How Games Can Help Us Access and Understand Archival Images

Mary Flanagan and Peter Carini

Abstract

A lack of quality metadata is a key problem encountered with mass-digitization projects as institutions strive to "go digital." This paper reports on a pilot study of Metadata Games, a software system that uses computer games to collect information about archival images in libraries and archives as these institutions digitize millions of items across national collections. Games offer a unique advantage for collecting metadata because they can entice users who might normally be inclined to visit archives to explore humanities content and, in the process, contribute to vital records, and they can work in a wide-scale, distributed fashion to collect much more metadata than a typical archives staff member could contribute alone in the same time frame. Metadata Games can be used to enhance knowledge about images associated with particular disciplines and fields, or in interdisciplinary collections. This open-source system is easily customized to meet each institution's needs. By inviting mass participation, Metadata Games opens the door for archivists, researchers, and the public to unearth new knowledge that could radically enhance scholarship across the disciplines. Metadata Games expands what researchers, students, and the public can encounter in their quest to understand the human experience. Games offer great promise for humanities scholarship by uniting the culture of the archives with a diverse user base, including researchers, hobbyists, and gamers.

> Effective democratization can always be measured by this essential criterion: the participation in and access to the archive, its constitution and its interpretation.
>
> —Jacques Derrida[1]

Derrida's note on the archive is astute. The key to civic engagement for society and scholarship is in the access to, and creation of, its own archive. With the advent of the twentieth century, and increasingly since World War II, there has been exponential growth in the volume of information. This tremendous increase has made finding relevant, useful materials in libraries, archives, and special collections increasingly challenging for scholars, researchers, and students. Archives and special collections face particular challenges because of the unique nature of their holdings, the varied formats of the materials that make up their collections, and the sheer volume of data within those collections. The volume of materials in archives and special collections has, for a long time, far surpassed the staffing provided to create access to these holdings.

The move to electronic recordkeeping and the ability to create digital surrogates of analog data offer archives and researchers both challenges and opportunities. For researchers, one primary opportunity is the ability to mine textual materials for data that would otherwise be difficult to find. For instance, a digitized version of Vilhjalmur Stefansson's *The Friendly Arctic* allows a researcher to search the text for the word "games," a concept not present in the book's index. This kind of access has the potential to change radically the way research is conducted. Digitization offers wonderful possibilities for improved access to printed textual materials, but handwritten documents, audio, and visual materials provide ongoing challenges because they require staff-intensive activities involving either transcription or the creation of metadata.

As a way around this problem, libraries and archives have taken the completely legitimate, albeit less than ideal, approach of applying minimal descriptive metadata across bodies of images (at collection, series, and file levels). For example, a single identifier such as "Dartmouth College" might be used to describe an image collection or series. In the best case, some description might make it clear what aspects of the college an image represents, but in many cases there may only be a single point of entry to potentially hundreds of images. Applying minimal metadata as a "blanket" across large collections provides a basic level of categorization, but visual images are full of details and can depict more than one object, or reference several topics at once. Stated differently, an image can have a focus, but may comprise multiple themes or subjects. Images also have physical details—format, color information, date, and so on—few of

---

[1] Jacques Derrida, *Archive Fever: A Freudian Impression* (Chicago: University of Chicago Press, 1996), 4.

which can be provided with this minimal method. It seems obvious that item-level description is highly desirable when it comes to large, diverse sets of photographic images. The problem is how to provide this in a simple, cost-effective manner.

Literature Review

Past practice in applying metadata to analog images focused on metadata derived from the materials and their context, whether at the collection or item level. These elements generally include creator, title, dates, physical description, subject, access and use conditions, and identification number.[2] Very little metadata related to the content depicted in the images are captured other than at a very high level. Recommended description is supplied primarily through subject headings that rely on a high-quality knowledge organization system to supply controlled vocabulary. In 1996, Nancy Bartlett noted in her article "Diplomatics for Photographic Images: Academic Exoticism?" the great importance of the intrinsic, essential characteristics of photographic images used in research. Photographic records differ from other records, and "the role of the content of the image is . . . of fundamental importance. Yet this remains without analysis by archivists."[3] Bartlett urges archivists to study images more completely, thinking beyond "a linear sequence of verbal language" to present information about the record.[4]

More recently, the expense of developing high-quality knowledge organization systems for libraries and archives has led to questions about the viability of continuing to utilize and maintain them. This has led a wide range of professionals to consider acquiring content descriptive metadata through crowdsourcing at the item level. Several projects, most notably the Library of Congress Photos on Flickr Pilot Project[5] and the New York Public Library's What's on the Menu? project,[6] have already explored crowdsourcing as an option for metadata enhancement.[7]

---

[2]   Helena Zinkerman, "Description and Cataloging," in *Photographs, Archival Care and Management*, ed. Mary Lynn Ritzenthaler and Diane Vogt-O'Connor (Chicago: Society of American Archivists, 2006), 168.

[3]   Nancy Bartlett, "Diplomatics for Photographic Images: Academic Exoticism?," *The American Archivist* 59 (Fall/Winter 1996): 491.

[4]   Bartlett, "Diplomatics for Photographic Images," 491.

[5]   Library of Congress Photos on Flickr, Library of Congress, http://www.loc.gov/rr/print/flickr_pilot.html, accessed 10 June 2012.

[6]   What's on the Menu, New York Public Library, http://menus.nypl.org/, accessed 10 June 2012.

[7]   Besiki Stvilia and Corinne Jörgensen, "Member Activities and Quality of Tag in a Collection of Historical Photographs in Flickr," *Journal of the American Society for Information Science and Technology* 61 (December 2010): 2477–78.

Prior Work

Because the human capacity to transform massive amounts of raw data into information and knowledge is limited, archivists and designers must work to invent IT tools to specifically assist with this challenge. Ninety-eight percent of all children play computer or video games, and many Americans are familiar with everyday digital experiences that are increasingly spatial, graphic, sonic, and game related.[8] In addition, a game-based metadata capture system has the potential to help organize the massive amount of raw data that could be collected by participants drawn to games and computer play. These interdependent advantages come together to make games an ideal tool for crowdsourcing for archives.

