance Geoportal, which is hosted by the University of Minnesota Libraries.

Metadata schema: [OpenGeoMetadata](#)

Content Organization: All record entries in GeoBlacklight are of the same type. It does not have separate types of records for collections or groups. It is essentially a flat system, whereby items are related to each other via metadata fields, including "Member Of," "Is Part Of," "Is Version Of," "Replaces," "Source," and a general "Relation." These linked fields can present records as hierarchies with parent/child/grandchild relationships or as siblings. This flexibility allows collections to be nested (i.e., a collection can "belong to" another collection). These relationships can also connect data layers about similar topics from different years within an organization.

Geoblacklight Content Organization Model

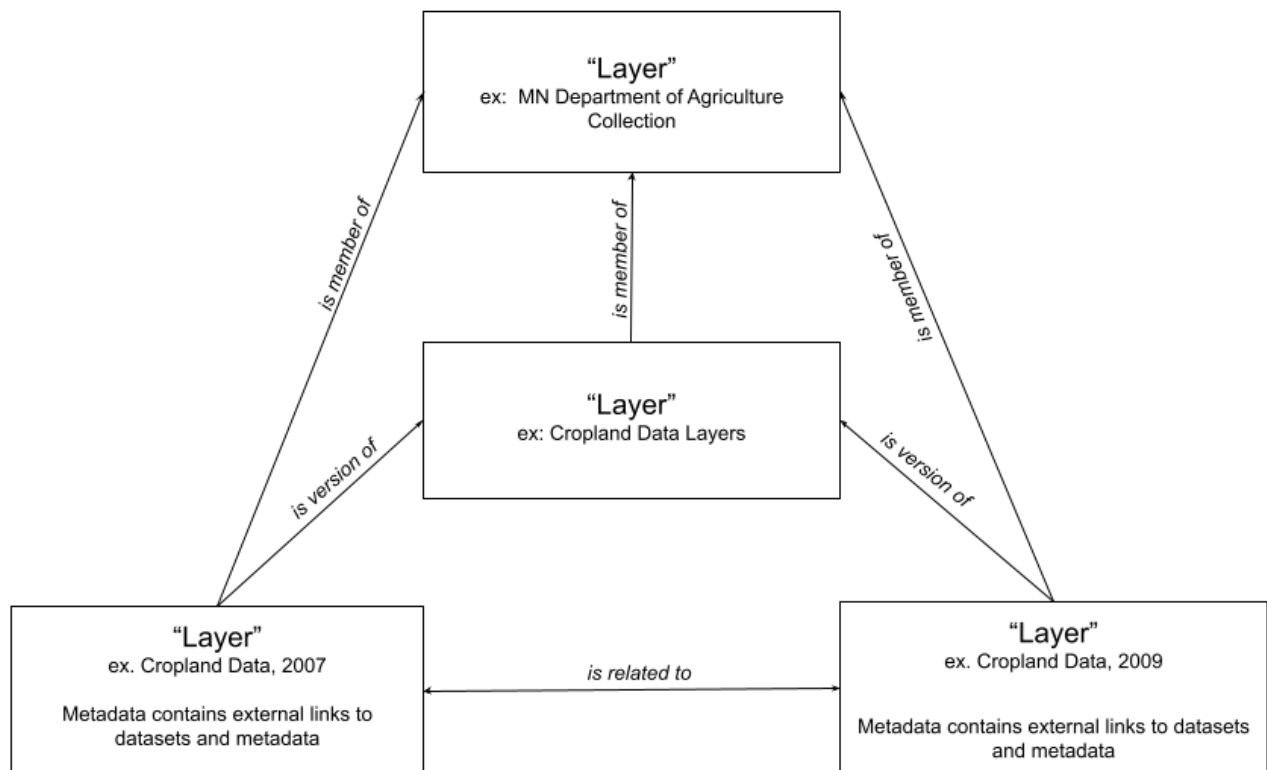


Figure 3.1: GeoBlacklight Content Organization Model

Workflow

GeoBlacklight has an optional backend administration tool, GEOMG, which is a graphical user interface for manual record creation, batch creation via spreadsheets, bulk editing, and metadata validation.

To add a single record, an administrator would log into GEOMG and click “New Document.” This action opens a form-based view for manually entering metadata. The required fields are Title, Access Rights, and ID. Fields with controlled values, such as “Resource Class,” feature dropdown boxes. Items can be previewed in the public interface before they are made public. Since GeoBlacklight does not include file access, there is no dataset upload page. Instead, there are several designated fields for external links, including Download, Information, Documentation, various metadata standards, and geospatial web services.

To add multiple records, an administrator would create a spreadsheet of metadata. This spreadsheet is then imported into the GEOMG tool. Although it is easier to label each column with a predetermined set of terms, the import function features a crosswalk page that allows an administrator to map fields manually. GEOMG also exports records in batches to a CSV format. An administrator can then make item-by-item changes and re-import the CSV for batch changes.

GEOMG can display dashboards showing the results of custom queries, such as the total number of records, records added during a set time period, or a count of items by a specified metadata field.

Screenshots

[Clear search](#) [Back to Search](#)

Cropland Data Layers

Creator: National Agricultural Statistics Service (NASS) and United States Department of Agriculture (USDA)

Description: This data series includes annually created raster, geo-referenced, crop-specific land cover data layer. The layers are from the United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is

Place: [Minnesota](#)

Resource Class: [Collections](#)

Resource Type: [Vector data](#)

ISO Topic Category: [Farming](#)

Temporal Coverage: 2007-2020

Provider: [Agriculture Department](#)

Rights: This data is subject to the Minnesota Government Data Practices Act. All data that is made available as part of the Archive is free for any use.

Location

Links

[Cite](#)

Belongs to collection...

[Agriculture Department](#)

Has version...

