MINNESOTA GEOSPATIAL ADVISORY COUNCIL

Minnesota Geospatial Advisory Council - Archiving Imagery Workgroup
Final Report

About the Archiving Imagery Workgroup

Background

In the fall of 2018, the GAC voted to approve the creation of an Archiving Workgroup to define the policies, best practices, and procedures for archiving geospatial data in Minnesota so that valuable geospatial data can be preserved and available for future use. The latest iteration of this project is the Archiving Imagery Workgroup, active June - December 2022, tasked with investigating best practices for archiving aerial imagery data and continuing to build on the work of the Archiving Pilot Workgroup, Archiving Implementation Workgroup, and Archiving Workgroup.

Objectives

Part 1: Research

- Research the historical and current formats of imagery data
- Identify any major challenges with archiving imagery data, such as space considerations due to the size of the files
- Evaluate existing practices and potential strategies for transferring, storing, transforming, documenting, and archiving imagery data

Part 2: Community outreach and survey

- Survey Minnesota's geospatial community on their imagery and existing practices
- Promote the activities and findings of the archiving workgroup
- Assess the monetary investment involved in the original creation of Minnesota imagery

Deliverables

- Final Report of our findings
- A poster (or presentation) at the 2022 MN GIS/LIS Conference

Workgroup Members

Group Membership:

- Melinda Kernik - University of Minnesota Libraries (Vice Chair)
- Carol Kussmann - University of Minnesota Libraries
- Karen Majewicz - University of Minnesota Libraries (Chair)

Resource People (consulted as needed):

- Jennifer Corcoran - DNR Remote Sensing Program Consultant
- Brent Lund - MNIT / MnGeo
- Ryan Mattke - University of Minnesota Libraries
- Matt McGuire - Met Council & Image Service Sustainability Committee

Summary of Findings and Challenges

- Historical aerial imagery is highly valued and used by the community.
- Multiple agencies are spending hundreds of thousands of dollars per year collecting new imagery.
- Digital image files are stored inconsistently in a variety of file formats and storage locations and are at risk of loss.
- Digital image files take up considerable server space.
    - As the resolution of collected imagery increases, so does the size of data files; this could pose a challenge for the archive over time.
- Digital image files are often unknowingly stored on duplicate drives across the state.
    - Imagery in multiple file formats or with related versions may further complicate identification of duplicates and decisions about what to preserve.
- There is no clear workflow for retroactively archiving images from geospatial image services.
- The archive will need a plan for dealing with proprietary and legacy file formats.
- Ensuring inclusion of metadata and indexes (if relevant) with imagery files is critical for maintaining the value of the imagery and allowing future reuse.

Introduction

The history of aerial photography began in the 19th century with large format cameras attached to hot air balloons, kites, pigeons, or even rockets, and the brief flights yielded a single photograph. The technology of the early 20th century, including airplanes and widely accessible film cameras, made long-range aerial photography more practical. By the 1920s and 1930s, commercial entities began offering flights for hire that would take a series of film photographs (air photos) to cover entire regions. These air photos were in the form of stereographs: each photo slightly overlaps the next and can be lined up to create panoramic views.

Minnesota's earliest air photos come from this period and their existence extends into the later 20th century. Many of these air photos are currently stored at state agencies, university repositories, or the Minnesota Historical Society as physical objects: film rolls, slides, or physical prints. Scanned copies of hundreds of thousands of these images can be accessed through the University of Minnesota's MHAPO[1] or the DNR's Landview[2] applications. These map-based search portals use centerpoints or grids associated with the photographs. Although they have spatial information connected with them, these images are not spatial raster files (they have not been processed for use in a GIS).

Judging from website analytics, air photos from this time period are highly valued by the public. In 2022, MHAPO received an average of 5,300 site visitors and 17,200 image downloads per month.[3] Although Minnesota's early air photos could benefit from being added to an archive program, the images from this time period are not currently at a high risk of loss.

