ViSTA Testing: Gemini Metadata Quality

Initial Report, February 2025

Drew Facklam, Metadata and Digital Projects Supervisor

Overview

Commented [SS1]: A brief paragraph, or just a sentence or two, would help to orient the reader to the purpose of the document

Testing description

The goal of this report is to describe the initial testing of the ViSTA tool for Google Gemini, a Python-based API created by AI the Digital Collections co-ops, Shoumik Majumdar and Rahul DMello Kamath. ViSTA for Gemini uses the fronts (images) and backs (text) of photographs to generate descriptive metadata to support discovery and access in the DRS. This initial testing is meant to be scalable, repeatable, and actionable—providing guidance that will support further testing, prompt

re-engineering, quality control decision-making, and guide next steps in integrating this tool into general metadata workflows.

The testing set is a group of 100 photographs from the Boston Globe Library collection (M214). This collection, which already contains over 75,000 undescribed photographs and will continue to grow, is one of the main collections intended to employ the ViSTA Gemini tool, and therefore quality of metadata output for these images is of paramount importance.

This round of testing focuses on establishing baseline expectations for completeness, accuracy, and consistency of metadata output. Additional considerations such as descriptive bias will be covered in a future report which will use a larger and more diverse dataset and involve additional library staff to review metadata output.

Testing set
100 image fronts and backs from the Boston Globe collection

Archival Folder titles for photographs in testing set
Massachusetts--Islands--Plum Island—Erosion
Massachusetts--Islands--Plum Island--AA

Elements tested
- Title
- Abstract
- Photographer
- Date[s]
- Raw transcription of back of photograph

Testing parameters
- Completeness: the comprehensiveness of the output for each element
- Accuracy: the degree to which the element value or content is correct
- Consistency: regularity of the accepted use of each element

Raw Gemini output
Gemini_output_2025-01-15_16-05-46.csv

Results

Title and abstract

Completeness

Measured by comparing essential keyword access. 3-6 keywords for each image were selected. These terms represented words that should be present in the title and/or abstract in order to support keyword search discovery. Then, each photograph's Gemini-generated title and abstract were reviewed for the presence of each keyword.

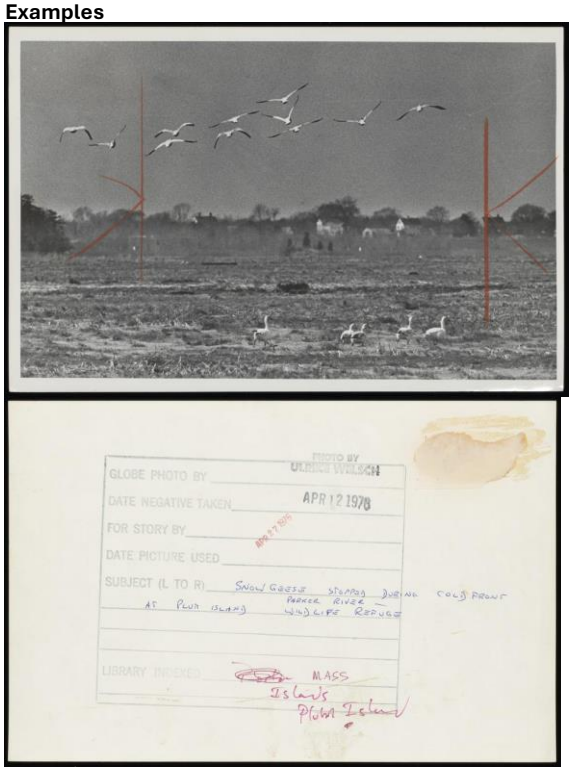
- Total terms generated by cataloger: 441
- Total terms missing from Gemini descriptions: 50

        - Percentage missing terms: 11.33%
    - Number of records with 100% of the terms: 60 (of 100)

