

Purposeful Gaming and the Biodiversity Heritage Library

Playfully Crowdsourcing Transcription: A Proposal from the Tiltfactor Research Laboratory at Dartmouth College

20 May 2014



Submitted by Dr. Mary Flanagan
Tiltfactor Research Laboratory
22 Lebanon Street
246 Black Family Visual Arts Center
Dartmouth College
Hanover, New Hampshire 03755 USA
(603) 646-1007
Website: <http://www.tiltfactor.org>
Portfolio: <http://www.tiltfactor.org/games>
Contact: Mary.Flanagan@Dartmouth.edu
Project Manager: Sukie.Punjasthitkul@Dartmouth.edu

The design plans and design approach contained in this proposal have been submitted in confidence and contain trade secrets and/or privileged or confidential information and such data shall only be disclosed for evaluation purposes, provided that if a contract is awarded to this Proposer as a result of or in connection with the submission of this proposal, the BHL shall have the right to use or disclose the data therein to the extent provided in the contract.



Tiltfactor Executive Summary

Tiltfactor invents new ways of thinking about important social issues through engaging games and play. Founded and directed by leading innovator Dr. Mary Flanagan, Tiltfactor is the award-winning design studio and research laboratory that takes on problem areas of national need and creates solutions through playful design. Our team harnesses the power of storytelling, systems thinking, and empathy in games to create powerful experiences to make positive differences in people's lives. The Tiltfactor team conducts rigorous studies that demonstrate the impact of these solutions and verify the efficacy of their methods. The group actively publishes results in scholarly journals. As trailblazers in the field of meaningful game design, Tiltfactor asks big questions about the intersections between psychology, learning, design, and technology. We are shaping not only the path of interactive design, but also 21st-century cognitive, social, and ethical skills.

Tiltfactor focuses on three key areas:

GAMES PROMOTING PUBLIC HEALTH

Our health-related games foster self-care and disease prevention, while promoting a greater understanding of modern healthcare delivery systems. They help players think deeply about important issues such as vaccination, and the health conditions and risk factors faced by hundreds of thousands of people. The wide range of games we have created includes board games that educate the consumer, digital games that encourage vaccination, and even a physical sport: *RePlay Health*, which explores solutions to the American health care system cost crisis. We have worked with the Robert Wood Johnson Foundation, the Dartmouth Center for Health Care Delivery Science, the Minister of Health of Rwanda, and the Rippel Foundation.

GAMES ADDRESSING COGNITIVE BIASES

These games aim to improve communication and cooperation skills, to reduce players' prejudices and biases, and to empower local communities in collaborative decision-making. For example, the pro-social party card games *Buffalo* and *Awkward Moment*, were developed as part of the National Science Foundation-funded project "Transforming Science, Technology, Engineering, and Math (STEM) for Women & Girls: Reworking Stereotypes & Bias."

CROWDSOURCING

In collaboration with national libraries and museums, Tiltfactor is creating game-based engines for knowledge generation. Our team's pioneering crowdsourcing work with Metadata Games, an internationally-recognized open-source game-based platform to help collect tags and other information for archive and library holdings, has brought us to the center of the dialogue about crowdsourcing at cultural heritage institutions. Ranging from simple gamified tagging tools to engaging arcade-style games, Tiltfactor's Metadata Games are designed to reward players for generating or verifying high quality metadata. The project is supported by an NEH Digital Implementation grant; more information can be found at <http://metadatagames.org/>.



TILTFACTOR RESEARCH

Our studies demonstrate that Tiltfactor produces meaningful games with cutting edge, empirically supported positive social and cognitive outcomes. Our randomized experimental study on *Buffalo* showed that the game significantly decreases players' prejudice and improves their understanding of the complexity of social identities. *Awkward Moment* triples players' association between women and science, and inspires greater assertiveness of response in confronting social biases. *POX* research demonstrates that a majority of players gain a nuanced understanding of herd immunity (disease prevention), and both *POX* and *ZOMBIEPOX* heighten players' perception of the value of vaccination and significantly improve systems thinking. Recent studies on Metadata Games explore questions like: "What are the key motivational factors that drive users to play a crowdsourcing game, and how can we frame the play experience in a way that maximizes player motivation and engagement?" and, "Does participation help create a psychological connection between crowdsourcing participants and a particular cultural heritage institution, media collection, or subject matter?" Flanagan and the Tiltfactor team have deep experience with crowdsourcing tools and have evidence of the efficacy of our designs across several research studies. The team is highly qualified and extremely familiar with the issues and processes involved (see: Flanagan & Carini 2012; Flanagan et al. 2013; Flanagan et al. 2014; Manzo et al 2014).