By the 1990s, digital cameras replaced film for creating aerial photography. Although they are more recent, the preservation status of the born-digital images is less clear, as storing and preserving digital media is more complex than storing boxes of physical media and paper maps. Additionally, born-digital aerial imagery typically undergoes postprocessing in the form of georeferencing and orthorectification.[4] Each image may be transformed into a spatial raster dataset, such as a GeoTIFF, and ultimately combined with adjacent images to transform into a geospatial image service for public access. Along the way, the raw data and the GeoTIFF might be stored on unreliable drives or even lost. This era of aerial imagery will serve as this report's main subject of investigation, as it is at the highest risk of being lost.

---

[1] MHapo, https://apps.lib.umn.edu/mhapo/

[2] https://www.dnr.state.mn.us/maps/landview/index.html

[3] Mattke, Ryan; personal communication, December 2, 2022

[4] Due to the curvature of the earth and distortions at the edges of camera lenses, aerial imagery needs to be adjusted to be used in a GIS application.

Part 1: Research

References

Our first task was gathering reference material for archiving geospatial imagery. The majority of cited resources related to geospatial data archiving continue to be primarily from the 2008-2012 era, when the National Digital Information Infrastructure and Preservation Program (NDIIPP)[5] funded several geospatial archiving projects, such as the Geospatial Multistate Archive and Preservation Partnership (GeoMAPP) and the National Geospatial Digital Archive (NGDA). See the Archiving Strategy Final Report (2019) for more details on individual projects.[6] Despite the preponderance of older documents, we were able to discover more recent scholarship that can serve as additional references for geospatial archiving in general and aerial imagery in particular (see Appendix 1).

Geospatial Imagery File Formats

In addition to reference materials, we also examined a selection of aerial imagery inventories at the University of Minnesota John R. Borchert Library collection and from MnGeo. These activities aided us in producing an extensive (though not entirely comprehensive) list of geospatial imagery file formats.

Some of the most common file formats for imagery include:

File Format	Usage
GeoTIFF and TIFF (.tif)	the most common format for geospatial imagery; recommended format for archiving images (Library of Congress)
JP2(.jp2) and JPEG (.jpg)	less commonly used for archiving because of its compressed format can be useful for large files and mosaics
MrSID (.sid)	commonly used format for orthoimages in need of compression

Considerations when archiving imagery file formats:

Avoiding proprietary formats where possible

Most geospatial file formats, even when they have no formally published specification, can be opened within open source software. For example, even though .sid and .gdb formats are proprietary, they can be read in QGIS. While open source file formats are preferable for long-term access, proprietary

[5] http://www.digitalpreservation.gov/

[6] https://www.mngeo.state.mn.us/workgroup/archiving/Archiving_Strategy_Report.pdf

formats could be acceptable as long as they can be read by Open Geospatial Consortium (OGC) tools (such as QGIS and gdal).

Contending with legacy raster file formats

There are a variety of legacy imagery formats, such as ESRI GridFloat Output file (.flt) and ESRI ArcInfo Grid / Coverage file (.adf). At the moment these can still be opened in major geospatial software, but with increasing limitations on what programs can still interpret them. For example, coverage files will not be supported in ArcGIS Pro. Legacy file formats may need to be converted for long-term access to data.

Ensuring inclusion of geographic location with imagery files

Sometimes information about the geographic location of the image is contained within the primary file format (GeoTIFF) or in a related file (.tfw). Other times it may be stored in a separate index, such as in a shapefile or geodatabase. This information is absolutely crucial to being able to interpret and make use of the files. Effort should be made to ensure that location information is present with or within the files.

Part 2: Community Outreach and Survey

The second part of our objectives was to engage with Minnesota's geospatial community. We endeavored to learn about what kind of imagery various organizations have, how they store it, and to gauge their interest in archiving it. This work took two forms: direct interviews with selected state agencies that collect or produce a large amount of imagery and an online survey distributed to the wider community that reached municipalities, counties, reservations, universities, and the nonprofit and private sectors.

We only considered digital aerial photography, including scanned photos, georeferenced / georectified image files, or image services. Examples of file formats are TIFF, JPEG, SID, GRID, and their associated spatial info files. We did not include information about satellite imagery or LiDAR files in the survey.