Types of missing terms

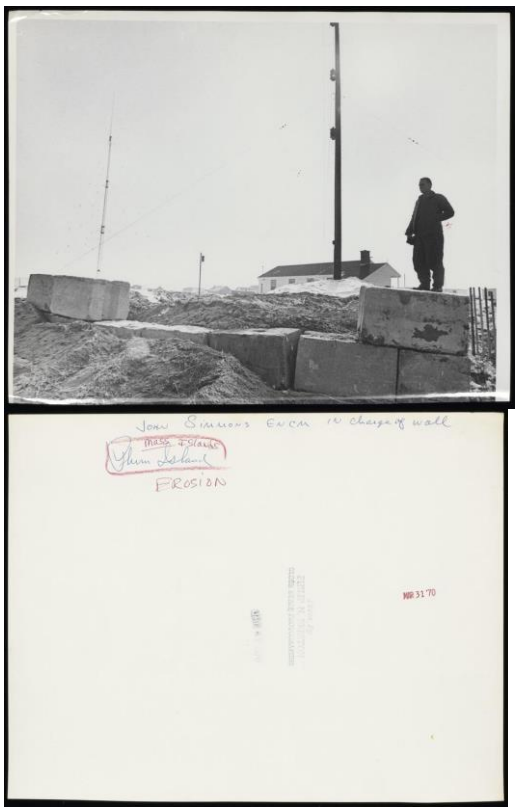
- **Terms missed because of transcription errors**
  For example, if the transcription misspelled or missed a reference to an archival folder title written on the back, then keywords based on the folder title would be either missing from or incorrect in the title/abstract.
- **Seasons**
  Times of year and their associated seasons and weather are commonly searched keywords when using image archives, but the vision tool did not provide/rarely provided this detail.
- **Jargon vs. general terms**
  Often there is a slight difference between the specificity of the term used and a more general term that could also be mentioned to support discovery. For example, many of the images depict birds in flight, and the titles/abstracts often repeat more specific terms such as geese, flock, avian life.

Examples



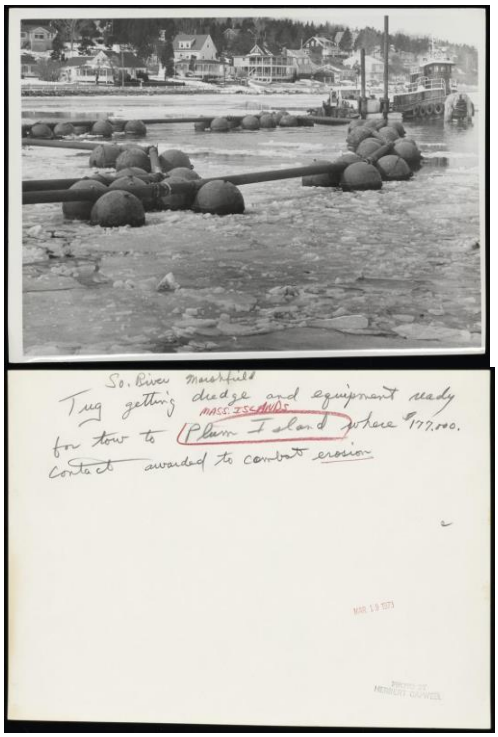
Filename: M214_1015936037_0001_0057

Title: Snow geese on the ground and in flight at the Parker River National Wildlife Refuge on Plum Island, Massachusetts
Abstract: This black and white photograph depicts snow geese at the Parker River National Wildlife Refuge on Plum Island, Massachusetts. A flock of geese flies across the top portion of the frame, while a smaller group walks on the ground in the foreground. The background features a flat, sparsely vegetated landscape with buildings visible in the distance. The image is marked with red lines, likely for cropping or layout purposes. The photograph documents the presence of snow geese on Plum Island, likely during a cold front.
Keywords: Plum Island, birds, geese, wildlife refuge, landscape
Number of missing keywords: 1



Filename: M214_1015936037_0002_0004
Title: John Simmons, ENCM in charge of the seawall, stands atop concrete blocks on eroding Churn Island, part of the Mass Islands
Abstract: This photograph, taken by Philip N. Preston of the *Boston Globe* on March 31, 1970, depicts John Simmons, identified as ENCM (likely an abbreviation for Engineer in Charge of Maintenance), standing on a section of seawall on Churn Island, part of the Mass Islands. The image documents the erosion occurring on the island, with large concrete blocks forming a barrier

against the sea. The composition features Simmons as the focal point, positioned against the backdrop of a utility pole and a small building, with additional utility poles and scattered debris in the foreground. The stark landscape and the presence of the seawall emphasize the impact of erosion on the island.
Keywords: Plum Island, blocks, sand, erosion
Number of missing keywords: 2



Filename: M214_1015936037_0002_0017
Title: A tugboat and dredge prepare to tow dredging equipment to Plum Island, Massachusetts, to combat erosion.
Abstract: This photograph depicts dredging equipment, including large spherical buoys connected by pipes, being prepared for transport by a tugboat in the icy waters of the South River in Marshfield, Massachusetts. The equipment is destined for Plum Island to combat erosion, a project with a $177,500 contract. Houses line the shore in the background, providing context to the riverfront location. The foreground is dominated by the dredging equipment and ice chunks, while the tugboat and a smaller dredge with an excavator are visible in the mid-ground. The background features the residential buildings along the shoreline.
Keywords: Plum Island, dredge, erosion, ice, winter
Number of missing keywords: 1

Accuracy

Measured by assigning 3 different grades of accuracy to each title and abstract:

- 3: 100% accurate, displaying a clear understanding of the subject matter of the photograph (both the image and additional context provided on the back).
- 2: Contains some small misunderstanding of the image or the transcription, but it does not essentially change the interpretation of the central aboutness of the photograph.
- 1: Contains some form of misunderstanding of the image or coverage of the transcription that leads to an inaccurate description which changes the essential subject matter of the photograph.

