Citizen Archivists at Play: Game Design for Gathering Metadata for Cultural Heritage Institutions

Mary Flanagan, Sukdith Punjasthitkul, Max Seidman, Geoff Kaufman
Tiltfactor Laboratory, Dartmouth College
HB 6194, 246 Black Family Visual Arts Center
Hanover, NH 03755
603-646-1007
mary.flanagan@dartmouth.edu; sukdith.punjasthitkul@dartmouth.edu; max.seidman@dartmouth.edu; geoff.kaufman@dartmouth.edu

Peter Carini
Dartmouth College
Rauner Special Collections Library, 6065 Webster Hall
Hanover, NH 03755
603-646-3728
peter.carini@dartmouth.edu

ABSTRACT
In this paper, we detail our design process for the Metadata Games project and discuss a number of design challenges involved in making a "metadata game," such as incentivizing players to offer accurate information, devising and deploying methods for verifying the accuracy of data, and introducing effective motivations for ensuring high replay potential. We present our "Outlier Design" model for creating effective crowdsourcing applications, and offer the Metadata Games prototype *One-Up* as an example. This game's design addresses the challenges of gathering increasingly higher quality metadata while creating a compelling play experience.

Keywords
digital humanities, humanities, metadata, games, design, crowdsourcing, motivation, archives, museums, cultural heritage

INTRODUCTION
Across universities, archives, libraries, and museum collections, millions of photographs, audio recordings, and films lie waiting to be digitized. This is because scanning is only half the battle: while digitization can now be carried out fairly easily, in many cases in bulk, there are few resources to document this material properly. Scanned artifacts are thus often added to collections with a bare minimum of descriptive metadata, the informative tags that provide details regarding an item's content, context, and creation. Metadata includes general descriptors (e.g., information about the subject, setting, theme, etc., of an image, such as "Greenland" or "Pop Concert") as well as specific details (e.g.,

Proceedings of DiGRA 2013: DeFragging Game Studies.

© 2013 Authors & Digital Games Research Association DiGRA. Personal and educational classroom use of this paper is allowed, commercial use requires specific permission from the author.

"Bar Harbor Buoy Bell" or "Justin Bieber") that are attached to an image or other artifact. Although metadata can be constituted from shared standards (such as Library of Congress Subject Headings) or culled from a donor's own descriptive materials about the object, the reality is that an adequately sized professional staff with sufficient time and knowledge is needed to identify archival images and tag them properly—a precious commodity in an era of shrinking public support for heritage institutions. As a result, an incredible number of collections unfortunately have little identifying information other than what might be written on a photo or on a box of material. In the absence of accurate, thorough metadata, the valuable items in archives' collections become difficult or impossible to retrieve—and, for all intents and purposes, inaccessible to scholars, researchers, and members of the general public.

To help stem this growing tide and save cultural artifacts worldwide from falling into digital oblivion, our team has created Metadata Games, an open-source, customizable software system that uses computer games to collect information about archival media through crowdsourcing. Prior work has illustrated the project's early efficacy (Flanagan and Carini 2012). In this paper, we detail our design process for the making an effective metadata game. We present our "Outlier Design" model, which guided the development of the game *One-Up,* and demonstrate how our design team approaches the challenges of incentivizing players to offer accurate information, devising and deploying methods for verifying the accuracy of data, and introducing effective motivations for ensuring high replay potential. We include results from playtests with *One-Up* that indicate that the game will likely yield high quality metadata for crowdsourcing applications.

The Potential of Crowdsourcing and Games for Knowledge Acquisition

Crowdsourcing has been used in a number of ways in the past decade, exploiting the increased availability of networked technologies—and the enthusiasm of their users—to acquire new knowledge and solicit solutions to specific problems. One of the most dramatic and well-publicized illustrations of the problem-solving capabilities of crowdsourcing is the case of *Foldit*, an online puzzle game in which players aim to fold proteins (University of Washington 2008). In 2012, *Foldit* players helped solve a 15-year-old mystery by deciphering the structure of an AIDS-causing virus—and they did so in a matter of days (Khatib et al. 2011). The power of crowdsourcing has likewise been channeled to the collection of metadata. For example, the New York Public Library's "What's on the Menu?" project invited users to assist in the transcription of historical restaurant menus (New York Public Library 2011), and the Library of Congress's Flickr pilot project enlisted the assistance of users to tag hundreds of images from the library's extensive collection (Library of Congress 2008). The Australian Newspaper Initiative from the National Library of Australia asked users to search and correct digitized newspaper articles dating from 1803; in a little over a year after the service was released to the public, 7 million lines of text in 318,000 articles were corrected (Holley 2010). A wealth of related citizen science projects, such as Galaxy Zoo (Citizen Science Alliance 2007), complement this humanities-focused work and further the field.