Interviews with state agencies

1. MnGeo

Summary of an interview with Brent Lund, GIS Architect, Geospatial Information Office, on September 13, 2022

Overview

MnGeo is the main organization in Minnesota that coordinates and collects aerial imagery from a variety of sources, including federal, state, and county agencies. MnGeo is a prominent contributor to aerial imagery access through its Geospatial Image Service, which provides Web Map Services (WMS) for dozens of Minnesota air photos, hillshades, and scanned topographic maps. The raw image data for these services comes from various sources, including the USGS, Farm Service Agency, U.S. Department of Agriculture, Minnesota Department of Transportation, Minnesota Department of Natural Resources, and many counties. See the list of MnGeo WMS Data Layers for a full inventory of the service and each layer's originator.

As time goes on, MnGeo faces the challenge of maintaining this growing collection of services. In response, the GAC charged the Image Service Sustainability Committee to regularly assess and prioritize the availability of current layers. This work involves identifying layers that should be retired from the Image Service, as each layer requires substantial server space and memory allocations. However, there is no clear workflow for retroactively deconstructing the mosaicked web services back into a format suitable for archiving, such as a group of orthorectified TIFFs. Alternatively, the original images could be archived, but this requires finding and processing them; they may be on a MnGeo hard drive, they may have been returned to the source organization, or their location may be unknown.

2. Metropolitan Council

Summary of an asynchronous interview with Tanya Mayer, Senior GIS Coordinator, Metropolitan Council in August 2022

Overview

The Metropolitan Council coordinates imagery collection in the 7-county Twin Cities Metropolitan area to analyze changes in land use designation and to aid their regional planning. The Council follows a contract process to hire an aerial imagery firm to carry out this work. Once collected, the imagery is sent to MnGeo to be included in their Geospatial Image Service.

The Council has been collecting aerial imagery for many decades. The original photographic prints from 1970-1996 are stored at the University of Minnesota or the Minnesota Historical Society.

How frequently is imagery collected?

Every five years

How much has your organization spent on collecting imagery?

In the most recent year of collection (2020), the Council spent $156,680 on aerial imagery collection.

How do you fulfill requests from the public?

When the Council receives public requests for imagery, they are referred to MnGeo.

3. MnDNR Resource Assessment

Summary of an interview with Jennifer Corcoran, DNR Remote Sensing Program Consultant on August 10, 2022

Overview

Resource Assessment (RA) is part of the DNR's Forestry Division. It acts as an "enterprise" group by providing various public consulting services and generating revenue for the agency. It is also the group that coordinates the statewide Color-infrared (CIR) photography. This imagery is taken in the autumn season to provide the fall color of the tree canopy. Since the time period of collection is limited, RA uses a ten-year cycle, with a section of the state covered each autumn. It is produced annually by external vendors, who need to bid on the project. This imagery is stored on DNR servers and mirrored to a non-public MN.IT server. The original prints are in multiple boxes and locations.

The RA group also has photography equipment that can be used for project-based collections, such as the Wetland Monitoring Program or tracking the effects of Oak Wilt. This imagery is not fully inventoried and may be stored in various drives around the agency.

How frequently is imagery collected?

Annually

How much has your organization spent on collecting imagery?

The cost of producing this imagery varies depending upon the year, but likely averages to several hundred thousand dollars annually.

How do you fulfill requests from the public?

Users generally use Landview to access the imagery and only contact the DNR if something is missing or broken.