- [Cropland Data, 2015](#)
- [Cropland Data, 2007](#)
- [Cropland Data, 2009](#)

Figure 3.2: Dataset Series page in the GeoBlacklight public interface. The view shows both Dublin Core and custom geospatial metadata fields. There is also a map inset showing the dataset extent. The bottom of the screen displays item relationships. For this record, item relationships provide links to a larger Collection page or to child records for dated versions.

Clear search

Back to Search

« Previous | 2 of 3 | Next »

Cropland Data, 2015

Creator:

National Agricultural Statistics Service (NASS), United States Department of Agriculture (USDA)

Description:

The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land cover data layer. This is the 2015 NASS USDA CDL dataset for Minnesota. The dataset was clipped and downloaded using the USDA NASS CropScape online web application: <http://nassgeodata.gmu.edu/CropScape/The> Minnesota Department of Agriculture (MDA) did minimal processing on this dataset to make more useful for Minnesota-specific work. The lineage section describes these steps taken, but the full description of the Cropland Data Laye...
[Read more](#)

Place:

[Minnesota](#)

Subject:

Land cover

Resource Class:

[Datasets](#)

Resource Type:

[Vector data](#)

ISO Topic Category:

[Environment](#)

Date Published:

2016-05-13

Temporal Coverage:

2015


Provider:

[Minnesota Geospatial Commons](#)


Format:


Shapefile

Location



Links

 [Metadata](#)

 [Cite](#)

Downloads

Original Shapefile

Belongs to collection...

[Agriculture Department](#)

Related Records

[Cropland Data, 2007](#)

[Cropland Data, 2009](#)

Is version of...


 [Cropland Data Layers](#)

Figure 3.3: Item page in the GeoBlacklight public interface. This record is for a dataset layer with a file download. It links to Related Records that are from the same series, but a different year.

You searched for:

Limit your search 1 - 5 of 5

Sort by Relevance ▾ 20 per page ▾

Place ▾

[Minnesota](#) 5

Resource Class ▾

[Datasets](#) 4

[Collections](#) 1

Resource Type ▾

[Vector data](#) 5

Subject ▾

[Census](#) 1

[Land cover](#) 1

Year ▾

-

Time Period ▾

[2005-2009](#) 3

[2015-present](#) 2

Creator ▸

Provider ▸

Public/Restricted ▸

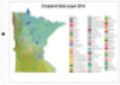










-  [Cropland Data, 2015](#) 
2015. National Agricultural Statistics Service (NASS), United States Department of Agriculture (USDA). The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-refe...
-  [Cropland Data Layers](#) 
2007-2020. National Agricultural Statistics Service (NASS) and United States Department of Agriculture (USDA). This data series includes annually created raster, geo-referenced, crop-specific land cover data layer. The layers are from the United States Depar...
-  [Cropland Data, 2007](#) 
2007. National Agricultural Statistics Service (NASS) and United States Department of Agriculture (USDA). The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-refe...
-  [Cropland Data, 2009](#) 
2009. National Agricultural Statistics Service (NASS) and United States Department of Agriculture (USDA). The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster, geo-refe...
-  [Fall Nitrogen Fertilizer Application Restrictions: Minnesota](#)  
2020. Minnesota Department of Agriculture. This Vulnerable Groundwater Areas Feature Class illustrates quarter sections that meet at least one of 3 conditions across the state. The first con...


Figure 3.4: Search page in the GeoBlacklight public interface with the list view toggled on. This option allows users to filter by year or time period.

The screenshot shows the GeoBlacklight public interface search page. At the top, there is a search bar with the text "Search..." and a "Search" button. To the right of the search bar is an "Advanced" link. Below the search bar, there is a section for "You searched for:" which includes filters for "Place > Minnesota" and "Resource Class > Datasets". A "Clear search" button is located to the right of these filters. Below the search filters, there is a bounding box input field containing the coordinates "-103.051912 39.13006 -84.287264 52.509535".

The main content area is titled "Limit your search" and includes a "Previous | 1 - 20 of 1,200 | Next" navigation bar. Below this, there are two columns of search results. The left column is titled "Place" and lists various locations in Minnesota, including Minnesota (1,200 results), Hennepin County, Minnesota (408 results), Anoka County, Minnesota (360 results), Ramsey County, Minnesota (339 results), Dakota County, Minnesota (264 results), Carver County, Minnesota (256 results), Scott County, Minnesota (248 results), and Washington County, Minnesota (189 results). The right column is titled "Resource Class" and lists various datasets, including 1-Foot Contours, Apple Valley: Dakota County, Minnesota; 1-Foot Contours, Burnsville: Dakota County, Minnesota; 1-Foot Contours, Eagan: Dakota County, Minnesota; 1-Foot Contours, Farmington: Dakota County, Minnesota; 1-Foot Contours, Inver Grove Heights: Dakota County, Minnesota; 1-Foot Contours, Mendota Heights: Dakota County, Minnesota; 1-Foot Contours, Rosemount: Dakota County, Minnesota; 1-Foot Contours, South St Paul: Dakota County, Minnesota; and 1-Foot Contours, West St Paul: Dakota County, Minnesota.

On the right side of the page, there is a map view toggle. The map view is currently selected, and the map shows a view of Minnesota and surrounding areas. The map includes a search bar with the text "Search when I move the map" and a "Search" button. The map also includes a "Map view" button and a "List view" button.


Figure 3.5: Search page in the GeoBlacklight public interface with the map view toggled on. This view allows users to zoom and pan the map to retrieve results based on the map’s spatial extent.