Grade	Number of titles	Number of abstracts	Number effected by transcription error
3	87	72	0
2	7	15	8
1	6	13	17

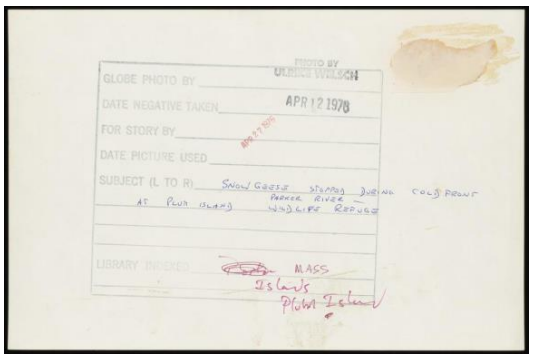
***66 records scored a 3 for both title and abstract.

Types of inaccuracies

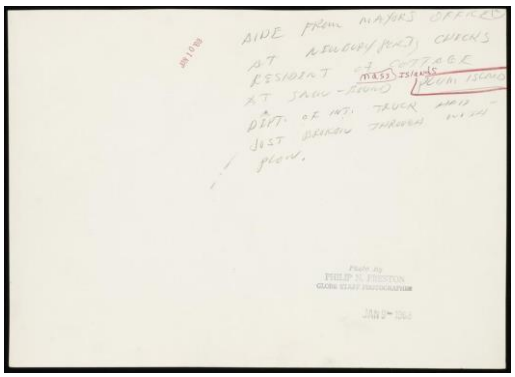
- **Doesn't affect overall meaning (grade of 2)**
  - Miscounts number of objects in an image (e.g. if there are 18 birds flying in an image and the abstract mentions 12)
  - Use of awkward or unused language describe something (e.g. calling a pond a "water pool")
  - Small misunderstanding/interpretation of the image or the transcription (e.g. describes a truck as a snowplow, refers to a small dinghy as abandoned or derelict when it just looks slightly worn)
- **Does affect overall meaning (grade of 1)**
  - Use of incorrect proper nouns based on transcription error (e.g. if the transcription named the location of the image by an incorrect name)
  - Incorrectly labels the figures in an image (e.g. if a small animal is described as a child)

Examples

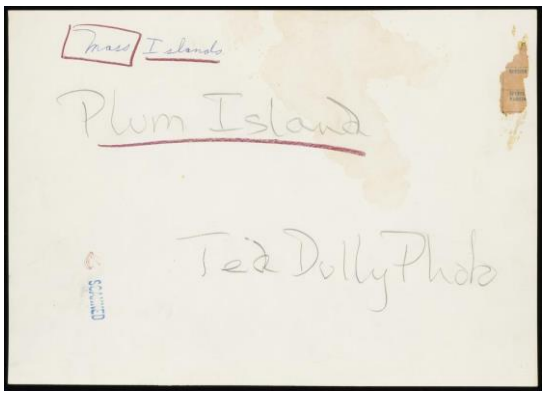
Commented [SS2]: Thinking ahead, if we end up creating a statement (like the Statement on Inclusive Description, separate from item-level statements) about our use of AI to create some metadata, this is a good example of what users should look out for



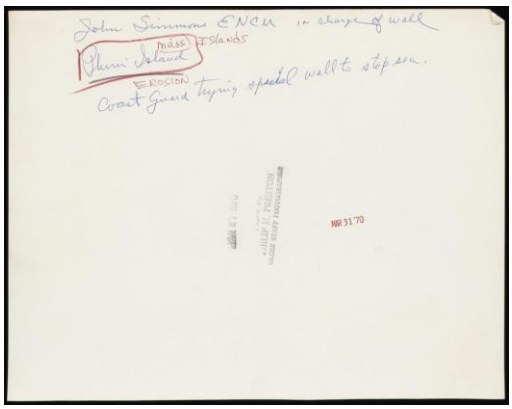
Filename: M214_1015936037_0001_0057
Title: Snow geese on the ground and in flight at the Parker River National Wildlife Refuge on Plum Island, Massachusetts. **(GRADE: 3)**
Abstract: This black and white photograph depicts snow geese at the Parker River National Wildlife Refuge on Plum Island, Massachusetts. A flock of geese flies across the top portion of the frame, while a smaller group walks on the ground in the foreground. The background features a flat, sparsely vegetated landscape with buildings visible in the distance. The image is marked with red lines, likely for cropping or layout purposes. The photograph documents the presence of snow geese on Plum Island, likely during a cold front. **(GRADE: 3)**
Accuracy: 100% accurate.