Online survey

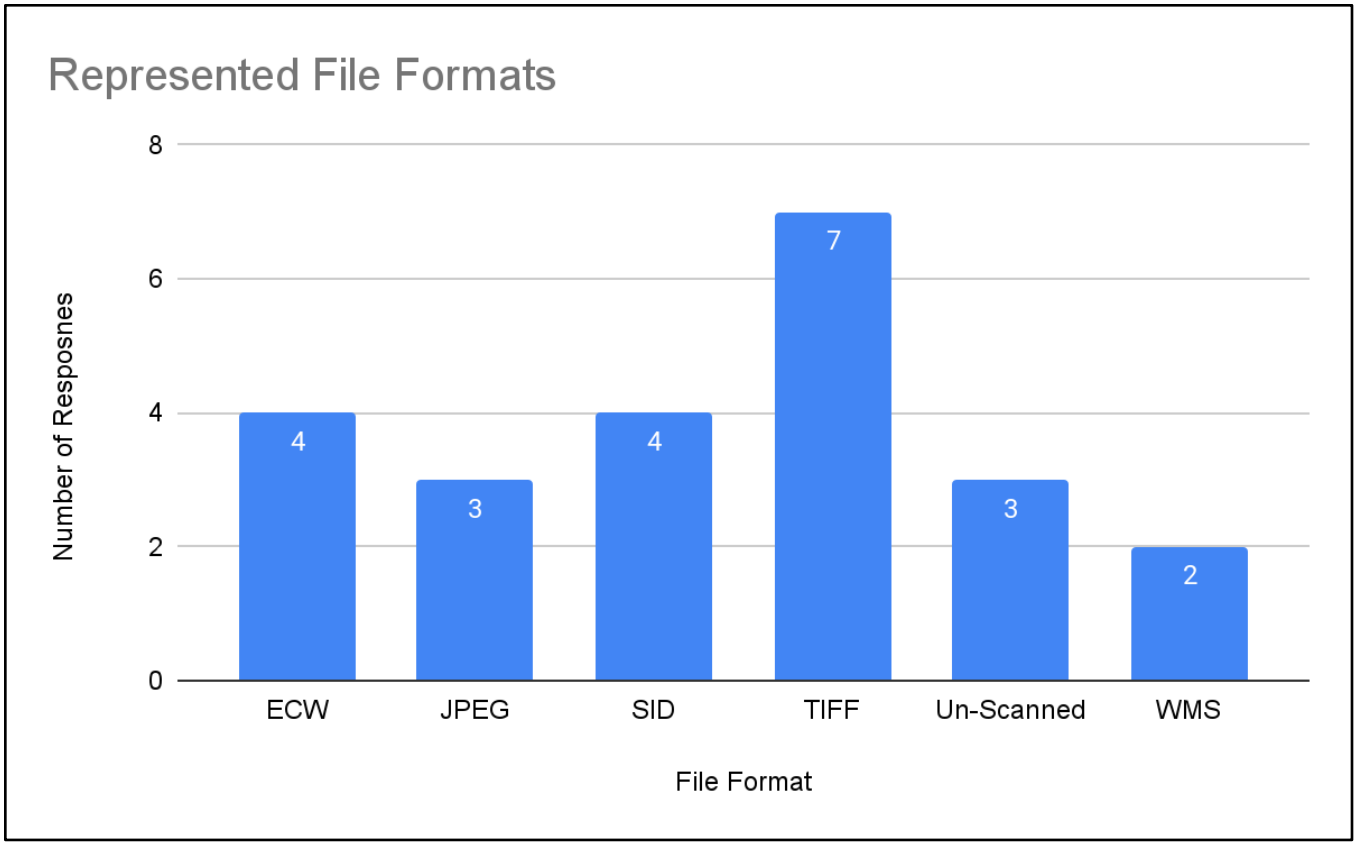
To reach a wider audience, we created a survey to gather an inventory of the types of imagery resources across the state that would be suitable for archiving. This survey was announced with a poster presentation at the Minnesota GIS/LIS 2022 Annual Conference and sent out by email to the geospatial community via the MN GIS/LIS Consortium E-Announcement and the Minnesota Geospatial Information Office newsletters. Responses were collected from November 2-28, 2022. Twelve responses were received in this short time period.

The survey asked about organization type, coverage of spatial data, types of imagery, file formats, date of collection, storage location, value, and frequency of requests for data.

The 12 responses were from a variety of organization types including county (6), city (2), university (1), private (1), tribal (1), and state department (1). Each organization focused on their general jurisdictions

with the exception of the university, which covered the southern part of the state, and the private organization which has materials from both Minnesota and other states.