 Cropland Data, 2015 Published

Identification

Descriptive

Title *

Cropland Data, 2015 

Theme: city, state, temporal coverage

Alternative Title

Cropland Data Layer 2015, Minnesota Remove

[Add another Alternative Title](#)

Description

The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a raster Remove

[Add another Description](#)

Language

eng Remove

[Add another Language](#)

Credits

Creator

National Agricultural Statistics Service (NASS), United States Department of Agriculture (USDA) Remove


[Add another Creator](#)

Publisher

Remove


[Add another Publisher](#)

Provider

Minnesota Geospatial Commons 

Categories

Resource Class *

Datasets  Remove

[Add another Resource Class](#)

Resource Type

View in Geoportal

Identification

- Descriptive
- Credits
- Categories
- Temporal
- Spatial
- Relations

Distribution

- Object
- Links

Administrative

- Codes
- Rights
- Life Cycle
- Accessibility
- Institutional Access Links

Figure 3.6: Editing discovery metadata about an individual data resource in the GeoBlacklight/GEOMG administration interface

27

test · Field Mappings

Source header * Title	Destination field * dct_title_s	<input type="checkbox"/> Delimited
Source header * Date Accessioned	Destination field * b1g_dateAccessioned_sm	<input checked="" type="checkbox"/> Delimited
Source header * Alternative Title	Destination field * dct_alternative_sm	<input checked="" type="checkbox"/> Delimited
Source header * ID	Destination field * geomg_id_s	<input type="checkbox"/> Delimited
Source header * Description	Destination field * dct_description_sm	<input checked="" type="checkbox"/> Delimited
Source header * Language	Destination field * dct_language_sm	<input checked="" type="checkbox"/> Delimited

Figure 3.7: Mapping metadata fields for an imported spreadsheet in the GeoBlacklight/GEOMG administration interface

[Edit Import](#) | [Back](#)

Import

Name: [update] Geology topic

Filename: [geomg-export-2022-03-08-2.csv](#) (text/csv)

State: imported

Source:

Description:

[Data Mapping](#)
[Import Results](#)

[Failed](#)
[Success](#)

Displaying successful documents **1-100** of **594** in total

[< Prev](#)
[1](#)
[2](#)
[3](#)
[4](#)
[5](#)
[6](#)
[Next >](#)

State	Title	Identifier
success	Preliminary Landform Subdivisions of Pennsylvania	fc080a56-994c-43b4-a837-95ba71560ea1
success	Interactive map of coal mines in Iowa	03a-01-001
success	Geomorphology - Mississippi River - St. Paul District	40f05973-f1a0-4218-b5bc-fbd5468b254f
success	Historic Oil and Gas well locations from Bureau of Oil and Gas Planning and Program Management PADEP, 2013 WPA Mines, K Sheet, H Sheet	2133d4bc-34f6-43f6-9bd4-e87c5ebfa1bb
success	Historic Wetland Habitats of the National Wetlands Inventory (NWI) for Indiana: Indiana	00b647c9-8bc0-47a2-aa1d-fdc302da0b2f
success	Historic Wetland Habitats of the National Wetlands Inventory (NWI) for Indiana: Indiana (Web Map Service)	da31f8e181fa4273b3d22bea3daea599_0
success	Geochemistry Samples - Coal: Ohio	a59a6e0d7810419da5f8c816536cbdcf_5

Figure 3.8: Batch importing metadata in the GeoBlacklight/GEOMG administration interface

Appendix 4. Preservica Application Details

Evaluation Instance: We tested a free instance of Preservica Starter, to see how data could be ingested and made available. We noted that more functionality with GIS data is on their development roadmap, but the timeline and details are unclear.

Metadata schema: [MODS](#)

Content Organization: Preservica has developed the Open Exchange Format, OPEX, which is essentially a system of nested folders. The system was designed to house various types of material organized by collection. As a result, using the public interface is not the most user-friendly way to find GIS data. The easiest way to currently find something is by knowing the record creator and knowing what type of data you are looking for.

Preservica Content Organization Model

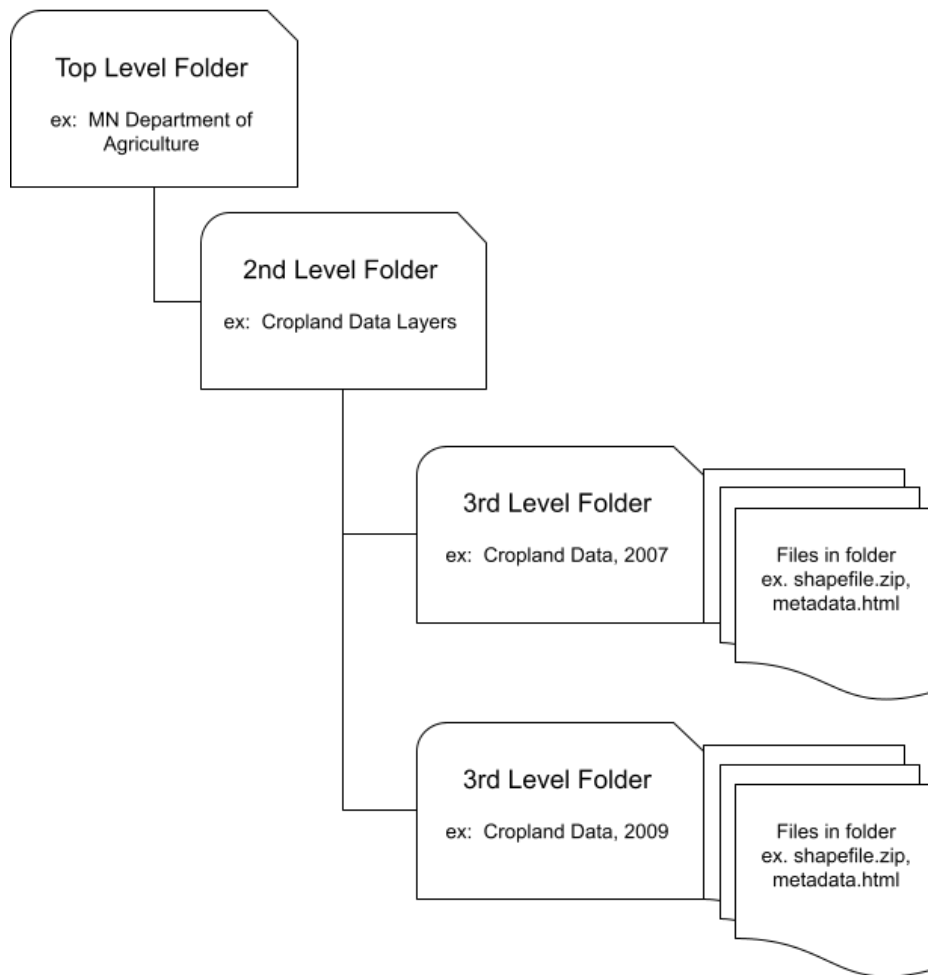


Figure 4.1: Preservica Content Organization Model

Workflow

There is both a public front-end and an administrative back-end to manage in Preservica. Materials marked as public will be shared on the front end.