Filename: M214_1015936037_0001_0053
Title: An aide from the mayor's office checks on a resident at their snow-bound cottage on Plum Island, Massachusetts, as a Department of Interior truck plows through. **(GRADE: 3)**
Abstract: This photograph, taken by Philip N. Preston of the Boston Globe on January 10, 1968, depicts the aftermath of a snowstorm on Plum Island, Massachusetts. A Department of the Interior truck, having just plowed through the deep snow, is prominent in the foreground. An aide from the Newburyport mayor's office checks on a resident of a snow-bound cottage. The scene shows snow-covered houses and a sign partially buried in the snow. The composition emphasizes the truck and the deep snow, highlighting the difficulty of access and the isolation of the residents. **(GRADE: 2)**
Accuracy: Title is 100% accurate. Abstract explains truck in image as a snowplow (which is not pictured), and discusses a sign buried in snow which is fully visible above the snowline.



Filename: M214_1015936037_0001_0061
Title: An adult and child walk along a beach on Plum Island, Massachusetts, as waves crash against the shore. **(GRADE: 1)**
Abstract: This black and white photograph, taken by Ted Dolly, depicts a windswept scene on Plum Island, Massachusetts. A lone figure in a hat walks along the beach as choppy waves crash against the shore. A small dog or other animal sits near the water's edge. The rough sea dominates the composition, filling most of the frame, while the dark silhouette of the land in the distance provides a stark contrast to the bright, churning water. The figures of the man and animal provide a sense of scale and emphasize the vastness of the seascape. **(GRADE: 3)**
Accuracy: Inaccurate title, labels a dog as a child. Abstract is 100% accurate.



Filename: M214_1015936037_0002_0010
Title: John Simmons, ENCM, observes a concrete seawall under construction by the Coast Guard to combat erosion on Thurni Island, Massachusetts. **(GRADE: 1)**
Abstract: This photograph depicts John Simmons, an Engineering Construction Mechanic (ENCM) with the Coast Guard, observing a concrete block wall being constructed on Thurni Island, part of the Massachusetts Islands. The Coast Guard implemented this wall as a measure against coastal erosion. The foreground features Simmons standing on sandy terrain, with the partially buried wall extending towards the background. In the distance, buildings and a lighthouse are visible, suggesting a nearby settlement. The image documents the Coast Guard's efforts to combat erosion in this island community. **(GRADE: 1)**
Accuracy: Both the title and abstract label Plum Island as “Thurni Island” due to a transcription error.

Consistency

Both title and abstracts show generally consistent format and descriptive elements. But there are some small inconsistencies which could be improved upon. Consistency will often affect accuracy and completeness so several of these issues are discussed previously in this document. However, consistency can reveal patterns which make general cataloging patterns, and therefore prompt re-engineering, clearer.

Consistent content

- Use of transcription to add meaning to titles and abstracts
  - Addition of locations and names
  - Addition of meaning behind documentary image
- Inclusion of date and photographer
- Information about distinctive figurative focus in images (persons, animals, buildings)
- Description of type of landscape
- Inclusion of notes on fore, middle, and background
- Artistic descriptions and notes on the mood displayed in the images

Inconsistent content

- Mention of weather and season

- Acknowledgement of source of description ("based on the handwritten notes on the back of the photograph…")
- Mention of editorial crop marks
- Mention and understanding of text captured in the image

Analysis

Almost 90% completeness likely rivals the variability of results we would find between two average catalogers (which can be confirmed in future larger-scale testing). Although additional prompt engineering could further correct the keyword variation seen in this testing, the current term inclusion is strong enough to be considered an adequate level of completeness. Furthermore, the addition of LCSH terms would be a generally accepted way to improve keyword access, and additional research on other workflows that could support this inclusion are currently being discussed. What has become immediately apparent is that the titles and abstracts created in this process would be essential in the process of generating additional subjects if we continue to leverage AI in our descriptive workflows.