Metadata crowdsourcing efforts have begun to incorporate the use of games, such as Google's Image Labeler game (Google 2006) and the Finnish government's experimental translation program, *DigiTalkoot*, in which players assist in detecting and correcting textual errors in digitized archival materials (National Library of Finland and Microtask 2011). To date, the most influential work in the crowdsourcing game space has been the pioneering research of Luis von Ahn, who has shown that a mere five minutes of play per day can generate thousands of tags from a much smaller group of participants than traditional methods of data collection typically require (von Ahn 2004). Von Ahn demonstrated this approach in his “ESP Game” (which later became Google Image Labeler), an online game in which randomly paired players are presented with the same image and must try to “guess” what the other player might be typing about it. If one player correctly guesses what the other player types while avoiding a “taboo” word list, both players receive points and are presented with another image (von Ahn 2004). Because the players are randomly paired, and are not permitted to communicate, the assumption is that tags are probably accurate if there is mutual agreement between players. In this paper, we describe this method as peer-to-peer verification.

What motivates users to participate in crowdsourcing efforts, and what unique contributions might games offer? There tend to be three core, potentially overlapping sources of motivation among crowdsourcing game players. One aspect of player motivation is to assist a particular institution, or simply contribute to a good cause. These engaged enthusiasts first and foremost like the idea of helping (Owens 2012). A second motivation for players is the love of a subject area—for these citizen archivists, the game’s content aligns with their personal interests or areas of expertise, such as players who intrinsically enjoy tagging photographs depicting architectural structures or dogs from various breeds. A third player motivation is to compete— to win by being the best, the fastest, and most accurate. While the first two motivations are satisfied by virtually any metadata application, the third is unique to games. The game element is important, as the draw of a game could attract individuals who wouldn't necessarily need to be intrinsically altruistic or interested in a particular subject matter to be enticed to participate.

As the above examples illustrate, crowdsourcing can be a viable strategy to help heritage institutions gather information about the items in their collections, and a particularly alluring option given that most institutions lack the personnel and resources to tag archival and library collections adequately.

An Overview of the Metadata Games Project

Inspired by crowdsourcing work, our design team endeavored to create an innovative suite of games that could quickly gather valuable tags, offer an enjoyable experience for the player, be freely available and accessible to nonprofit institutions, and reward higher levels of accuracy and participation. Our team aims to gather tags for a range of media types (images, video, and sound) as well as foster new play mechanics and styles (single versus multiplayer, cooperative versus competitive, and more). Funded by the US National Endowment for the Humanities, Metadata Games is designed to be a free, open-source, customizable software package that would be available to a wide range of cultural heritage institutions without expensive licensing fees or contracts. The games included in the suite are data gathering “portals” that collect individual tags and store them in a larger database that the institution hosts (or one to which the institution has easy access). Our team’s goals with Metadata Games are to help cultural heritage institutions gain useful data for their collections, assist scholars in learning how to interact and utilize collections in new and possibly unexpected ways, and provide a host of opportunities for the public to interact with cultural heritage institutions.

Our novel design approach employs unique game play mechanics, dynamics, and reward schemes that aim to solicit accurate, specific contributions from players and ensure high levels of player investment and engagement. By employing novel designs and giving players agency in selecting subject matter, Metadata Games simultaneously attracts avid gamers with the promise of a compelling play experience and experts with the promise of intrinsically rewarding content. At the same time, the aim is that even inexperienced gamers will find the gameplay dynamics immediately accessible and enjoyable, and that non-expert players will feel compelled to engage with (and possibly even learn more about) the content the games present.

To illustrate our general design approach and the strategies we employed to satisfy the goals outlined above—namely, to generate accurate, specific metadata and inspire high player engagement and investment—we will focus on *One-Up*, a competitive two-player mobile game that has been iteratively designed and tested at Tiltfactor.

SECTION 1: THE DESIGN OF *ONE-UP*

One-Up

One-Up is a multi-round mobile app game in which players score points for submitting single-word tags and try to accrue more points than their opponent (see Figure 1). In this game, a player can challenge either a friend or a random player.