Others' efforts have shown how games or crowdsourcing can be used to generate data. The influential work of Luis von Ahn, a computer scientist at Carnegie Mellon University, helped inspire the Tiltfactor game design research lab at Dartmouth College to investigate games for gathering metadata.[9] In his research on metadata gathering efforts for applications now in common use at Google, von Ahn designs and creates small applications and games that help solve problems. Some of his test cases include ReCaptcha, software that decodes the *New York Times* archives while providing secure Internet login checks. In his work, von Ahn notes that just a thousand players can feasibly tag well over eleven thousand images in a single day if each plays for just an hour. An employee would have to tag nine hundred images a day for more than 125 days, or around four months of full-time work, to achieve the same level of output. As a comparison, the Tiltfactor lab's past game *Layoff* attracted over a million players, with an average play time of ten minutes, during its first week of release in 2009. Extrapolating from these figures, even short-burst casual games of five to ten minutes would be extremely effective in generating large amounts of data per player. Useful high-quality data, in accordance with our design, would eventually emerge through the sheer scale of such an initiative. As valuable and informative as von Ahn's programs are, shortcomings make them less than viable solutions for archives and institutions to implement. They are not free, open source, customizable, or sharable.

In addition to studying von Ahn's work, our team examined the progressive experiments conducted by the Library of Congress with the online commercial photo-sharing software Flickr as a possible model for using crowdsourcing to augment archival metadata. In 2008, the Library of Congress uploaded more

---

[8] Amanda Lenhart, Joseph Kahne, Ellen Middaugh, Alexandra Macgill, Chris Evans, and Jessica Vitak, *Teens, Video Games and Civics* (Washington, D.C.: Pew Internet and American Life Project, 2008).

[9] Luis von Ahn and L. Dabbish, "Labeling Images with a Computer Game," *Proc. CHI'04* [Association for Computing Machinery, Special Interest Group on Computer-Human Interaction, Conference on Human Factors in Computing Systems] (New York: ACM Press, 2004), 319–26. See also http://www.tiltfactor.org/metadata-games, accessed 10 July 2012.
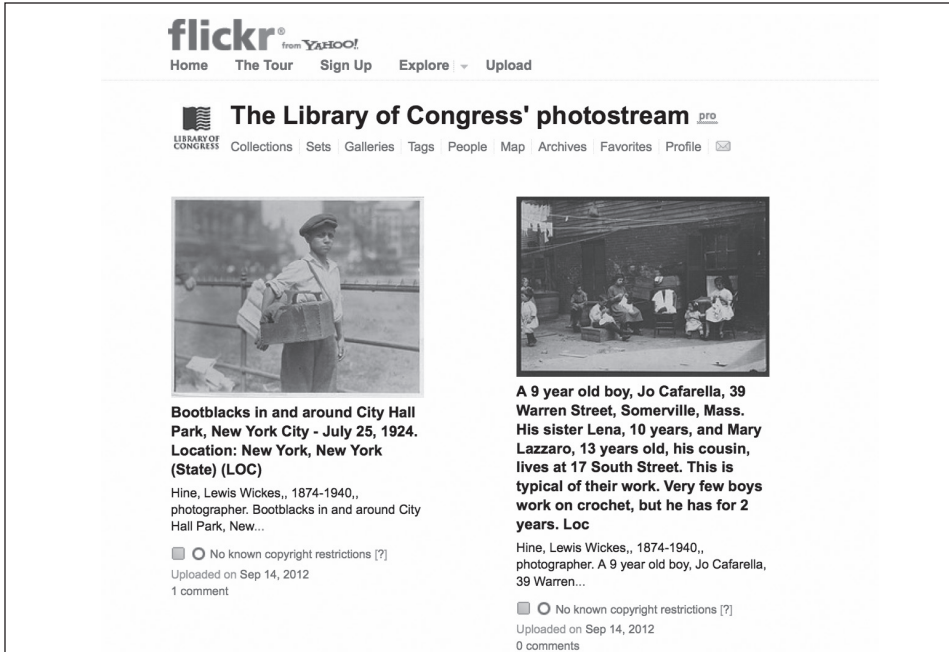
F I G U R E   I .   **Entries from the 2008 Library of Congress's tag-gathering Photos on Flickr Pilot Project**

than four thousand images from the Farm Security Administration/Office of War Information and the George Grantham Bain News Service on the photo-sharing site, encouraging users to apply their own tags and comments.

This popular crowdsourcing experiment was a success. The public tagged posted images at a high rate: the Library of Congress reported, "67,176 tags were added by 2,518 unique Flickr accounts" and "less than 25 instances of user-generated content were removed as inappropriate."[10] Data provided by the general public included corrections and additions to the existing metadata through the identification of places, events, individuals in the images, and the dates on which the images were originally created. Additionally, the public added colloquial expressions and informal keywords and tags that would not have been included in traditional library terminology. "Popular concepts like 'Rosie the Riveter' (added to 77 of the FSA/OWI photos) provide another avenue for retrieval beyond the Library's controlled vocabulary terms of 'Women—employment' and 'World War, 1939–1945'."[11] At the conclusion of the Flickr experiment, the Library of Congress found that 80 percent of the tags contributed by users closely matched metadata already in the library's well-

[10] Michelle Springer et al., *For the Common Good: The Library of Congress Flickr Pilot Project* (white paper, 30 October 2008), 4–5, Library of Congress, http://www.loc.gov/rr/print/flickr_report_final.pdf, accessed 7 November 2011.

[11] Springer et al., *For the Common Good*, 5–6.

FIGURE 2. **Flickr comments on a Coney Island, N.Y., poster for Ruth the Acrobat**

documented records for the images, or that the collected data described features that were visible in the images themselves. This means that for images about which little is known, crowds could indeed match or nearly match metadata that previously only highly skilled staff could provide. Annotations made directly on the photos were less useful than the tags left by users, because image tags that identify particular photo features or components tend to attract graffiti-like writing and smart-aleck humor.

To continue the Flickr project, the Library of Congress staff estimated they would need a minimum of one additional half-time employee, and, preferably, an additional full-time employee. Most cultural institutions cannot afford a half-time employee at times of institutional budget cuts and reprioritization. So, while the Library of Congress's Flickr project was surprisingly successful, it may not be a realistic alternative for most archival institutions because it requires new paid positions and it relies on commercial software with limiting restrictive rights options. Libraries and archives need another solution.