With the majority of the inaccuracies due to transcription errors, additional efforts must be made to discover these errors via human intervention during the quality control process. However, the titles and abstracts that are either 100% accurate or with small inaccuracies provide effective and well described context to the images.

The consistent application of use of the transcription coupled with thorough descriptions of the figures and imagery of the photographs creates dependably ordered and logical titles and abstracts, but it also means that transcriptions must be closely monitored to insure accurate and complete descriptions. Subjective descriptions of visual symbolism and moods are also consistent, but do not adhere to general cataloging practices, so should be removed to save tokens and avoid unintended partiality or bias. Inconsistencies such as mentions of weather and seasons, descriptions of editorial crop markings, and explanation of how the transcription is used in the abstracts could be improved with further prompt engineering, and/or with the addition of collection notes and folder-level subjects, providing users with additional context about the use of AI in generating descriptions as well as the history of the collection.

Photographer

Completeness, accuracy, and consistency

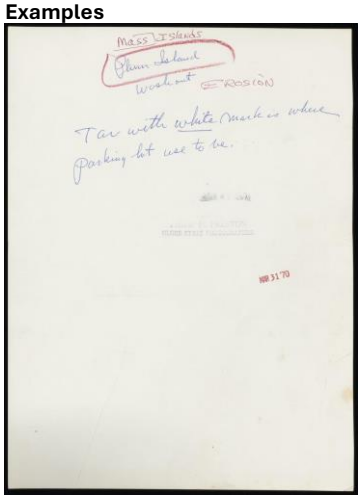
Measured by human transcription of the photographer's name, then comparing that transcription to the output of Gemini. Additional testing was completed to check if output was an exact match, partial match, misattribution based on unrelated description written on the back, or totally missed by Gemini.

- Number of photographs with a photographer name: 94
- Number of photographs with a photographer name provided by Gemini: 79
  - Number of Gemini name was 100% accurate: 73
- Errors:

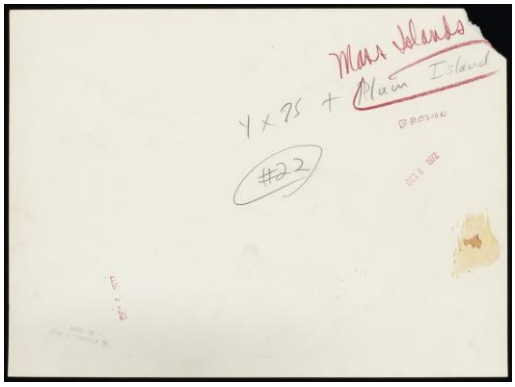
Type	Number
Missed by Gemini	15

Slight misspellings or missing middle initials	9
Name of someone besides photographer or hallucination	3
TOTAL	27

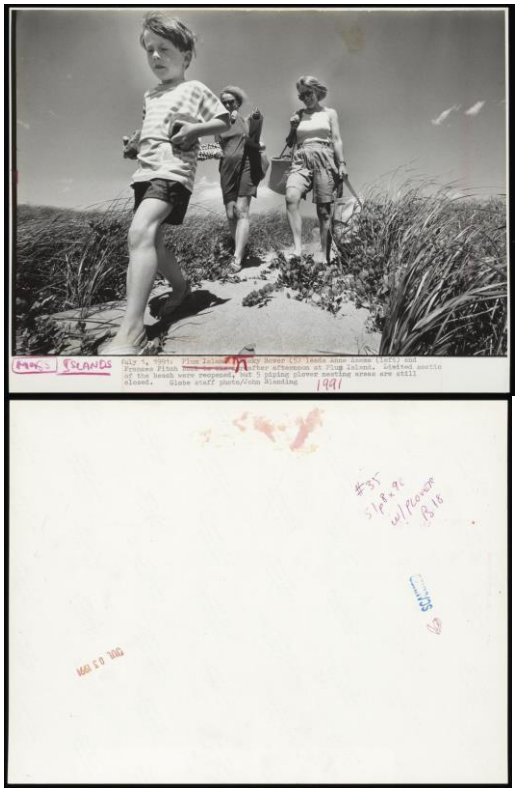
Examples



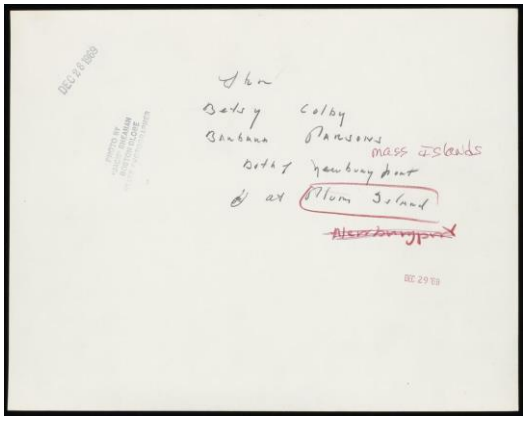
Filename: M214_1015936037_0002_0023
Human transcribed photographer name: Philip N. Preston
Gemini provided photographer name: Philip N. Preston



Filename: M214_1015936037_0002_0019
Human transcribed photographer name: John V. Connolly, Jr.
Gemini provided photographer name: John Connolly, Jr.