Figure 1. Screenshots of *One-Up* game prototype.

At the start of **Round 1**, two players are simultaneously presented with the same image on their respective devices. Each player individually submits three tags that describe the image (e.g., for the image in Figure 1, players might enter words such as "tree" or "bloom"). In this round, players are awarded 1 point for each tag they enter. At the same time, players have the opportunity to earn an "accuracy bonus," additional points awarded if a tag that a player submits matches one submitted by players in previous games, which are stored in a database. Once both players have submitted their 3 words, Round 1 ends and Round 2 begins.

In **Round 2**, gameplay is similar to the first round, with two important exceptions. First, the accuracy bonus increases. Second, if players submit a tag that their opponent previously submitted in the first round, their opponent captures a point from them (in which case, players are informed that they have been "One-Upped"). That is, players are penalized one point and their opponent receives an additional point.

In **Round 3**, the accuracy bonus once again increases, but so too does the penalty for submitting a tag one's opponent already submitted in Round 1 or 2.

After all of the rounds are played, the player with the highest score is declared the winner.

Design Implications

In using a multi-player, competitive format for gathering metadata, a design challenge arises: how to prevent or dissuade players from subverting the system in order to earn the most points and intentionally contribute unwanted data. Ways to prevent game subversion include randomizing player pairing and anonymizing player names. While effective, these tricks make it very difficult to play directly with friends and minimize the social gaming experience.

Unlike in peer-to-peer verification models, which permit friends to cheat to both maximize points and contribute unwanted data, in *One-Up* players can try to subvert the system by focusing on earning the highest number of points, or they can intentionally try to contribute unwanted data, but not both. This uniquely positions *One-Up* to be able to tap into the motivation of competition among friends, but at the same time to retain the ability to gather useful metadata: the design reduces friend-players' tendency to collude on answers (to generate points by "gaming the system") or by entering "junk" or inappropriate tags. Collusion is a particular challenge for crowdsourcing games, and, to our knowledge, *One-Up* is among the first to address this issue directly through its design.

One-Up structures player rewards such that, instead of rewarding a player for matching another's entry, it rewards a player for matching the tags of previous players in the database. In this way, the game still aligns player incentives with the collection of accurate data.

Designers often focus on specifying the reward and the risk in their games. For example, in *Pac-Man* the reward is gaining points, at the risk of getting killed by one of the ghosts roving around the level. In *World of Warcraft*, the reward for taking on a challenging quest is gaining valuable loot, at the risk of getting one's character killed. In many games involving metadata, there is little risk involved; taboo word lists remove risk because the player knows what tags to avoid entering. But what if instead of banning obvious words in order to collect specific data, players are allowed to submit obvious words, but punished if they match an opponent's entered terms? *One-Up* utilizes such a method and, by doing so, creates a tension between the desire to submit a tag for a particular image that is likely to get an accuracy bonus, and the risk of getting points stolen for typing a word an opponent has already given in an earlier round. *One-Up* incentivizes players to avoid exclusively entering obvious tags. Both players have to try to guess what strategy the other is using. One key play strategy to resolve the tension is for a player to enter some obvious terms (to try for an accuracy bonus) but also to make sure to include more specific tags as well. This tension guides players to enter obvious terms early on in the game, and save the accurate and specific tags for later rounds (to avoid getting "One-Upped").

SECTION 2: ALIGNING PLAYER INCENTIVES WITH THE SUBMISSION OF HIGH-QUALITY METADATA

Many games employ a number of techniques to incentivize desired behaviors in players. Crowdsourcing games often utilize a points reward system; that is, the behaviors we reward with points are the encouraged behaviors. Players are less likely to submit correct metadata unless it is incentivized with rewards.

High-quality metadata is data that is accurate and specific (see Figure 2).