The New York Public Library's What's on the Menu? project is another example of using a crowdsourcing platform to collect data. This case involved transcribing the New York Public Library's collection of forty thousand restaurant menus.[12] By 2012, users of the platform had transcribed nearly seven hundred thousand dishes from 11,264 menus. But this project's strength is also its limitation: it is topic- and collection-specific and is not meant thus far to be

---

[12] What's on the Menu? Help Transcribe The New York Public Library's Historical Menu Collection, http://menus.nypl.org/, accessed 18 August 2012.

distributed as a tool that can be tailored to a wide range of collections or audiences.

The diverse tags generated by crowdsourcing projects may also mean that institutions will need to reflect upon the new types of knowledge that might surface—new classifications, observations, descriptions, narratives, and practices. Another challenge relates to the accuracy of information given by the public. Crowdsourcing scholar Eric von Hippel asks the vexing question, "Can we truly democratize innovation in crowdsourcing?"[13] In other words, our task is to design a system that engages the public while not diluting the accuracy of our records. But, in the case of library and archives metadata, would it not be better to know something, as opposed to nothing, about an object? Our team endeavors to engage many people working together to create and verify accurate tags to ensure that the data are useful. We also aim to broaden participation while maintaining high standards for data accuracy. Indeed, issues of access, openness, and public participation are deeply value laden. Those traditionally adding metadata to library materials would be primarily white, middle class, English speaking, educated, "wired," and highly trained in the information sciences.[14] Our team approached the project with the supposition that diverse voices will affect data for the better, uncover different points of view, reveal different uses for objects, and contribute specialized knowledge otherwise difficult to collect.

In the design of the Metadata Games project, the challenge has been to develop a software system able to entice participation and reward contributors for offering accurate information while at the same time providing an easy way for archivists and data managers at cultural institutions to interface with, and use, the data—all at little to no cost. This is not as easy as it seems. Verifying accuracy, especially on materials about which little is known, is a particularly worrisome problem in projects that open up archival input to the public. Scholarship on tagging has contributed to our understanding of the relationship

---

[13] Eric von Hippel, *Democratizing Innovation* (Cambridge. Mass.: MIT Press, 2005).

[14] Amanda Lenhart, John Horrigan, and Deborah Fallows, "Content Creation Online: 44% of U.S. Internet Users Have Contributed Their Thoughts and Their Files to the Online World," Pew Internet and American Life Project, http://www.pewinternet.org/Reports/2004/Content-Creation-Online.aspx, accessed 18 August 2012. Amanda Lenhart and Mary Madden, "Teen Content Creators and Consumers: More Than Half of Online Teens Have Created Content for the Internet; and Most Teen Downloaders Tink that Getting Free Music Files Is Easy to Do," Pew Internet and American Life Project, http://www.pewinternet.org/Reports/2005/Teen-Content-Creators-and-Consumers.aspx, accessed 18 August 2012.

between user tagging and professional indexing.[15] Marshall notes that narrative descriptors might offer better metadata than tags, but they are more difficult to parse.[16] Metadata Games incorporates both types of inputs.

With an open-source back-end, and flexible, extensive system architecture, Metadata Games provides a way to motivate participation with the ease of use of the Library of Congress's approach to image tagging; Metadata Games can be used to focus interests as the New York Public Library's menu project does; and it goes further than either of these projects by providing a broad and flexible base upon which crowdsourcing projects can grow. Further, Metadata Games overcomes the limitations and shortcomings of previous efforts because the software is designed to meet the unique needs of vastly different institutions. A core ethos of the project is to create an effective crowdsourcing tool that is freely available for cultural institutions. The ways in which the Metadata Games software will fit into existing institutional structures are therefore core design considerations for the project.

The Promise of Games

The thesis behind the Metadata Games project is that games and gamelike activities can be used to attract the public to participate in providing valuable descriptive metadata for archival images. As previously mentioned, crowdsourcing can produce thousands of tags with just five minutes of play per day by a thousand visitors from around the globe. Therefore, a game approach that attracts participants to a site and facilitates tagging in an enjoyable way could revolutionize the way additions are made to an institution's knowledge base.

The Metadata Games team created a suite of games catering to different play styles, as a proof-of-concept system to be shared with others as a free and open-source system. To test proof of concept in early versions of the system, the team used proprietary software (Adobe Flash) to create some of the "front-end" games, or games that would be visible to participants. These were then rewritten in HTML 5 to conform to open-source standards. The Metadata Games

---

[15] Joan Beaudoin, "Flickr Image Tagging: Patterns Made Visible," *Bulletin of the American Society of Information Science and Technology* 34 (October/November 2007): 26–29, ASIS&T, http://www.asis. org/Bulletin/Oct-07/beaudoin.html, accessed 22 June 2012; Chris Landbeck, "Trouble in Paradise: Conflict Management and Resolution in Social Classification Environments," *Bulletin of the American Society of Information Science and Technology* 34 (October/November 2007): 16–20, ASIS&T, http:// www.asis.org/Bulletin/Oct-07/landbeck.html, accessed 7 May 2012; Elaine Ménard, "Image Indexing: How Can I Find a Nice Pair of Italian Shoes?," *Bulletin of the American Society of Information Science and Technology* 34 (October/November 2007): 21–25, ASIS&T, http://www.asis.org/Bulletin/Oct-07/ menard.html, accessed 1 August 2012.

[16] Catherine C. Marshall, "No Bull, No Spin: A Comparison of Tags with Other Forms of User Metadata," *Joint Conference on Digital Libraries (JCDL'09)*, 15–19 June 2009, Austin, Texas.

installation kit releases this game code as well as the underlying database, so others can open, alter, and recompile the project as they wish. In the first eight months of the project, the team developed seven games to appeal to different types of players, ranging from single player, open-ended tagging activities that were barely "gamelike" to two-player, very "gamelike" networked guessing games.

In Zen Tag, players input, at their own pace, words and phrases that describe the image

FIGURE 3. **Zen Tag**

before them. This is an example of the simplest activity in Metadata Games: a tagging window that rewards players with points for each tag they enter, with higher points awarded for tags that prior players have provided. Matching a certain number of other players then "verifies" that the term is relatively reliable.