Adding content to Preservica is easy using drag and drop functionality. Upon ingest, Preservica will perform several actions: create technical metadata (including checksums, number of files, file sizes, structure/organization), identify and validate file formats, and capture file properties for some file formats. Optionally, it can also upload additional XML-based metadata, normalize to pre-selected file formats, and sync to a catalog. Preservica can identify many file formats and automatically create a derivative copy using a different file format if appropriate for preservation purposes. However, this action is not yet available for GIS data.

Before ingesting records, care must be taken to decide how to organize the collections and files. For example, to upload a set of nested folders, they must be in the exact order that they should appear in the repository. To be more selective and upload individual files at a time, an administrator would need to create folders in advance within Preservica.

Once files are in the system, they can be moved around or copied, and an administrator can add metadata about the collection and individual files. The more metadata added to each item, the easier it will be for users to find what they are looking for. However, adding metadata within Preservica is a time-consuming process. Bulk metadata editing is available, but only if the changes are all the same. This is not as useful as using a spreadsheet to associate different pieces of metadata to individual assets.

Preservica also creates a dashboard of the entire archive. At a glance, the dashboard shows: the total amount of space used; the last time an administrator logged in; a link to the resource library, a link to the public portal; a list of recently migrated files; a list of the most recent recycled items, a list of the files with recently edited metadata; and a graphic showing the most common file formats in your repository.

Screenshots

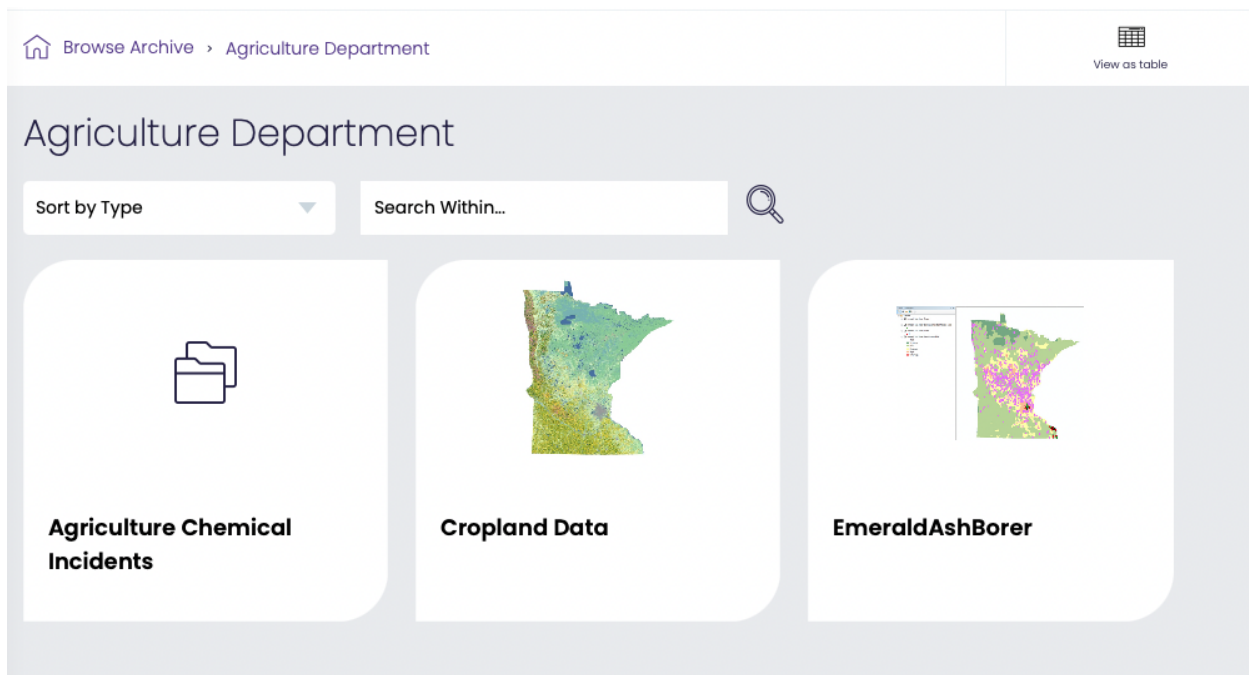


Figure 4.2: Top level folder in the Preservica Starter public interface. This shows the folders that have been added to the Agriculture Department directory.