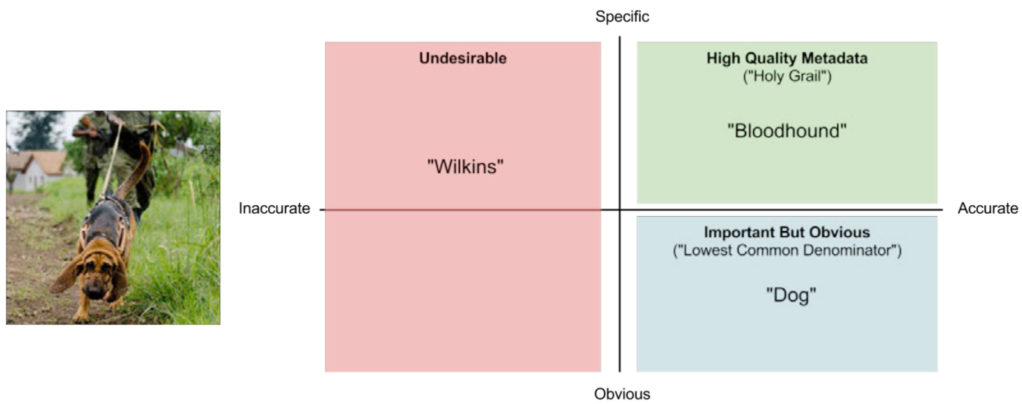

Figure 2. From left to right: As we collect data about an image, the goal is to move from inaccurate to accurate, and from generic (or obvious) to specific tags. We want to avoid collecting inaccurate tags. Example tags are shown in each sector for the displayed picture of a bloodhound.

Generic or obvious but accurate tags are necessary, and they are also fairly easy to collect. Such tags help us classify the artifacts and can serve as fodder for the creation of basic topic categories and other sorting mechanisms. Even the most rudimentary game or tagging system can foster player input such as "dog" for a picture of a bloodhound. The real "Holy Grail" in crowdsourcing, however, is the acquisition of specific, accurate information.

As the following discussion illustrates, while there are particular incentive schemes that metadata games could employ to reward players based on either the specificity or accuracy of the tags they submit, with *One-Up* our team devised a unique design approach that incentivizes players to provide metadata that is both specific *and* accurate.

Method 1: Rewarding Players Based on Specificity

How does the system identify correct answers so that it can reward players for submitting specific tags? Players of crowdsourcing games might submit many types of words to be used as tags. After they are entered, natural language processing (NLP) techniques are common verification tools that can determine whether player-submitted metadata is specific. Simple syntactic accuracy checking, such as spelling matches in *SpellTower* and *Words With Friends*, serves as a bare minimum for checking whether a word is actually a word. NLP techniques, however, can be further employed to determine the specificity of submitted terms, for example by using inside-outside information in what is called "term hierarchy" generation (Ryu and Choi 2006). Using inside information (the characteristics of component words or the internal structure of terms) and outside information (contextual terms), one can build hierarchical relationships among terms and measure the quantity of domain/subject specific information contained in a given term. In another approach, we can reduce the clutter of unimportant words and cull "keyphrases" (Chuang, Manning, and Heer 2012). These approaches incentivize players to provide highly specific metadata: when a tag is determined by NLP to be specific, players could be rewarded with bonus points.

Unfortunately, NLP techniques cannot easily determine the semantic accuracy of a particular tag. No matter how closely related a user-submitted word is to a base word, or how scientific a tag may be, the system cannot verify how well the tag describes the image. This is because NLP software can search databases of word specificity ratings to be able to determine that "bloodhound" is more specific than "dog," but is not capable of determining whether either is an accurate tag for a particular image. Thus, rewarding a player for submitting a brand new tag with a high specificity rating is not properly aligning incentives with collecting accurate and specific data. This is a weakness of this method.

To conclude, using computational language techniques as core tools in metadata crowdsourcing may help verify specificity and accuracy—indeed, crowdsourcing tags is a process that requires some NLP to, for example, exclude purely garbage tags. It is important to remember, however, that even at its best this technique merely gauges specificity and accuracy, but does not help increase the accuracy of players' submissions.

Method 2: Rewarding Players Based on Accuracy

Peer-to-peer methods for verifying the accuracy of submitted data are used extensively across crowdsourcing applications—indeed, they provide the foundation on which these applications rest. Using peer-to-peer verification in a metadata games context, if multiple players are shown the same image, and they both submit the same tag within a short period of time, it can be assumed that the tag is accurate—provided the players cannot communicate with each other and conspire to "cheat" by entering identical, inaccurate tags.

When players enter matching tags for a particular image, they are typically rewarded with points and allowed to move on to a new level. This is an excellent match of aligning player incentives with the crowdsourcing goal of collecting highly accurate metadata—each matched tag reinforces the veracity of the other tag (von Ahn 2004).