Once a base set of tags is accumulated, they become off limits to players, and thus, through a process of elimination, we are able to validate less-frequently used tags as they rise to the top in terms of player usage. Zen Tag has no fail state and no inherent risk, so by a strict definition of a game, this would be a rewarded "activity" rather than a game per se. Some players truly love this activity, however, and spend significant time with Zen Tag, including a surprising number of young male gamers who belie stereotypes about their dislike of noncompetitive play. We found that this application generates the most tags per player in the time played. For those players who prefer a more competitive experience, there are other games from which to choose.

In the multiplayer game Zen Pond, we introduce real-time networked collaboration. Without seeing each others' entries, players enter terms that describe the image and that also might match the other players' entries in real time. Direct matches mean points for both players, with additional points awarded to players for matching complex words and/or phrases.[17] At the end of each round, players find out which items they have matched. Zen Tag and Zen Pond serve to gather a large amount of new metadata and to verify terms: terms

---

[17] This essay describes research conducted on the first complete version of the software, Metadata Games 1.0. The next iteration of Metadata Games, Version 2.0, will also include the ability to give special points for inputting valid scientific or discipline-specific terms by matching these to customized technical word lists.

are "verified" for an image when a certain number of repeated uses are recorded. These can be set by each institution (e.g., twenty duplicate entries of "snow" likely indicates the validity of the term). Both Zen Tag and Zen Pond use "words to avoid" lists to motivate players to dig deeper and provide novel tags.
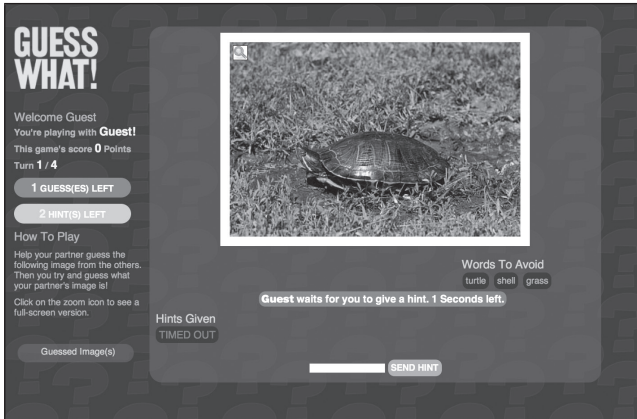
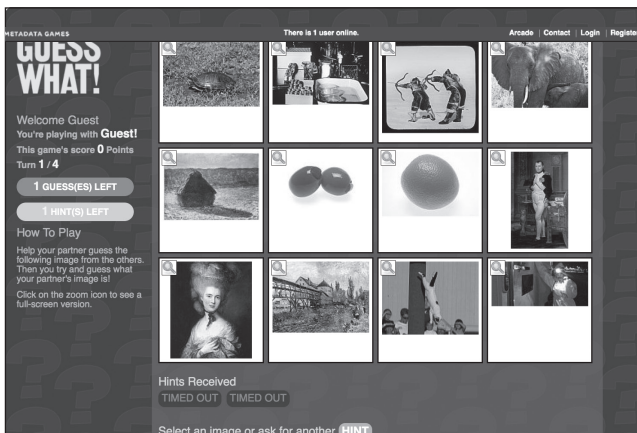**FIGURE 4.** **Guess What!**

FIGURE 5. **What player two sees in Guess What!**

Guess What! is a real-time, collaborative, two-player game. In Guess What!, player one sees one image on the screen such as a mammal or hunters. Player one must then describe the image to another player across the network, so that the other player could correctly select the image from an array of images on his or her screen.

Therefore, if several mammals or hunters are depicted in the array of images, player two will have a difficult time choosing the correct image unless he or she receives specific tag hints from player one. Figure 5 shows an example of the array of images that player two might see in Guess What!. Players enjoy giving each other accurate but arcane hints, thus raising the specificity of the vocabulary terms used. The game serves the goal of the project in two ways: it helps to collect new metadata on images that have none prior, and it helps verify existing terms by monitoring their frequency of use among players.

If players wish to play this type of game against the computer, as in the case of the game What's That!, the computer randomly selects one of the metadata

terms associated with one image in an array of twelve. The player has to guess from which image the term originated. In these ways, players earn points by verifying the tags associated with an image.

In Cattygory, players are prompted to describe an image. The game uses specific categories for users to input tags, rather than being completely open ended. Cattygory allows the archivists and designers to create prompts to direct the particular types of data input that might be most useful.

These categories help the team to "drill down" in terms of achieving a desired level of specificity among tags. Thus, a game could show an image of a woodchuck and the prompt "Mammal," with the challenge to supply a more specific term, thus

FIGURE 6. **Cattygory**

requiring the player to add "woodchuck," or the Latin "*Marmota monax.*"[18] The match to the classification system in the archives would be "Mammal— Woodchuck, or Mammal—Woodchuck, *Marmota monax.*"

Scholars are beginning to question the biases embedded in the way archival institutions have traditionally recorded metadata.[19] Scholarship in the digital humanities helps us reflect upon which data are considered "valid" for the archives and which are not, such as feelings, indigenous knowledge, beliefs, and so on—these examples could be either valid or invalid depending on the needs of the archives or the type of content contained in a given institution or collection. Archivists often designate a user community and "project" the values of that group on the metadata they record for posterity.[20] Games could be used to collect metadata such as feelings, colors, and shapes from particularly unique

---

[18] In the future Metadata Games Version 2.0, the system prompts might take players further away from traditional classification systems used in libraries, but semantic relationship building can help ensure that tags conform to specific vocabularies.

[19] Mark Gahegan and William Pike, "A Situated Knowledge Representation of Geographical Information," *Transactions in GIS* 2006, 10, no. 5 (2006): 727–49.

[20] KKari Kraus, "Conjectural Criticism: Computing Past and Future Texts," *Digital Humanities Quarterly* 3 (Fall 2009), http://www.digitalhumanities.org/dhq/vol/3/4/000069/000069.html, accessed 6 February 2012.

perspectives and to add these to a collection's lexicon.[21] For example, a photo archives may wish to seek out tags related to the moods or feelings the images portray or evoke, and such an organization could customize a game's prompts accordingly to mine for that type of information.