Our research studies on crowdsourcing games demonstrate that our unique approach to the design of such games can increase accurate tags and provide evidence of increased findability. For example, one study revealed that participants searching with access to traditional metadata took six times longer to search for each image (M = 154.1 secs, SD = 98.84) compared to participants that were given access to both traditional and folksonomic metadata (M = 25.08 secs, SD = 19.39), and successfully retrieved significantly fewer items overall.

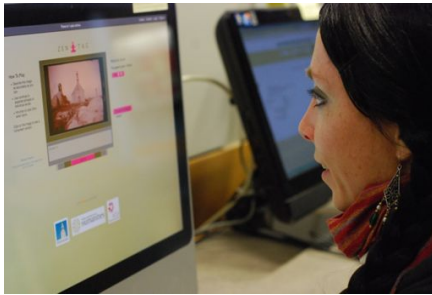
TILTFACTOR AS A CROWDSOURCING PARTNER

Informed by our experience with the Metadata Games project and our participation in the crowdsourcing community at large, Dartmouth College proposes to build two games and a tool for the Biodiversity Heritage Library to help facilitate high accuracy online transcriptions. As we have created the Metadata Games platform (with support from the NEH), we have made connections across national libraries and we have surveyed institutional needs at over 30 libraries and archives ranging from small to large. We have investigated firsthand which technological platforms are reusable for others building onto systems and which can most effectively tie into larger digital platforms like the DPLA. We have made international connections with crowdsourcing efforts at the British Library and other institutions worldwide. More broadly, PI Flanagan is engaged across members of the diverse NEH, ACLS, NSF, NIH, and IMLS communities. Because of this cross-sector involvement and expertise with new technologies and crowdsourcing, Flanagan and her colleagues are especially well positioned to bring together the diverse voices in the field (e.g., library scientists, archivists, humanities scholars, computer scientists, etc.). The team is currently working with the IMLS to create a collection of open crowdsourcing resources for nonprofit museums, archives, and libraries.



Tiltfactor Games Relevant to This Project

At Tiltfactor we've developed a suite of game apps for crowdsourcing.



[Zen Tag](#) (and specialized skins, such as *Ships Tag*)

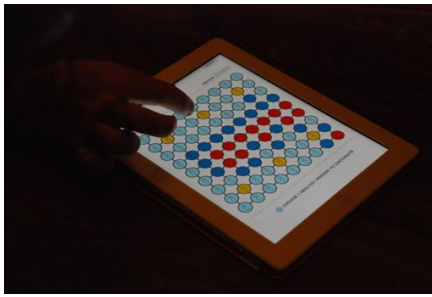
Our simplest metadata gathering game, *Zen Tag*, is designed to engage primarily non-gaming audiences who are interested either in seeing the images from our content partners, or in contributing to the our partners' data. Studies has shown that *Zen Tag* collected more tags per user on average (0.84 tags per image per contributor) than the Library of Congress *Flickr* Initiative (0.006 tags per image per contributor), and that more than 97% of the tags gathered by the game were accurate. Further research on this game has revealed how to minimize the impact of the psychological phenomenon of social loafing on player motivation.



[Stupid Robot](#)

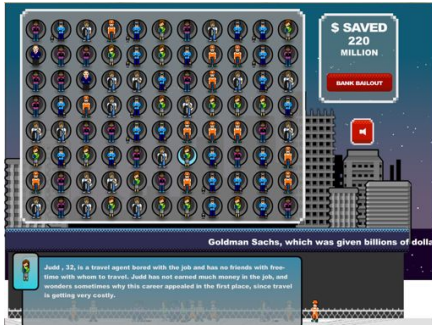
In this soon-to-be released single-player browser game, players are challenged to teach *Stupid Robot* (who looks at everything but understands nothing) as much as they can about an image from one of our partnering libraries' digital collections. *Stupid Robot* is built to engage gaming audiences while collecting high quality metadata about the images it's played with.