Unfortunately, a reward scheme that is based on peer-to-peer verification also incentivizes players to submit the most obvious tags to describe a given image. For example, players are much more likely to attempt to match a game partner's tag of "dog" than a more detailed tag of "bloodhound." The *ESP* game attempts to circumvent this problem by having a "taboo" list that includes the most frequently entered tags from previous play sessions. Instead of inspiring players to then enter a tag of "bloodhound," however, this overused tag ban only incentivizes players to submit the next most obvious tag, which might be something like "grass," or to give up on tagging the image entirely. This method disincentivizes players from entering more specific metadata.

A reward scheme based on peer-to-peer verification can be extremely effective for collecting obvious tags, but a prohibitively long "taboo" list would be needed to collect very specific tags. Nuanced and expert tags are the last to be cultivated from crowdsourcing using these methods.

Thus, a peer-to-peer verification method might help determine the accuracy of submitted metadata (and reward based on it), but its flaw is that while it can get specific data, it incentivizes players to supply less specific data in the service of matching their partner's input.

Method 3: Rewarding Players Based on the Outlier Design Model

A new strategy is to create a system that attracts specific and accurate tags—tags that, in other systems, might be either interpreted as "junk" tags or as "outliers." It is far more challenging for the system to differentiate between accurate specialty tags, the "outliers" (see Figure 3), which are less frequently used, from "junk," or inaccurate entries.

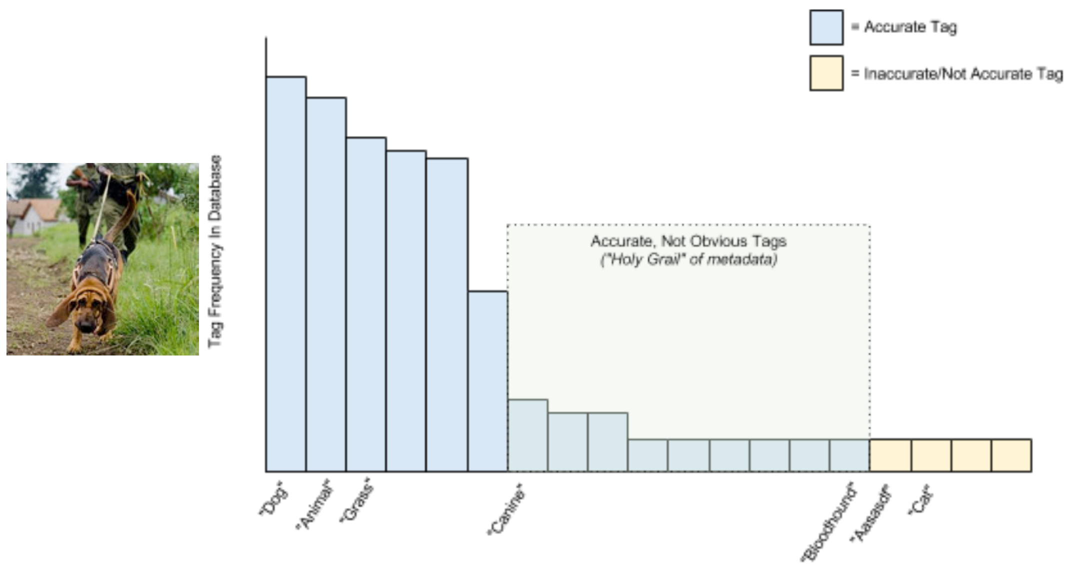

Figure 3. Standard metadata crowdsourcing methods result in datasets like this one. Obvious tags have a high frequency of appearance, and are thus easier for the system to recognize as accurate. More specific tags appear with lower frequency, and are more challenging for the system to distinguish from inaccurate tags.

So how might designers incentivize players to input more high quality tags—that is, more "outlier" tags in the "Holy Grail" region (see Figure 4)? One possible approach is a two-player design, which rewards players for inputting tags that are already in the database, and thus likely accurate, but also penalizes them for matching tags their opponent has submitted in previous game rounds. This design, which is employed in *One-Up*, adds a sense of uncertainty and excitement for the player in two ways: first, players perceive risk in the game because they don't know what the other player will do, and second, there is suspense because players aren't being told explicitly what tags to avoid.