Filename: M214_1015936037_0001_0060
Human transcribed photographer name: John Blanding
Gemini provided photographer name: [MISSING]



Filename: M214_1015936037_0001_0061
Human transcribed photographer name: Jack Sheahan
Gemini provided photographer name: Betsy Colby

Analysis

Photographer names are important for the Boston Globe images for several reasons. Researchers, academics, students, and other groups may want the photographer names to research particular Boston Globe staff, subject matters a person covered over time, etc. Additionally, as the creator of the resource, it follows both local and national cataloging conventions to include the name of the photographer if available. But perhaps most importantly, photographer names are essential to determining copyright for these images and therefore being able to make images available to the wider public.

We are currently working through a process that will automate matching Gemini-generated photographer names and dates with spreadsheet data about photographers employed over time provided by the Boston Globe. With 73% of the photographer names being an exact match, that drops the total number of records that require additional human intervention significantly, saving processing time. The photographer names that do not have an exact match to Boston Globe data, will need to be checked by a cataloger to determine copyright, so fixing these inconsistencies will be an expected part of the quality control process.

Date

Completeness, accuracy, and consistency

Similarly to photographer name, date was measured by transcribing the earliest date written on the back of the image, then comparing that human transcription to the output of Gemini. Gemini provides two dates if more than one is transcribed from the image. Multiple dates on the back of Boston Globe images are not uncommon, so many records will have date ranges that likely represent the time between the shoot and publication.

- Number of photographs with dates written on the back: 99
- Number of photographs with dates provided by Gemini: 73
    - Number correct: 64 (85% recorded were correct)
- Number of photographs with dates that Gemini didn't record: 27
- Incorrect dates: 9 (12% of total dates recorded)
    - Dates inaccurate by more than 1 year: 3
    - Dates inaccurate by more than 5 years: 1

Analysis

Dates are one of the most common ways that users facet photograph collections, and therefore necessary for effective discovery. Although 12% of the dates were incorrect, only 3% dates were off by more than 1 year and only 1% was off by more than 5 years, meaning that the recorded dates will still be consistently valuable to our users. There is still an issue of missing dates, but with the majority recorded by Gemini, this still saves catalogers quite a bit of time.

As mentioned in the photographer name discussion, both names and dates will be used in combination to decide copyright status. Some photographers worked for the Boston Globe as staff

Commented [SS3]: I think this statement is saying something different than the last bullet above about incorrect dates? But it's confusing - I'm not sure if the numbers are wrong or if this is a different point of analysis than the 6 incorrect dates within a year/8 within 5 years stat above?

photographers their entire careers and some acted as freelance photographers for a portion of their tenure. This distinction affects copyright and therefore both the presence and accuracy of dates is important. Records with missing dates will need to be checked to complete the copyright process, but with 3% of the dates off by more than a year, it will be important to decide if that error rate is low enough to merit relying on Gemini generated dates to determine copyright status.

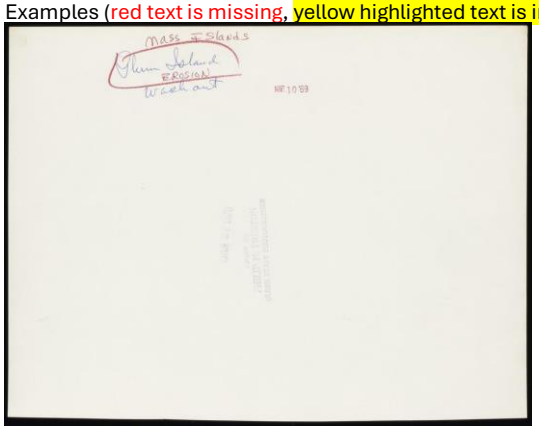
Transcription

Completeness and accuracy

Both were measured on the character-level. Two errors were recorded: One, when a character was wrong, and 2, if a character was missing from the Gemini provided transcription. This was done by comparing the images of the backs of the photographs to the Gemini transcription and recording every time a character was incorrect or missing.