Stupid Robot is highly successful for two major reasons: first, its narrative is designed to act as a conceptual model for teaching gameplay and tagging. The simple narrative of teaching the adorable robot concepts and the explanation of its learning process streamlines accessibility and new player uptake. Second, *Stupid Robot* engages players by rewarding them with polished audio, animation, and score for submitting quality metadata. The game uses several metrics to assess the metadata, including dictionary checks (to identify common words), character sequences (to throw out nonsense), agreement with past players, and when all else fails educated guessing. These persuasive rewards are key to keeping players playing for long periods of time, resulting in optimal data collection.



[POX for iPad](#)

An award-winning iPad app and board game that helps players understand the concept of herd immunity through vaccination, and the importance of public health. The game has received 4.5/5 stars at the [App store](#) and won Best Digital Game at Meaningful Play 2012.



[Layoff](#)

A browser game for Tiltfactor's [Values at Play](#) project, *Layoff* is a casual game that also offers researchers an examination of how empathy operates in a game based on a time of crisis. In the game, players play from the side of management needing to cut jobs. At its peak, *Layoff* saw over a million plays in one day and received major media attention.

Our Clients and Collaborative Partners

Current and past collaborators and clients include:

- National Science Foundation (NSF)
- National Endowment for the Humanities (NEH)
- British Library
- Disney Imagineering
- Boston Public Library
- Digital Public Library of America (DPLA)
- Clark Art Library
- Open Parks Network
- Robert Wood Johnson Foundation
- Digital Mill
- Rippel Foundation
- The Dartmouth Center for Health Care Delivery Science
- Association of Professional Futurists
- American Council of Learned Societies

Standard Procedure of Game Development

In developing crowdsourcing games, design teams encounter many challenges to participation, accuracy, and providing feedback. Not only are there technical challenges in designing intricate systems that encourage high quality answers or data from players, but there are psychological and design issues that our team has experience addressing. The specific challenges our team identified on the BHL purposeful gaming project to which we have solid approaches.



Tiltfactor uses its unique human-centered design methodology, Critical Play, to incorporate fundamental human values, psychological principles to create engaging games that gather high quality data. Our design cycle happens in 6 broad steps, but in accordance with human-centered design principles, some steps may be revisited multiple times before proceeding:

1. Foundational research - through literature reviews and user interviews, our team identifies key problems and core target audiences for potential games to address.
2. Ideation and brainstorming - our team works with subject matter experts and collaborators to come up with many initial game concepts targeted towards users and built to solve the problems identified during the foundational research step. We jointly narrow these many concepts down to a few concepts to pursue by combining the most promising elements from each pitch.
3. Initial mockups and playtesting - using paper or low-tech digital tools such as Powerpoint and Game Salad, our team mocks up rapid functional prototypes in order to assess viability, datastructures, and playability. Successful playtests lead to further higher fidelity prototypes targeted at specific functionality made using more sophisticated tools. Art and graphic designers mockup game screens and proxy audio is created.
4. Design specification and game development - using data gathered from internal and external playtests with the prototypes, Tiltfactor creates a full design specification for the most successful concepts (rated both on fun and quality of data produced), and begins to program the games.
5. Alpha testing and iteration - early builds of the game using our graphic designers' game screens and art assets are playtested externally as they become available for further design enhancements.
6. Beta testing and iteration - further along builds of the game are playtested for structural integrity and polish
7. Research and assessment - as a final round of design iteration before release, our research team conducts A/B studies on beta game builds, manipulating psychological factors including narrative, scoring incentives, and text framing to determine their effects (as well as the effects of demographic and subject data) on quality outcomes. The manipulations with the highest levels of success are chosen for implementation in the final release(s).

When working with stakeholders, our team updates our collaborators weekly to ensure that development is on track and that design choices meet needs for all involved parties. We use a number of project management tools, including Asana, Skype, Google Hangouts, Google Drive, and Dropbox for communication and media asset management. Code is managed using Git and Github version control.



Proposed Timeline

We propose a nine-month development cycle beginning on approximately July 21, 2014 and ending with final games deployed and archived by May 1, 2015.