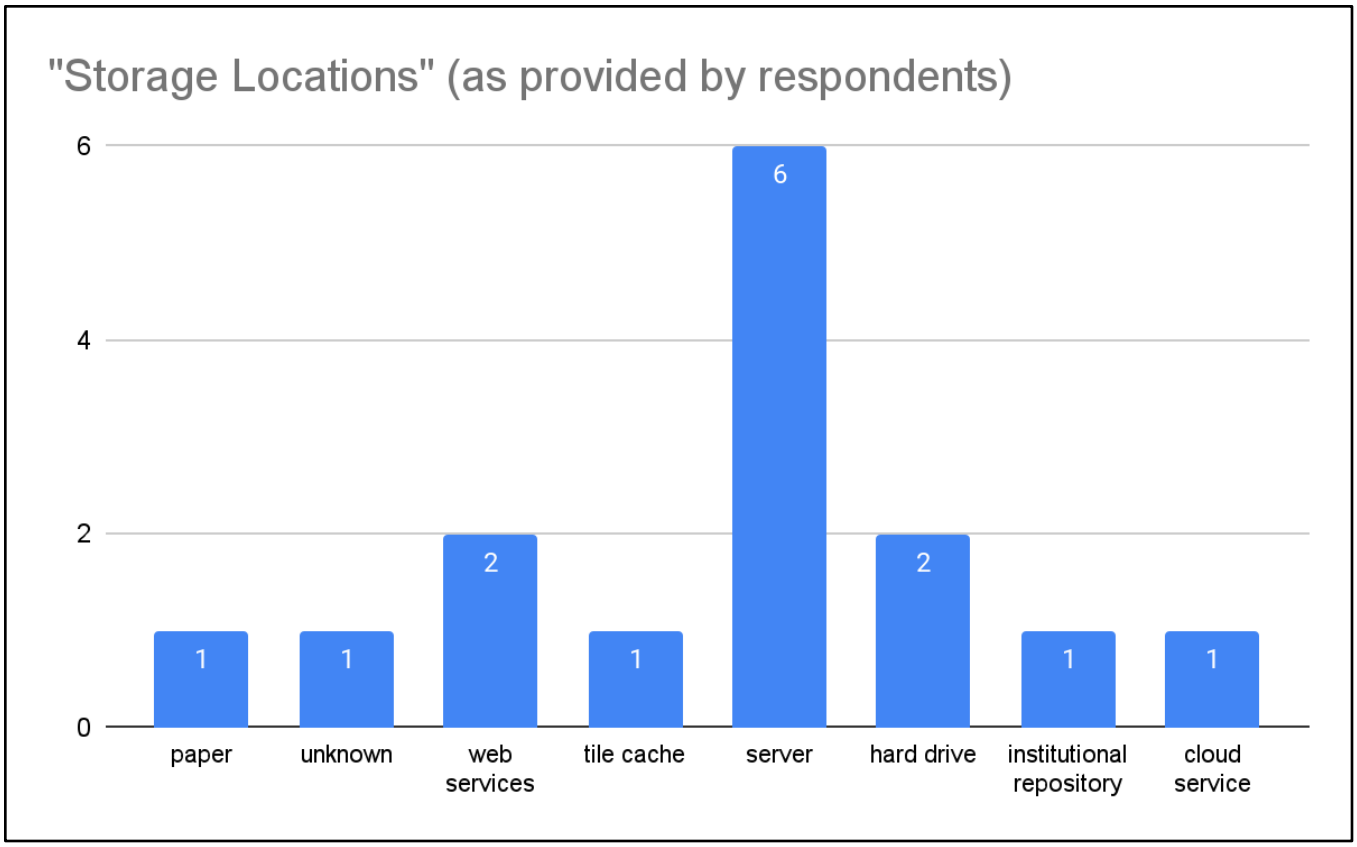
Most respondents have data in multiple formats. The formats below were provided by the respondents, no format options were provided in the survey. TIFF files were the most common, which is what our research showed as well.