Our Technical Model

System Architecture

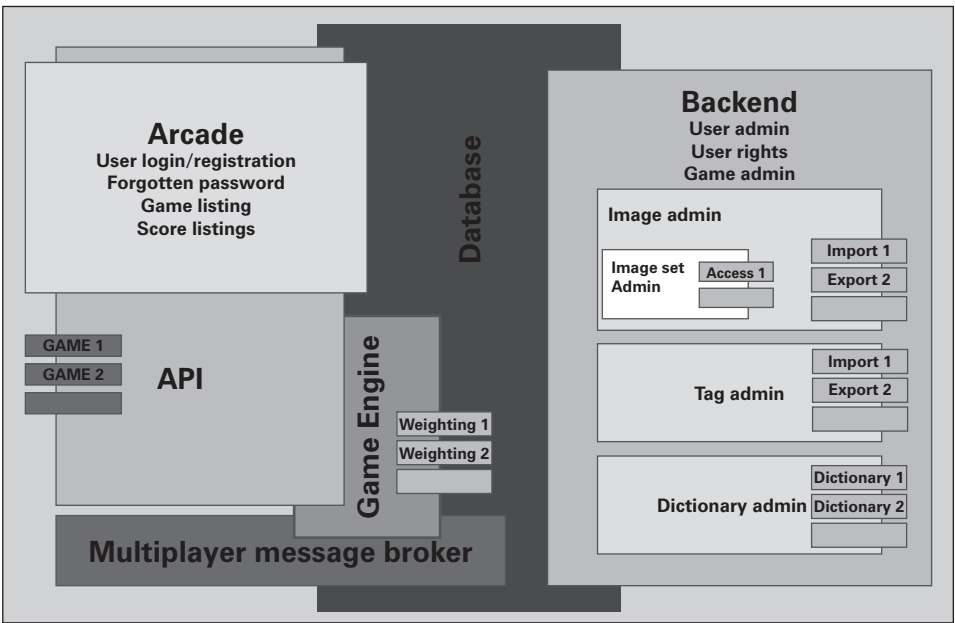Metadata Games is an iteratively designed proof-of-concept system.

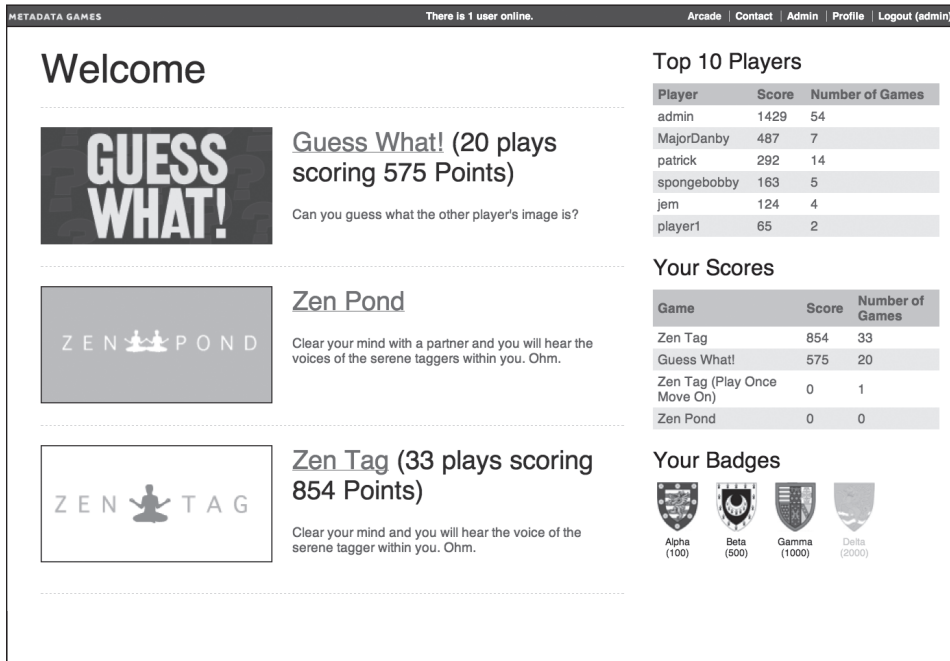**F I G U R E 7 .**   **Metadata Games system architecture**

The core system is created to have extended capabilities. Figure 7 shows the various components of the system and how they fit together. All components are extendable, but, in particular, the games, image import and export, weighting, and dictionary modules are written as plug-ins (software components designed

---

[21] Lacy Shultz, director of collections access at the Museum of the City of New York (MCNY), notes that some researchers might want to find photos that feel "happy" or were predominantly blue—and MCNY's website could not meet that need. As documented in a talk at Pratt Institute by John Tomlinson, "Crowdsourcing and Linked and Open Data: New Ways to Make Collections Visible," 14 October 2011, SLA@Pratt, http://mysite.pratt.edu/~sla/events/2011crowdsourcingandlinkeddata .html, accessed 2 January 2012.

to add specific abilities to the overall program) to allow for easy adjustment, tweaking, and expansion. Any number of new games can also be added as plug-ins using the Metadata Games application program interface (API).

A p p l i c a t i o n   P r o g r a m   I n t e r f a c e

The Metadata Games API uses Digest Authentication[22] for secure authentication because of its simplicity and because end users cannot be expected to run the system on a secure (https) server. The API also makes use of JavaScript Object Notation (JSON) to exchange data between server and client.

F I G U R E   8 . **Metadata Games: The Arcade**

The front-end of Metadata Games, called "The Arcade," is completely customizable. All the material that appears to a set of players or even subsets of players is controlled through the back-end. There is an area for badges in the design of the system. The badges function in the current system in a limited fashion and can be customized to each institution. Players' scores are saved and, if desired, displayed across the suite of games the host institution employs.

---

[22] For more information on Digest Authentication see *Wikipedia*, s.v. "Digest access authentication," http://en.wikipedia.org/wiki/Digest_Access_Authentication, accessed 25 June 2012.

Games

All games are implemented with HTML, CSS, and JavaScript to offer cross-platform support and to allow them to be open for further development. By making use of a consistent technological approach, it is very easy for other institutions or individuals to build games in HTML5 or mobile phone applications.