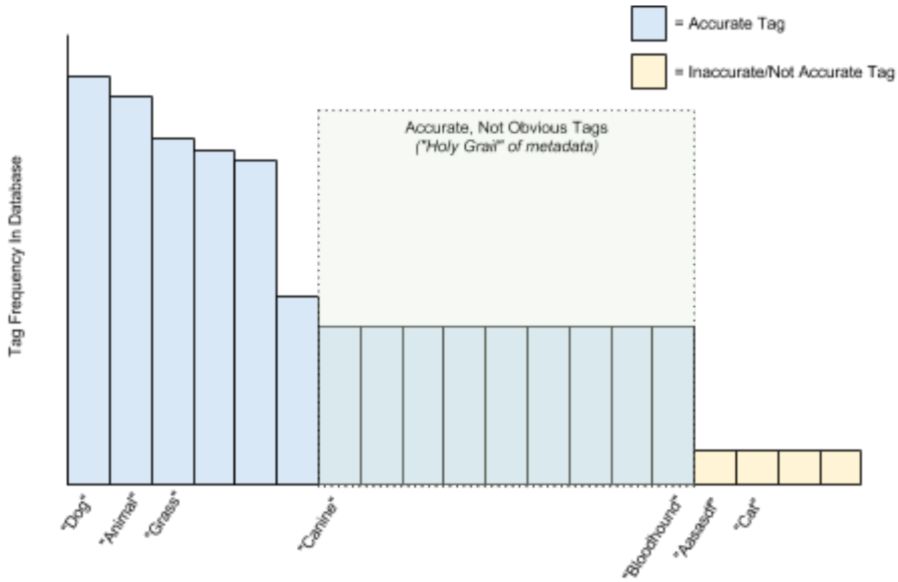

Figure 4. The goal is to incentivize players to input tags in the "Holy Grail" zone, such that the frequency of these specific and accurate tags rises above the frequency of the inaccurate tags, thus allowing the system to determine their accuracy (shown above).

The Outlier Design model requires players to balance their desire to enter terms that are likely already in the database with the need to strategize in order to avoid entering tags already entered by an opponent. Thus, the design naturally shepherds the players towards providing more accurate and more specific answers (see Figure 5).

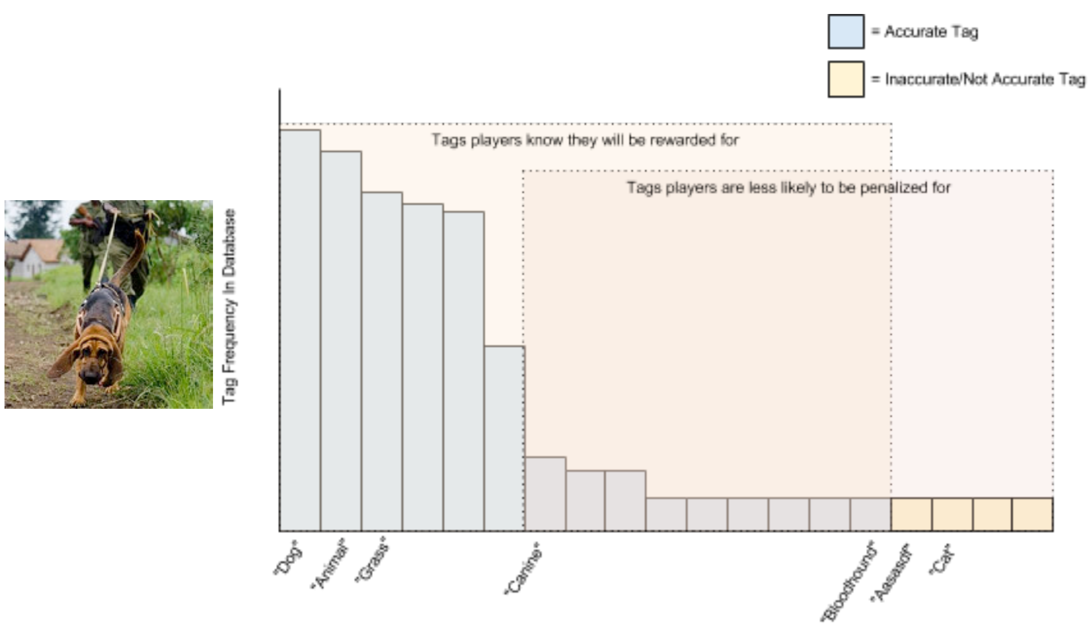

Figure 5. Using this method, players avoid inputting all the obvious tags, while also wanting to avoid putting in the inaccurate ones. Together, these constraints incentivize players to submit tags in the "Holy Grail" zone.