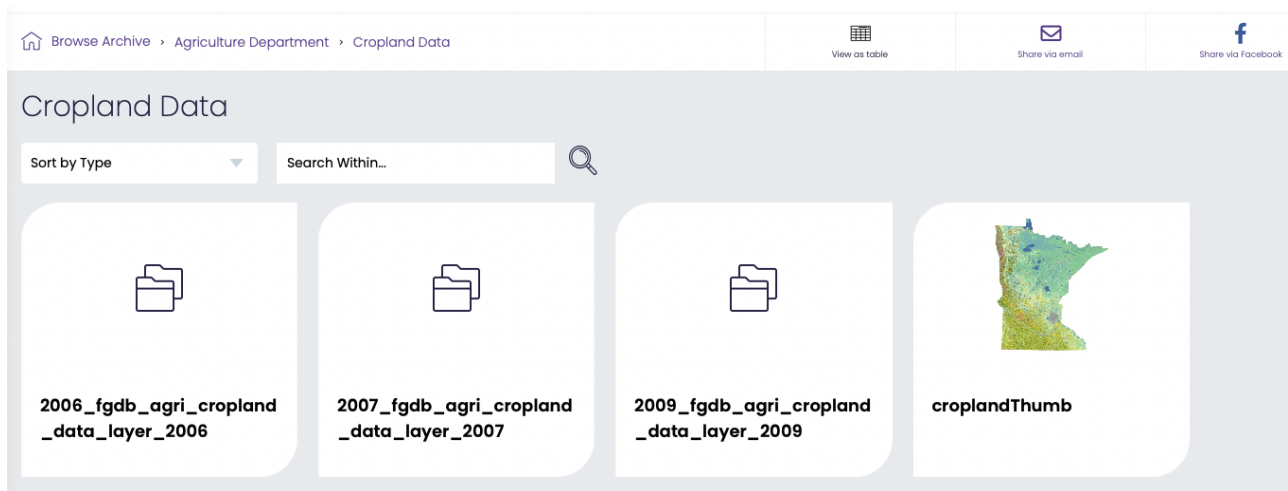


Figure 4.3: 2nd level folder in the Preservica Starter public interface. This shows the publicly shared content within the Cropland Data folder.

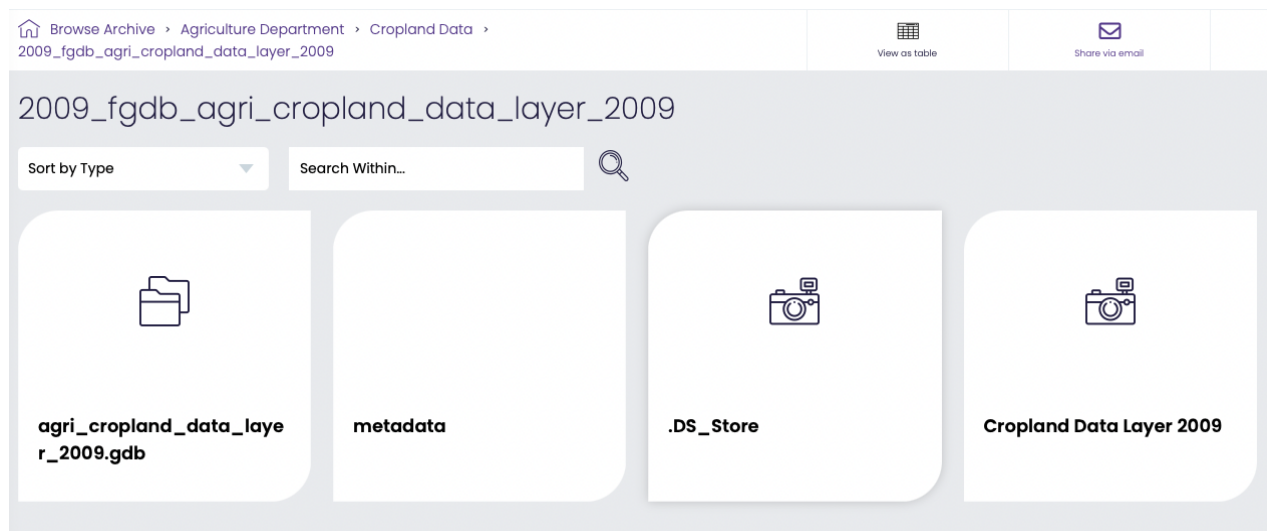


Figure 4.4: 3rd level folder in the Preservica Starter public interface. This shows the files within the Cropland Data 2009 folder.