- Average number of characters recorded per photograph: 180.9
    - Average number of characters incorrect: 1.88
    - Average number of characters missing: 2.44
    - Average percentage of character errors (incorrect and/or missing): 3.92%
        - Median percentage: .77%
- Number of records with 100% complete and accurate transcription: 42
- Number of times the transcription error affects some part of the description (title, abstract, date, and/or photographer): 37
- Number of times the transcription error creates a significant inaccuracy in the title or abstract: 17

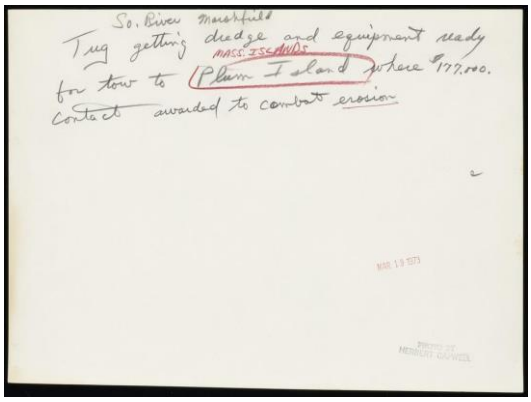
Examples (red text is missing, yellow highlighted text is inaccurate)



Filename: M214_1015936037_0002_0035
Transcription:

Mass Islands
Glenn Plum Island
EROSION
Wash out
MAR. 10 '69

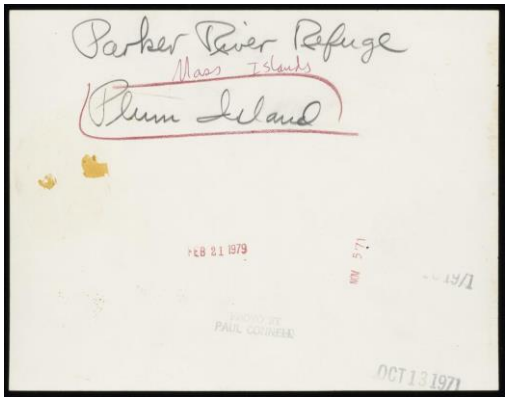
PHILIP N. PRESTON
GLOBE STAFF PHOTOGRAPHER
JAN 27 1970



Filename: M214_1015936037_0002_0017
Transcription:

So. River Marshfield
Tug getting dredge and equipment ready
MASS. ISLANDS
for tow to (Plum Island where $177,5000.
Contact awarded to combat erosion

Mar 19 1973
Photo by
Herbert Capwell



Filename: M214_1015936037_0001_0037

Transcription:

Parker River Refuge
Mass Islands
Plum Island

FEB 21 1979
PHOTO BY
PAUL CONNEELELL
NOV 5 ‘71
- 19/1
OCT 13 1971

Consistency

Given how inconsistently the information written on the back of the photographs is formatted, the consistent spacing and ordering of the text in the transcriptions is impressive. Single spaces between lines regularly indicate groups of text, and blank lines indicate larger spaces between text, helping to better visually organize the transcription for users. Additionally, the recording of the capitalization and formatting of handwritten cursive is especially notable. The most regular errors are character omissions because of faded stamps, upside down handwriting, or messy cursive, which do lead to missing or incorrect data, but that would be partially remedied by human intervention during the QC process.

Analysis

Many of these errors will be caught when reviewing photographer names and checking titles and abstracts for the presence of folder titles if that level of review was integrated into the quality control process (for example in M214_1015936037_0002_0035 the correct name of Plum Island, is the name of the folder, and therefore would be caught by a simple folder-wide keyword search). Additionally, further errors could be caught during general quality control (selecting a percentage of the records for full review), and checking the accuracy and completeness of the transcription is a reliable indicator of the quality of the rest of the metadata provided by Gemini. However, there will still be character omission or variation that couldn’t be caught without detailed and time-consuming item-level transcription review. The threshold of acceptable error must be considered, and the possibility of mis-transcribed text should be added to the metadata records in a publicly available content note.

Next steps

Bias testing

Testing for bias is an essential step in reviewing the output of Gemini data. Although this is done on the item-level when records are reviewed, there are also systematic ways to search Gemini output for language used that may show several kinds of biased description. This tool, created by the DLF Digital Accessibility Working Group (DAWG), called MaRMAT, uses a lexicon of terms related to several categories identified as commonly problematic according to current reparative description practice: aggrandizement, disability, gender, LGBTQ, mental illness, racial euphemisms, race, slavery and incarceration. Additionally, terms used to describe violent acts have been added to the lexicon terms in order to find images that may require additional review for content statements.