In the Outlier Design model, the goal is to disincentivize the entry of a range of obvious tags without explicitly forbidding them. Because obvious tags aren't forbidden, players can enter them if they aren't capable of discerning more specific details from an image. Yet generic tags only go so far in helping understand an image or piece of media. Indeed, one issue designers must struggle with is that the difficulty of obtaining high level, "Holy Grail" tags can be attributable to a lack of knowledge (e.g., knowledge of a specific species of plant or breed of animal) rather than a lack of motivation or incentive. It is likely that the "Holy Grail" tags are rare in large part because fewer members of the "crowd" have this specific knowledge. The strategy behind this design is that if players are incentivized to provide more specific tags, then they should be motivated to find a way to provide them, even if it means doing their own independent research or finding friends to help fill in gaps in their knowledge. The Outlier Design model helps increase accuracy and specificity within the system.

SECTION 3: A STUDY OF THE RESULTS OF *ONE-UP*

As a test of the efficacy of the Outlier Design model for generating specific, accurate metadata, a prototype version of *One-Up* was developed and playtested with adult participants. *One-Up* is intended to be a two-player, asynchronous, turn-based mobile game. For testing purposes, the prototype was not networked, so asynchronous gameplay was simulated using a "play and pass" model: Player 1 plays a turn, then passes the mobile device to another player (who, in our testing, was located in a different room). *One-Up* was developed with the to be played between friends, so players in the *One-Up* test were paired with another player they knew. To assess this game, players saw one of four images (shown in Figure 6) in either *One-Up*, or a game called *Zen Tag*, a simple one-player image tagging game. In *Zen Tag*, a player earns points for each tag submitted, with bonus points awarded for tags that are not already in the database. As past work has noted *Zen Tag's* efficacy in data collection (Flanagan and Carini 2012), we set out to investigate whether One-Up's competitive and social gameplay could have an effect on the quality of submitted metadata.

Figure 6. Images used for testing (clockwise from the top left: Car, Dog, Tree, Mask)

Metadata Playtest Results

Tag data collected from *One-Up* and *Zen Tag* playtests were rated on a scale from 0 to 2 for their accuracy and 0 to 2 for their specificity. Three independent raters scored the tags and exhibited 80% overall agreement for accuracy scores and 75% agreement for specificity scores; we then took the average of their ratings (see Figure 7).

Image	Avg Accuracy Rating (Zen Tag)	Avg Accuracy Rating (One-Up)	Avg Specificity Rating (Zen Tag)	Avg Specificity Rating (One-Up)
Car	1.31	1.82	1.53	1.39
Dog	1.14	1.69	0.91	1.25
Mask	1.15	1.36	1.47	1.67
Tree	1.02	1.24	1.30	1.41

Figure 7. Average accuracy and specificity scores for images tagged in games *Zen Tag* and *One-Up*. Accuracy and specificity were rated on a 0 - 2 scale (0=not accurate, 2=extremely accurate; 0=not specific, 2=extremely specific).

The average accuracy scores for all four images were higher for the tags generated in the *One-Up* playtests than those generated by players of *Zen Tag*. With the exception of the "car" image, the average tag specificity ratings were higher for the *One-Up* prototype than for *Zen Tag*. One possible explanation for the "car" image's lower specificity score from the *One-Up* prototype as compared to *Zen Tag* is the limit in *One-Up* on the number of words a player can submit per round and the total number of rounds per game. It is possible that players were not able to continue to enter more specific tags because there were not enough rounds to do so. Overall, *One-Up* yielded higher quality metadata.

Observations

In addition to scoring the gathered tags for their level of quality, observations of players' decisions and reactions during gameplay offered indications of the success of *One-Up*'s Outlier Design model. For example, *One-Up* generated the tag "Pike" for the "car" image, referring to Pike's Place Market in Seattle. In contrast, *Zen Tag*, utilizing the identical image, did not generate this tag; players identified only obvious characteristics (e.g., "market") in their submissions. While not a full-scale study, these results offer preliminary support for the view that *One-Up*'s Outlier Design model elicits more accurate and more specific metadata than that collected by the free-association tagging game model employed by *Zen Tag*.