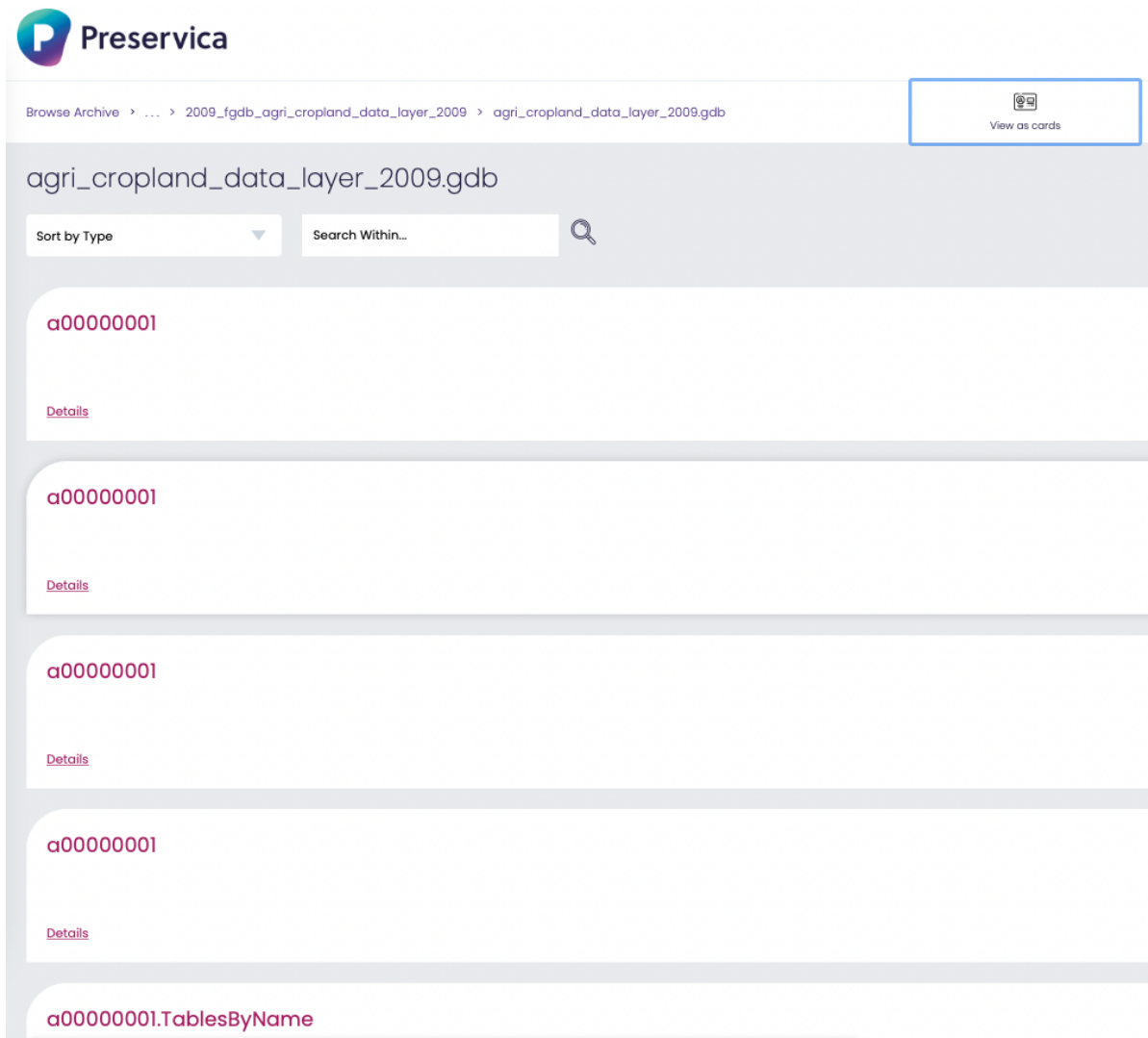


Figure 4.5: Files (in a geodatabase) in the Preservica public interface. A user can click on each of the files to either see a preview of it or a link to a page where the file can be downloaded. From this screen, a user cannot see file extensions. The curator would want to ensure the file name that is displayed would let users know what type of file they were going to open/download. This needs to be done on a folder by folder or file by file level.

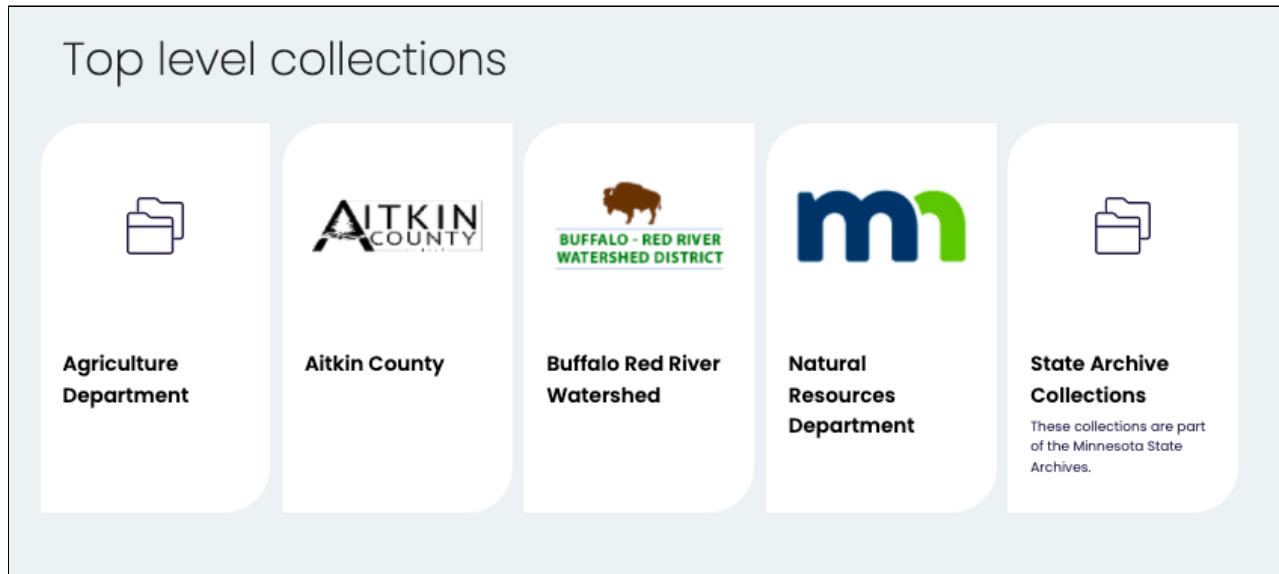


Figure 4.6: Main page of the Preservica Starter public interface. The search options are limited to browsing by folder.