- For more information about the MaRMAT tool: https://github.com/marriott-library/MaRMAT
- For DAWG's inclusive metadata toolkit: https://doi.org/10.17605/OSF.IO/2NMPC

This testing set did not have Gemini-generated descriptions containing language identified as containing bias, or with terms related to reparative description. However, this group of photographs contained maps, aerial views of rural coastlines, landscapes, animals, and images of white, cis-gendered adults in passive activities. So, to properly test possible bias in Gemini output, a larger testing set has been created with photographs of diverse persons identified in Boston Globe's folder descriptions as members of underrepresented and/or marginalized groups and people engaged in forms of civil action. Folder titles include:

Boston--Chinatown
Boston--Crime--Prostitution
Boston--Crime--Shootings, Stabbings, Murders
Boston--Demonstrations--Blacks (See Boston--Blacks)
Boston--Hispanics
Boston--History--Black
Boston--Hospitals and Clinics--Brigham and Women's
Boston--Parades--Gay People
Boston--Police Department--Women
Boston--Racial Violence and Racism
Boston--Riots
Boston--Schools--Public--Blacks
Demonstrations -- Massachusetts -- Boston
Gay rights -- Massachusetts -- Boston
Massachusetts State House (Boston, Mass.)

Output of this testing will be discussed in a separate report that will focus on the titles, subjects, and transcriptions generated by Gemini for this larger group of photographs. Additionally, the report will discuss how to minimize bias and descriptive harm inherent in the text written on the back of the photographs, as historic folder titles already reveal likely interpretive issues because of their use of outdated and offensive terms. If necessary, additional prompt-engineering will be tested to mitigate any possible harm and/or continued quality control using the MaRMAT tool will also be discussed.

Prompt re-engineering and additional testing

Based on concerns about Gemini output discussed in this report, the outcome of bias testing, and emerging research about prompt formation, updates to the prompts used to generate the Gemini output will be initiated before another round of item-level testing for accuracy, completeness, and consistency.

The second round of testing will review a larger group of records for the possible metadata quality issues described in this report but will also engage a larger group of testers to mimic real-world cataloging workflows, engage additional library professionals of differing expertise, and prevent bias. The output of that testing will best determine the quality of metadata output likely for full Gemini implementation and will provide additional data to assist in determining effective quality control procedures at scale.

Possible user studies

User studies may also be a necessary step in determining the manner of implementation. Testing discovery behaviors of various user groups searching, faceting, and browsing through a subset of the Boston Globe collection that's been described using Gemini may help improve metadata implementation. There is current research about user behaviors when engaging with historic photograph collections, but the size of and themes contained in a newspaper morgue may require additional research to best understand the needs of the DRS' variable user groups such as faculty, staff, students, historians and independent researchers.

There are also more options for limited user testing such as conducting interviews with a key user group, University Archives and Special Collections (UASC) staff. As the stewards of the Boston Globe photo morgue, as well as superusers of the DRS discovery layer for this collection, archives staff could provide valuable insight. UASC's reference archivists, Molly Brown and Grace Millet, would likely have useful information about the common keyword searches, description needs to support reference requests, and discovery challenges associated with their work supporting community inquiries into the collection.

Commented [KK4]: Maybe the user testing could be an interview with Molly and Grace on their behavior searching this collection (since I expect they are the ones who search it the most)?

Commented [FD5R4]: I love this idea Kim, thank you!

Resources

DLF Metadata Assessment Working Group. DLF Digital Library Assessment Metadata Working Group, https://dlfmetadataassessment.github.io/projects/framework/

Gavrilis, D. *et al.* (2015). Measuring Quality in Metadata Repositories. In: Kapidakis, S., Mazurek, C., Werla, M. (eds) Research and Advanced Technology for Digital Libraries. TPDL 2015. Lecture Notes in Computer Science(), vol 9316. Springer, Cham. https://doi.org/10.1007/978-3-319-24592-8_5

Kumpulainen, S., & Late, E. (2022). Struggling with digitized historical newspapers: Contextual barriers to information interaction in history research activities. Journal of the Association for Information Science and Technology, 73(7), 1012–1024. https://doi.org/10.1002/asi.24608

Late, E., Ruotsalainen, H., Kumpulainen, S. (2024). Image searching in an open photograph archive: search tactics and faced barriers in historical research. Int. J. Digit. Libr. 25, 4 (Dec 2024), 715–728. https://doi.org/10.1007/s00799-023-00390-1

Traill, S. (2021). Core Concepts and Techniques for Library Metadata Analysis. Code4Lib Journal (59). https://journal.code4lib.org/articles/16078