In another playtest of *One-Up*, a player tagged the "tree" image using "pollination," and was rewarded with an accuracy bonus, as "pollination" was a term in the database. She assumed that a previous player was a botany expert (as she was, but her opponent was not), and so in the following round she entered "Angiosperm," the scientific name for flowering plants. Angiosperm was not in the database, but the Outlier Design model of *One-Up* leveraged the player's uncertainty to input an accurate and extremely specific tag.

It is important from both a gameplay and crowdsourcing perspective that the games provide fun, engaging experiences so that players are motivated to return to play again and again. *One-Up* fosters a new type of player engagement in crowdsourcing applications through competitive gameplay that generates more specific or specialized terms. While *Zen Tag* pleased players as a point accumulator, players of *One-Up* relish in their winning, and report dismay when their points are "One-Upped" by an opponent. The example of the “botany expert” illustrates that the success may likely mean the melding of compelling gameplay with subject matter interest. *One-Up*'s design worked as intended to engage players with competitive gameplay while incentivizing players to input excellent tags.

CONCLUSION

In our research at Tiltfactor, we have found that the more overtly “educational” one tries to make a game, the less effective such a game might turn out to be. This seems counterintuitive, but from a psychological perspective, some “distance” allows players to connect more with the content. This is what makes our approach unique—we “tilt” the content of serious issues to make them more accessible, and can prove in controlled experiments that this is effective. While some players wish to play games for altruistic reasons to help institutions of their choice, others want the real excitement a good game can offer.

In this paper, we’ve discussed the challenges and strategies of obtaining high quality metadata—tags that are both accurate and specific. We presented our Outlier Design model for crowdsourcing applications. With the example game *One-Up*, we have shown how good game design can align player incentives with the acquisition of high quality metadata, while also creating a compelling play experience.

Metadata Games evokes (and is informed by) critical and theoretical questions concerning collections, data, and design. In our project we endeavor to discover how games can foster a curiosity about the humanities, motivate players to delve deeper into subjects, and diversify the types of knowledge that can be crowdsourced. Games also offer great promise for humanities and archival scholarship by engaging with a broader cross-section of players, ranging in interests and areas of expertise, that other platforms may miss. In addition to gathering valuable metadata, games also offer opportunities to physically draw audiences to cultural heritage institutions.

BIBLIOGRAPHY

Center for Game Science at University of Washington (2008) Foldit [PC, Mac, Linux Computer Game] University of Washington, Seattle, WA USA.

Chuang, J., Manning, C.D., and Heer, J. “‘Without the clutter of unimportant words’: Descriptive keyphrases for text visualization.” ACM Transactions on Computer-Human Interaction, vol. 19, no. 3 (October 2012), pp. 1-29.

Citizen Science Alliance (2007) Galaxy Zoo [Online Web Browser Project] Citizen Science Alliance.

Flanagan, M. and Carini, P. “How games can help us access and understand cultural artifacts,” in American Archivist vol. 75, no. 2 (December 2012), pp. 514-537.

Google (2006) Google Image Labeler [Online Web Game] Google, Mountain View, CA USA.

Holley, R. “Crowdsourcing: How and why should libraries do it?” in DLIB Magazine, vol. 16, nos. 3/4 (March/April 2010).

Khatib, F., Dimaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., et al. "Crystal structure of a monomeric retroviral protease solved by protein folding game players," in Nature Structural & Molecular Biology vol. 18, no. 10 (September 2011), pp. 1175-1177.

Library of Congress (2008) Flickr pilot project. [Online Web Browser Project] Library of Congress, Washington DC, USA.

National Library of Finland and Microtask (2011) DigiTalkoot [Online Web Browser Game] National Library of Finland, Helsinki, Finland.

New York Public Library (2011) What's on the Menu [Online Web Browser Project] New York Public Library, New York, NY USA.

Owens, T. (2012, May 20). "The crowd and the library." [Blog post] Retrieved from http://www.trevorowens.org/2012/05/the-crowd-and-the-library/ (accessed May 2013)

Ryu, P-M., and Choi, K-S. "Determining the specificity of terms using inside-outside information: a necessary condition of term hierarchy mining," in Information Processing Letters vol. 100, no. 2 (October 2006), pp. 76-82.

Tiltfactor (2012) Zen Tag [Online Web Browser Game] Tiltfactor, Hanover, NH USA.

Von Ahn, L., and Dabbish, L. "Labeling images with a computer game," in Proceedings of CHI '04, (Vienna, Austria, April 2004), ACM Press, pp. 319-326.