Pilot Test Background

The Test Collection

We tested the Metadata Games system and the content it collects against the detailed metadata created for images in the Stefansson Collection on Polar Exploration at the Rauner Special Collections Library at Dartmouth College. The Stefansson Collection is one of the world's most extensive bodies of research materials on the North and South Poles. Founded as the private research collection of the Arctic explorer Vilhjalmur Stefansson (1879–1962), the collection is an exceptionally rich body of materials for research on the history of both the Antarctic and the Arctic, with most of the Antarctic materials predating World War II. For the Arctic regions, the bulk of the collection covers events that occurred prior to 1930. The bulk of the photographic holdings relates to the Canadian Arctic Expedition of 1913–1918. Additional photographs document other expeditions, ships, events, peoples, flora, fauna, and equipment.

Image Metadata

Descriptive metadata for photographs is inherently complex. A number of systems are designed to assist librarians and archivists describe various aspects of photographic images, such as the Cataloging Cultural Objects (CCO) guidelines, the Art and Architecture Thesaurus (AAT), the Thesaurus for Graphic Materials (TGM), and the Library of Congress Subject Headings. But there are limitations to these controlled vocabularies. For instance, a number of items in the Stefansson Collection depict *Nalukataq*, a spring whaling celebration, yet it does not appear in any commonly used thesaurus or set of controlled vocabulary. In addition, because these resources have not always been applied, and many legacy systems utilized local terms, terminology across collections is inconsistent. Furthermore, while a controlled vocabulary makes for more consistent searching and retrieval, the time it takes to apply it at the item level is often beyond the means of an institution to provide.

RAUNER LIBRARY

[Playing drums and dancing by an up-turned umiak. Originally stored with glass plate Negatives.] Boat, umiak; Canadian Arctic Expedition; Child, Inuit; Individuals, female; Inuit; Individuals, male, Inuit; Recreation, dance, music.

FIGURE 9. **Different archives, different metadata**

ALASKA'S DIGITAL ARCHIVES

Dancing on blanket made of bearded sealskin outside during Nalukataq (spring whaling celebration). Well-dressed women dance with babies on their backs.

The metal bucket near drummers is filled with water used to keep drums moist for better sound. Man seated under right end of umiak, seen between two dancers on left is wearing wooden snow goggles. Mukluks were replaced yearly and unique to each community. Women in summer kuspuks.

Alaska Natives—Northern Alaska—Inupiaq

FIGURE 10. **Comparing the data: Rauner Library versus the crowd**

An example of inconsistent metadata creation can be seen in Figure 9, two examples of the metadata applied to similar images at two different institutions: Alaska's Digital Archives and Dartmouth College's Rauner Library. As one can easily see, the image metadata vary widely between the two collections. Much of the descriptive metadata applied to visual images is either not standardized or represents only a few aspects of the images.

Crowdsourcing image data does not promise to provide more consistency or controlled vocabulary, but it does open up the possibility of a broader range of metadata from widely varying perspectives, thus enhancing the ability of researchers to find images that might not have been accessible before.

All Responses: mountain (2), houses (2), landscape (1), house (1) precarious (1), decay (1), erosion (1), hillside (1), not falling down the hill (1), cliff side housing (1), stilt housing mountain side dirt(1), unstable slope (1), eskimo (1), degrading landscape (1), hill dwellers (1), houses on poles on hill side (1), deconstruction (1), hillside (1), settlements on hill (1), rainbow (1), alaska (1), bering strait (1), cliff dwellers (1), plants (1), window (1), clinging (1), sky (1), rocks (1), cliff (1), stilts (1), shacks (1)

From the Rauner Metadata:
[Hand-colored. Cliff dwellings. See also glass plate negative 207.] The Cliff Dwellers of Bering Sea. Eskimo settlements King Island Alaska Lomen Bros. Nome # 707. Bering Sea; Canadian Arctic Expedition; Cliff Dwellers; Habitation, cliff dwellings, stilt house; Individuals, male; Photographers, Lomen Brothers; King Island

FIGURE 11. Comparing the data: Rauner data and input from the crowd. The numbers next to tags represent the number of times players in the pilot study input the tag.

Pilot Study Data

For the pilot study, the team created two separate installations (one called the "red installs" and the other called "blue installs") of the same 194-image database, each with a unique URL and Arcade interface. The content for the test images depicted Arctic scenes from the Stefansson Collection. For the test, all the existing metadata were stripped from the images to see what players would come up with on their own.

The pilot study incorporated thirty-seven users, engaging participants from an English-speaking international audience in a two-month session. In a hidden, preprogrammed list, 445 "stop words" (from conjunctions to banned words) prevented obviously irrelevant material from entering the database. Players in this pilot study generated thousands of entries. The collected content was tested

against the existing, detailed metadata associated with the images. To gauge the impact of experts on metadata collection, graduate students in Arctic Studies at Dartmouth College were asked to participate in the pilot and their data was tracked against the data provided by the general group.

E x p e r t   P l a y e r   C o m p a r i s o n s

Figures 12 and 13 provide a valuable comparison between three different groups of data: that from players who are already interested in the topic (Arctic Studies players), the collected responses in the test (all players using Metadata Games), and the existing metadata from the images housed at the Rauner Library. Note that the existing metadata does not mention the white flowers or grasses, or use the term *dryas*. The details collected from the expedition notes (such as the presence of anemones at the precise location of Bernard Harbor, or the exact date when the photograph was taken), of course, are invaluable and much more difficult, if not nearly impossible, to crowdsource.