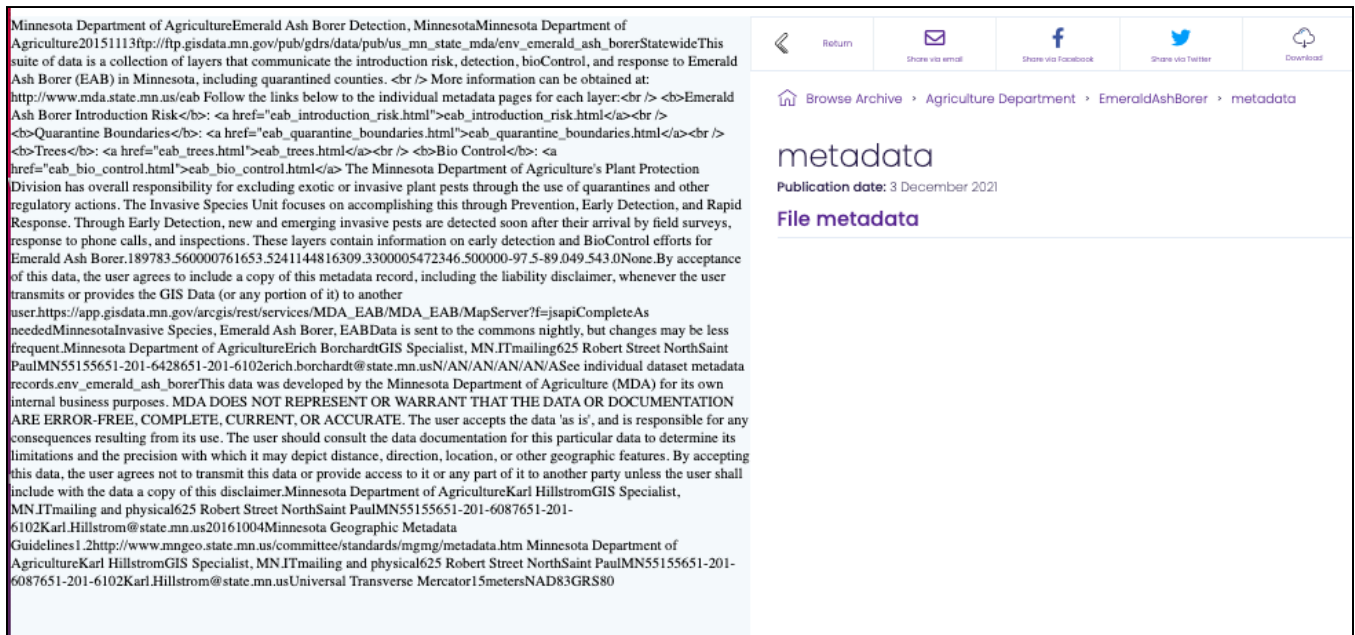


Figure 4.7: XML Metadata in the Preservica Starter public interface. This screenshot is an example of how the XML metadata file may be displayed for this type of file. It's available, but not pretty.

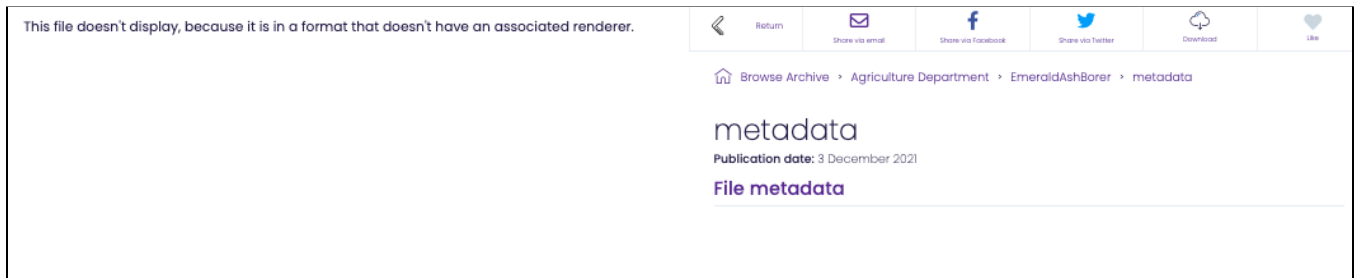


Figure 4.8: HTML Metadata in the Preservica Starter public interface. This is an example of what it looks like when there is no viewer. A note says it can't be displayed. Users can click the cloud/download button to download the file to their computers.

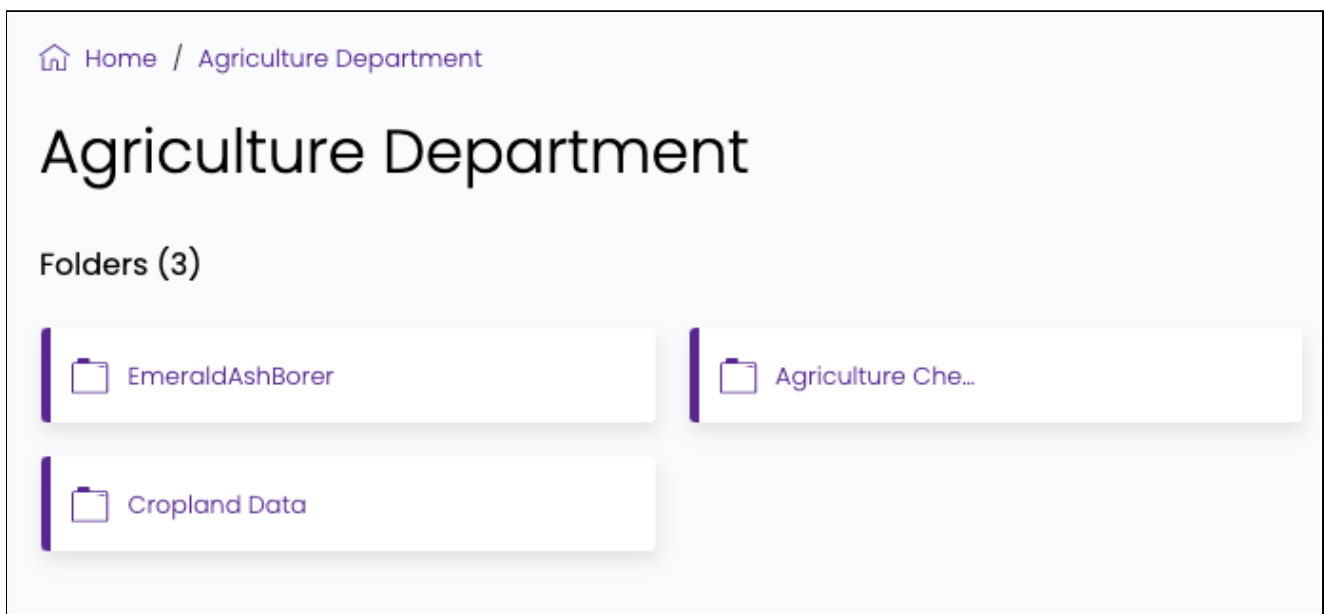


Figure 4.9: The Agriculture Department top-level folder in the Preservica Starter administration interface. It contains three Collection folders.