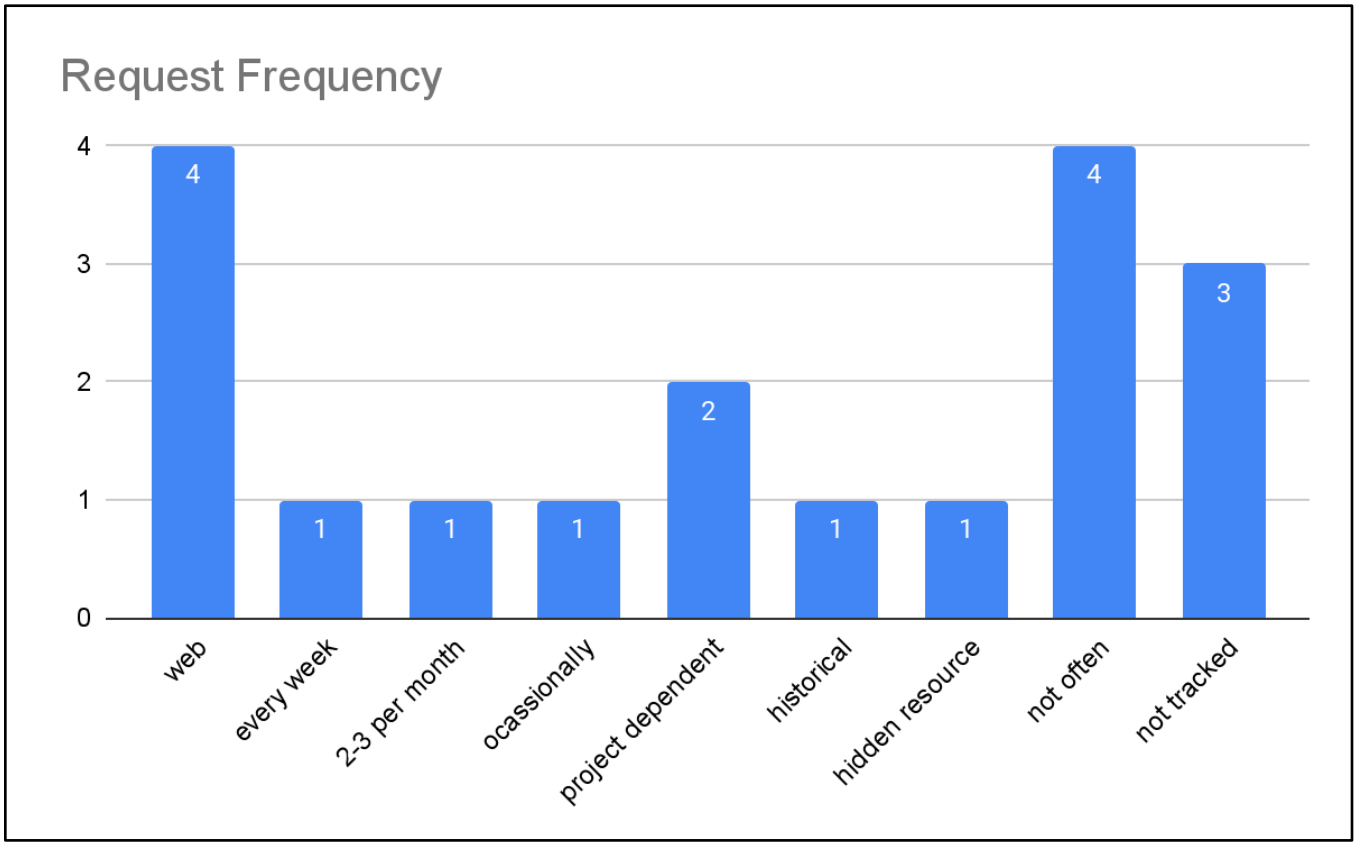
The respondents were also asked to describe the 'type of imagery' they had. Some of the responses provided by the respondents in their own words included aerial images (5), ortho imagery (4), georeferenced (3), black and white aerial footage (2), tape or film (2), scanned images (1), digital stereo imagery (1), oblique imagery (1), raw files (1), imagery services (1), and digital mosaic (1).

The collection date for imagery respondents are interested in archiving range from 1928-2022. Two respondents included only one year's worth of data from the 1960s. Six respondents seem to collect data every year, mostly from the 2000s but also earlier. Three respondents seem to collect data every few years on a 2-3 year cycle. Four respondents have data from prior to 1940.