From our Arctic Players: dryas (1), arctic (1), tundra (1), younger dryas (1)

All Responses: flowers (8), field (5), grass (3), white (3), hillside (2), flower (2), groundcover (2), dryas (1), low to the ground (1), pretty flowers (1), rocky hillside (1), low (1), younger dryas (1), arctic (1), rugged landscape (1), tundra (1), beauty (1), land (1), earth (1), white flowers (1), hill (1), grasses (1), wild (1), green (1), alpine flowers (1), moss (1), small white flowers (1), little white flowers growing on rocky hillside (1), hills (1)

From the Rauner Metadata: [Hand-colored. Arctic tundra landscape. See also Glass Plate Negative 238. Duplicate.] 42408 Anemones at Bernard Harbour, N.W.T. July 1916 to Johansen, C.A.E. Flora; Bernard Harbour; Canadian Arctic Expedition; Prairie; Summer; Tundra

F I G U R E 1 2 .   Comparing the data: Arctic Studies experts versus Rauner versus the crowd

From our Arctic Studies Players:
inuit extended family (1), arctic people in traditional clothing in posed setting (1);

All Responses: portrait (5), furs (4), family (3), native american (3), skins (3), fur (2), colorized (2), people with animal heads (1), animal skins (1), cold (1), unusual (1), skin (1), dressed up (1), north (1), hides (1), people (1), alaska (1), interior (1), posing (1), olden days (1), formal (1), portrait (1), arctic people in traditional clothing in posed setting (1), inuit extended family (1), native american group photo traditional clothing (1), landscape (1), somber (1), hoods and boots and gloves made from skin (1), masks (1), artifacts (1), box (1), sit (1), sombre (1), grey (1), old (1), tannery (1), tribal clan (1), seated (1), animal masks (1), full regalia (1), group portrait (1), hand colored (1), dress (1), solemn (1), sepia tint (1), ceremonial photograph (1), animal mask (1), animal heads (1);

From the Rauner Metadata:
Kaylagamutes of Alaska Lomen Bros. Nome #174 The Wolf [crossed out] Eagle Dance; Animal, wolf; Bird, Eagle; Clothing, ceremonial; Ceremony, The Wolf Dance; Fauna; Fur; Habitation, interior; Individual, female, Inuit; Individual, male, Inuit; Kaylagamutes; Pelt; Photographer, Lomen Brothers; See also, Large format photographs in Stefansson collection for this hand-colored image.

F I G U R E 1 3 .   Comparing the data: Arctic Studies expert input compared with Rauner's data and the data generated by the crowd

Figure 13 makes an additional comparison between Metadata Games players and the existing metadata from the Rauner listing. The existing metadata does not document that the photo depicts an extended family, that the photo is posed, or that the subjects are wearing masks, nor does it mention that the photo was taken in the summer.

F i n d i n g s

In analyzing our pilot study data, we have been able to generalize our findings into five themes and recommendations.

1. **Games can produce more entries per person in crowdsourcing than nongame systems.** The number of entries in the pilot test, and the quality of these entries, suggest great promise for the use of Metadata Games as a solution for gathering metadata. The thirty-seven user accounts were split between two installations of the software. Each account generated a specific number of entries (see Table 1). A game approach, as compared to the Library of Congress Flickr initiative, generated many more tags per person (see Table 2).

Table 1. Number of Contributions from Each Account During the Pilot Study

Blue Installation	Red Installation
1,350 unique tags	1,382 unique tags
3,224 entries overall	3,026 entries overall

Table 2. Average Number of Contributions from a Metadata Games Player Compared to the Library of Congress Flickr Project

Library of Congress Flickr Initiative	Tiltfactor Metadata Games Pilot Study
4,000 images	194 images
67,176 tags	6,250 tags
2,518 unique users	37 unique users
17–18 tags per image	32–33 tags per image
Average: .006 tags per person	**Average:** .84 tags per person

2. **Customizable systems meet more diverse needs.** The IMLS DCC Flickr feasibility study notes that rights concerns are the greatest barrier to

participation.[23] Metadata Games can be deployed locally for those collections with limited rights holdings, such as for use within a library, and/or deployed internationally for those collections whose rights are more liberal. Uploading sets of images in the Metadata Games system is easy to do, and access to the system is easy to manage in-house.

3. **Game interfaces significantly influence the data.** The types of data received varied quite broadly depending on the types of data prompts, game interfaces, and designed play experiences created for the players. The kinds of prompts we provide for players to answer will dictate the data that users provide.

4. **Game play and subject matter preferences both dictate player participation.** As expected, game play enjoyment and engagement varied from person to person. The game preferences revealed by players in the study were surprising. Some players refused to play the more open-ended games, such as Zen Tag (e.g., a seventy-year-old great-grandmother desired more competition and rewards), while others, including a ten-year-old boy, found the open-ended gamelike activities highly addictive. The team was also surprised that players were mostly unconcerned about the data they were generating—how it was used, whether it could be tracked and tied to their identity—and posed few questions about the results, such as how many tags they generated, which player tagged what image with what word(s), how the crowdsourced data differed from existing metadata, and so on. Players were very concerned, however, about clarity of play, the reward system, competing and collaborating with friends, and the overall feel of the gameplay experience.

5. **Curiosity and doubt are key design opportunities.** A player's own curiosity and doubt while using the system represent two interesting domains for future research that have thus far not received much attention in crowdsourcing literature. In a number of instances, players became so curious about the images they were tagging that they would tag images with inquiry phrases, such as "want to know more about this culture." This shows the power of linking games to our more traditional heritage holdings and suggests a powerful new way to involve fresh audiences and increase their discovery and use of unique collections. The team also monitored when players expressed uncertainty in games. The team found several instances in which players, unsure of the correct answer, continued to participate enthusiastically and wished to contribute even though their entries were only guesses. For

[23] Jacob Jett, Carole L. Palmer, Katrina Fenlon, and Zoe Chao, "Extending the Reach of Our Collective Cultural Heritage: The IMLS DCC Flickr Feasibility Study," *Proceedings of the American Society for Information Science and Technology* 47 (November/December 2010): 1–2.

F I G U R E   1 4 .   **Yak, water buffalo, or musk oxen?**

example, the image in Figure 14 was tagged as "possibly water buffalo" (in fact, the image depicts musk oxen). If players were able to mark uncertainty and certainty in their input, we might be able to gather more accurate responses.

6. **Trust is complex but must be designed for within such a system.** During the course of the design, the team created various avenues by which players could earn trusted status. In parsing "trust," we had to define several variations on what the word means in the system. Tapping into expertise and player motivation are priorities in developing trust and will be an important component of future work, covered in the next section.