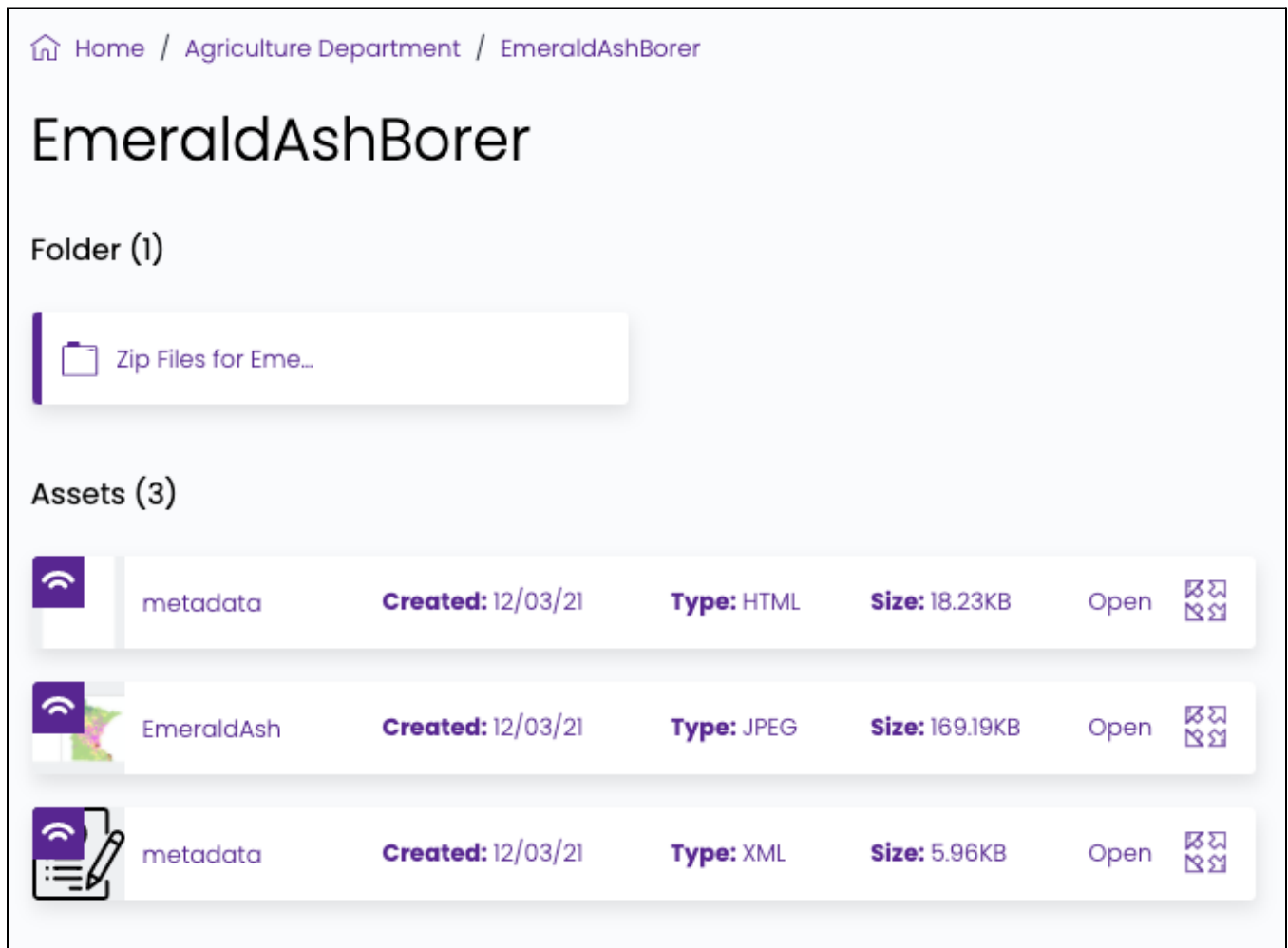


Figure 4.10: The second level folder in the Preservica Starter administration interface. It contains one folder and three files.

Appendix 5. Metadata Crosswalk

This chart maps the Minnesota Geospatial Metadata Guidelines (MGMG) to Dublin Core, the metadata schema used by most discovery interfaces. Note: This is not all-inclusive, as each application may have custom fields not represented here.

Bold = mandatory; ***Bold Italics*** = Mandatory if applicable; *Italics* = Desirable; Plain = Optional

MGMG	Dublin Core
1.1-8.4 Title	dcterms:title
1.1-8.8.2 Publisher	dcterms:publisher
1.2.1 Abstract	dcterms:abstract
<i>1.3 - 9.1.1 Time Period of Content Date</i>	dcterms:temporal
<i>Part of 1.2.3 Spatial Extent of Data</i>	dcterms:spatial
<i>1.3.1 Currentness Reference</i>	dcterms:temporal
1.5.1.1 West Bounding Coordinate	dcterms:spatial
1.5.1.2 East Bounding Coordinate	dcterms:spatial
1.5.1.3 North Bounding Coordinate	dcterms:spatial
1.5.1.3 South Bounding Coordinate	dcterms:spatial
1.6.1 Theme Keywords	dcterms:subject
1.6.1.1 Theme Keyword Thesaurus	dcterms:subject
1.6.2 Place Keywords	dcterms:subject
1.7 Access Constraints	dcterms:rights
1.8 Use Constraints	dcterms:rights
2.4.1.1 Horizontal Positional Accuracy	dcterms:spatial
<i>2.4.2.1 Vertical Positional Accuracy</i>	dcterms:spatial
<i>5.2.1 Entity and Attribute Overview</i>	dcterms:description
<i>5.2.2 Entity and Attribute Detailed Citation</i>	dcterms:description
<i>6.3 Distribution Liability</i>	dcterms:rights
No equivalent	dcterms:identifier