Most of the data respondents are holding reside on servers. Materials stored on hard drives are probably the most at risk. Other responses are shown in the chart below.



Tracking how often data is requested or used is one way to assess the value of an archive. Many of the respondents provide access to the data via web services, however, which means that the data is constantly available but access may or may not be tracked. Three respondents specifically stated that they do not track how often data is requested. Four stated data is not requested often, while two suggested requests are project driven, so there may be a lot of requests in a short period of time and then not many after the project is complete. Other respondent provided time frames are listed in the chart below.

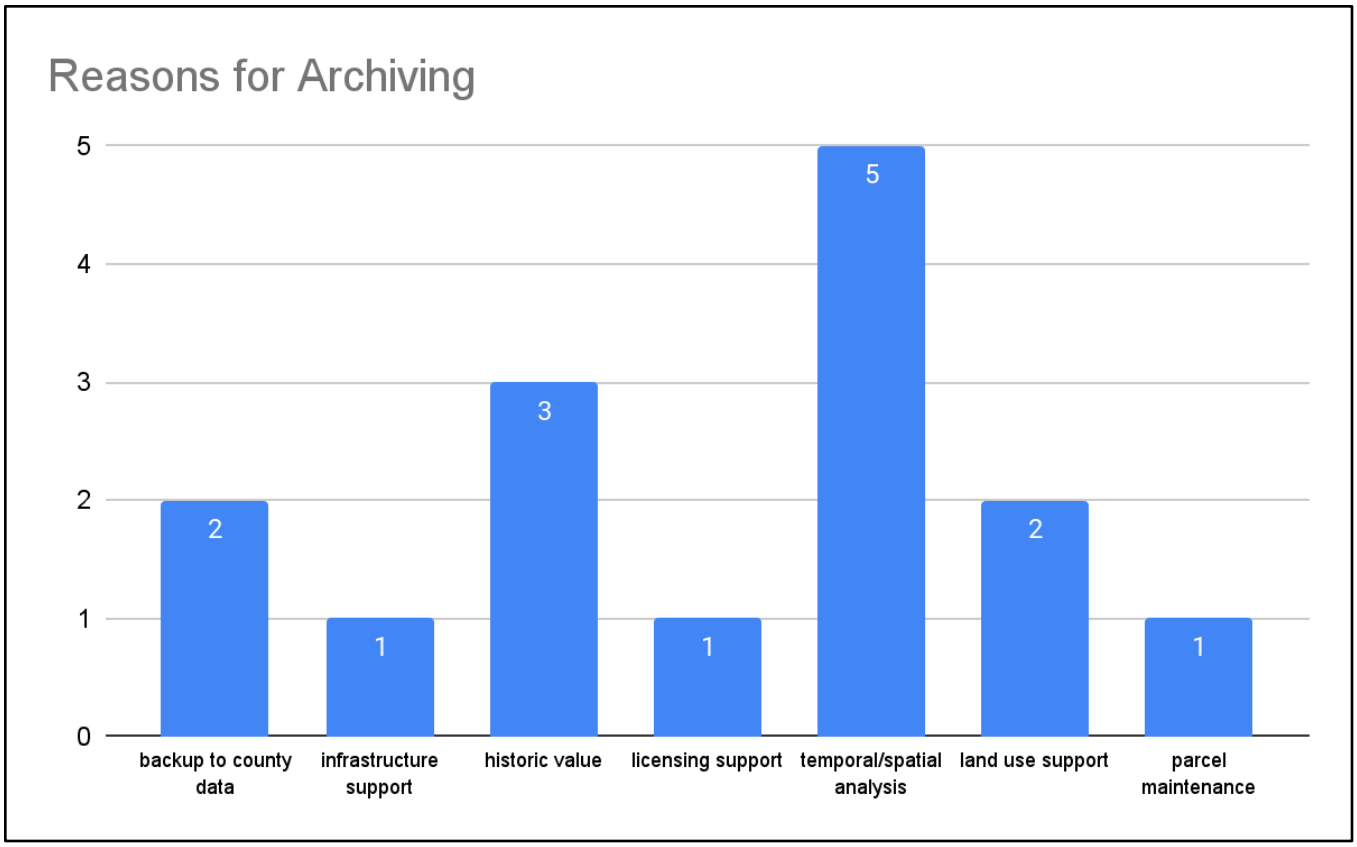


The number of times data is requested is not the only way to assess value. Previous archiving work groups have gathered stories showing the value and many uses of historical geospatial data, including to identify and preserve archeological sites, resolve issues with contested survey boundaries, and track lake level changes.