Current and Future Work

Many institutions and groups are interested in employing crowdsourcing games to gather translations, transcriptions, or metadata. The National Library of Finland, for example, has recently launched a new program to support efforts to digitize its archives. Current work on the Metadata Games project includes implementing trust algorithms to discover who is a trusted player. We are now creating an Iterative Trust Algorithm for monitoring and rewarding knowledge

network game behavior.[24] In addition, we continue to incorporate elements that address player motivation, including overall game rewards, such as sitewide achievements and treasure hunts within games.

Expertise remains an elusive concept in crowdsourcing projects. In certain circumstances, for example, hobbyists have more expertise than do information scientists. In the future, the system will provide a careful measure of expertise. The Metadata Games system will offer ongoing minigames that ask experts to rank valid answers. Determining a metric for trust and expertise, and solid metrics for other elusive concepts such as persistence, are next on our agenda. We also plan to increase the capacity of the system to allow for tagging of sound and video archives and to develop new games for these media.

Conclusion: Critical and Theoretical Issues

While our team acknowledges that this paper primarily focuses on the technical and practical concerns of Metadata Games, we also emphasize that the system, even at an early stage of development, evokes (and is informed by) critical and theoretical questions concerning collections, data, and design. We ask fundamental questions in our project about how to foster a curiosity about the humanities through Metadata Games, how to motivate players to delve deeper into subjects, and how to discover the types of knowledge that can be crowdsourced. Sarah Farmer notes that trends in historical research and publications over the course of the last ten years, for example, indicate a "visual turn" in historical scholarship.[25] Books such as Apel and Smith's *Lynching Photographs* respond to this fresh focus in humanities scholarship and prioritize visual evidence as a way to trace the violent activities of the past.[26] Visual evidence—what is depicted (shape, form, color, perspective, point of view), how what is depicted feels (emotion, mood), when it might have occurred (time

[24] Michael G. Noll, Ching-man Au Yeung, Nicholas Gibbins, Christoph Meinel, and Nigel Shadbolt, "Telling Experts from Spammers: Expertise Ranking in Folksonomies," in *SIGIR '09: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York: ACM, 2009): 612–19; Wai-Tat Fu and Wei Dong, "Facilitating Knowledge Exploration in Folksonomies: Expertise Ranking by Link and Semantic Structures," in *Proceedings of the 2010 International Conference on Computational Science and Engineering*, (Minneapolis: Minn., 2010); Ruogu Kang and Wai-Tat Fu, "Exploratory Information Search by Domain Experts and Novices," in *IUI '10: Proceedings of the 15th International Conference on Intelligent User Interfaces* (New York: ACM, 2010).

[25] Sarah Farmer, "Going Visual: Holocaust Representation and Historical Method, *American Historical Review* 115 (February 2010): 115–22. Farmer cites a variety of anthologies and readers that are defining "visual culture" scholarship, including Marita Sturken and Lisa Cartwright, *Practices of Looking: An Introduction to Visual Culture* (Oxford: Oxford University Press, 2001); Malcolm Barnard, *Approaches to Understanding Visual Culture* (New York: Palgrave Macmillan, 2001); Richard Howells, *Visual Culture* (Malden, MA: Blackwell Publishers, 2003); Vanessa R. Schwartz and Jeannene M. Przyblyski, *The Nineteenth-Century Visual Culture Reader* (London: Routledge, 2004).

[26] Dora Apel and Shawn Michelle Smith, *Lynching Photographs* (Berkeley and Los Angeles: University of California Press, 2007)

frame)—has become essential to the archival record. Metadata Games can enable and advance this scholarship by uncovering previously unknown information about visual evidence through a game system and increasing access to an ever-expanding number of visual collections.

Clearly, as digital technologies proliferate, the volume of data available outstrips scholarly research methods. The goals of this research are many, encompassing both practical and conceptual concerns. First, and most obvious, is the hypothesis that crowdsourcing may help institutions faced with dwindling budgets address resource constraints by involving interested participants in the process of contributing metadata. After all, participants already willingly feed systems such as Facebook and Twitter with their personal information, providing free marketing and profiling data. If the experience engages participants and they value it, the "labor" involved in the exchange can be considered a voluntary, in-kind contribution.

Second, the wisdom of crowds may help new vocabularies and new ways of categorizing, understanding, and linking materials to evolve in intuitive ways, augmenting the ways in which archives are currently described. The use of crowdsourcing also points to a number of significant epistemological and ontological questions. In archives, notes from the collection are often used to document an image carefully and accurately. Images are understood by their context: by the notes and various details surrounding the collection, and by the content and disciplinary expertise of the archivist. In crowdsourcing, images are understood by their visual evidence and the associations that evidence creates and evokes, and are only mapped to traditional indexing terminology after the public has input its terms.

Third, the primary concerns expressed by librarians and archivists about the project—namely, the practical and epistemological questions about "what are valid data to collect"—come to the forefront. Corporations collect a substantial amount of information to profile their participants and users. Thus far, creating *knowledge profiles*—tracking how someone goes through information to learn, or how one follows one's interests—has remained outside the realm of humanities and information science researchers for ethical, moral, and technical reasons. There are compelling reasons, however, to create *learner profiles* and to collect this information ethically and confidentially to foster new insights on learning in the humanities.

Fourth, information is being steadily lost. The new adage, "If it is not on Google, it is lost," is not far from the truth. New generations use networked technologies at an increasing rate, and digital materials are truly made available to new generations in a way that their everyday practices demand. This also means that we may inspire new generations to seek and use archival materials by introducing them to such materials through games. Our team speculates that

engagement with primary source documents, such as archival photographs, coupled with the use of games as a motivating factor, may kindle interest among users to seek out archives and historical materials. Libraries face an unprecedented level of competition for the attention of an online community that has a seemingly limitless choice of where to invest its time and interest. Perhaps making the materials available through games could lead to an increased interest in what could be perceived by some, unfortunately, as an irrelevant resource.

Finally, Metadata Games can be used to enhance knowledge about archival images associated with particular disciplines and fields, or with interdisciplinary collections. The software can serve as a model for the way that scholars, students, and general audiences can participate together in twenty-first-century cultural production. Crowdsourcing metadata opens up the possibility of a broader range of input from widely varying perspectives. This diversification enhances the ability of researchers to find—and contribute knowledge about—images that might not have been accessible before, and it increases the extent to which egalitarianism and inclusivity become integrated into the design of research technology.