When asked about the value of preserving files over the long term, survey respondents provided additional examples demonstrating use of historical data as well as comments about the overall value of historical imagery. A few specific examples are listed below:

- filling in gaps in imagery that is currently available through the DNR and other sources
- monitoring changes in infrastructure over time to assist with city inspections
- identifying pre-development land use and locating possible unsealed wells
- delineating wetlands and conducting environmental site assessments

In trying to summarize the responses into a chart the following was found.



Respondents also stated the value of archiving this data may be invaluable (1), debatable (1), or a novelty (1).

The last question in the survey asked respondents if they had questions for the survey creators. A few were forward thinking. One respondent asked about how the archived materials would be made available, another interested in LIDAR data, and another about other formats that could be used for presenting data.

Analysis and Conclusion

Anecdotally, we learned state agencies who produce aerial imagery are spending up to several hundred thousand dollars per year on imagery collection. This represents a significant investment, largely funded by public budgets. Though the most recent and highest resolution imagery is viewed most often, older imagery continues to be a valuable asset for land use, environmental, and infrastructure planning. It is important not to squander the money and effort spent collecting this information.

While many agencies rely on web services to store and deliver aerial imagery to the public, the original prints or digital files that were collected to make these services are not being stored systematically. It is important to manage the digital files used to create these resources so they can be accessed over time as technology and discovery platforms change.

MNGeo backs-up decades of imagery from both state and county entities and currently has about 38TB of imagery data files. Our survey has made clear, though, that there are many other smaller sets of imagery held by city, county, tribal, and other government agencies not included in this total. These smaller sets of images are the most at risk and could benefit the most from preservation efforts.

As our inventory efforts revealed, one of the challenges working with many agencies will be the potential complexities of handling duplicate imagery. For example, identical files may be stored by multiple agencies or multiple versions of the same imagery may be held in different formats. Other challenges in archiving imagery data include being able to manage the overall size of an archive, the varying file formats, and being able to make the files accessible to users in a way that is useful to the users. Providing access to archived files may look different than the way existing services provide access to imagery as the focus is on long term availability and preservation rather than immediate use.

Creating a central location in which to store older imagery is a way to protect past investments. Workflows can be developed to streamline the process. The existing imagery data files from MNGeo will provide a good baseline for testing processes as the data is in a variety of formats and covers both state and county areas. Smaller sets of data held by other city, county, tribal and government agencies can then be examined for unique imagery to be included in an archive. These smaller agencies may benefit the most from the archive as they often don't have the infrastructure to store multiple years or versions of the same datasets.

Appendix 1: Recent scholarship on archiving aerial imagery

Carol Patterson McAuliffe, Kathryn Lage & Ryan Mattke (2017) Access to Online Historical Aerial Photography Collections: Past Practice, Present State, and Future Opportunities, Journal of Map & Geography Libraries, 13:2, 198-221, DOI: 10.1080/15420353.2017.1334252

Bruce Godfrey (2019) Opportunities to Enhance Discovery, Explorability, and Access for Digital Aerial Imagery Collections, Journal of Map & Geography Libraries, 15:1, 28-44, DOI: 10.1080/15420353.2019.1661933

Bruce Godfrey (2021) Image Files, Web Services, Web Applications, and Partnerships: Two Decades of Managing Digital Georeferenced Aerial Imagery Collections of Idaho, Journal of Map & Geography Libraries, 17:1, 39-57, DOI: 10.1080/15420353.2022.2041529

Artefactual Systems and the Digital Preservation Coalition (2021). Digital Preservation Coalition. Preserving GIS: Data Types Series http://doi.org/10.7207/twgn21-16