Activities	Deliverable	Persons Responsible	Date
Game Development Kickoff		All	July 21, 2014
Initial design of game(s)	Design documents and Graphics	Game, art designers	July - September 2014
Iterative Back End Development	Working code	Programmer	August - October 2014
Iterative Game Development	Testable Game prototype(s)	All	September - December 2014
Testing, analysis of game data; further game revisions	Game data, game revisions	All	December 2014 - April 2015
Press	Marketing, Presentation Materials	Project Manager	April - May 2015
Official Release of Games; Archives delivered	Release Game	All	May 1, 2015
Additional Support, Maintenance		Programmer, Game Designer, QA	May - July 2015

Team Structure/Expertise Allocating to Project

Tiltfactor’s eclectic team has extensive experience in game design; we have 2 game designers on staff in addition to Flanagan, 2 social psychologists, and the equivalent of 3 project managers, programmers, and designers. Due to Flanagan's work in game development teams over the last 20 years, the lab has access to a strong network of programmers, designers, animators, and graphic artists who are available as contractors.

Dr. Mary Flanagan (Tiltfactor director, game design) is the Distinguished Professor in Digital Humanities at Dartmouth College and founder/director of Tiltfactor. She innovated in the game industry in the 1990s and began Tiltfactor over a decade ago to research and create games which have impact above and beyond entertainment. Her approach



She has led the team to revolutionize serious games to enact social change from a psychological approach. In the “Serious Games” domain, Flanagan has nearly 20 years of experience crafting award winning games; she is deeply invested in novel game design and what she calls “collaborative strategy” games. She will work on design issues, systems design, and institutional viability.

Max Seidman (game design) is Tiltfactor’s lead game designer. With over a dozen game titles to his credit, Max is an expert in creating compelling gameplay experiences, developing prototypes, and devising design solutions to optimize user experience.

Sukdith “Sukie” Punjasthitkul (project management, sound, backend development, QA) is a project manager and technical lead at Tiltfactor. He brings over a decade of experience in digital production and has worked on award-winning projects ranging from simulations to the creation of online social networks.

Dr. Geoff Kaufman (research) is Tiltfactor’s specializes in psychology and assessment and leads Tiltfactor’s research studies. Dr. Kaufman will advise on data accuracy checking methods, data collection for improved usability, and future studies.

Sarah Ettinger (art direction) is Tiltfactor’s graphic designer and animator who is responsible for the look and feel of the games and their polish elements.

Our team of programmers have capacities from backend development to HTML 5 and app development. We have used the following frameworks and software libraries for past and current projects:

- Yii - PHP-based MVC framework for rapid web development
- jQuery - javascript library
- createJS - A suite of Javascript libraries & tools for building rich, interactive experiences with HTML5

COST ESTIMATES

Below are preliminary cost estimates for BHL deliverables.

Backend Development

Backend development includes a database where player-submitted tags and profile data is collected and stored, and writing code to access BHL collection media and associated data (transcriptions). We estimate that the backend will cost approximately \$15,000.

Game #1

Game one targets non-gaming and altruism-motivated players. The team involves game design experts, handling the mechanics, player interactions, user interface, and theme; media experts, in charge of graphics, audio, and animations to flesh out the theme; programmers implement designers work across multiple platforms; and quality assurance (QA) testing to verify code stability. The design process also includes iterative game revisions based on test feedback and research to ensure polished, engaging gameplay that results in high quality data. Cost estimate is \$38,000.



Game #2 OR Transcription Tool

The second game would either be a game played by game-motivated players, or a non-game transcription tool to support game one. For the transcription tool, the team would involve interface design; graphics, audio, and animation to enhance the user experience; programming and QA testing to ensure a simple and easy-to-use experience. Cost estimate is \$38,000.

Outreach and promotion

We highly recommend devoting effort to publicize the games. Our moderate outreach support via social media and special interest websites will increase the player base. Cost estimate is \$4,000.

Optional: additional updates and maintenance up to 3 months after launch

These include bug fixes and minor feature tweaks after the official launch of the games, billed hourly in this category up to this amount. Cost estimate is \$5,000.

Dartmouth College will receive at least a 10% overhead charge for this contract work.

Design Techniques and Data Collection for the Project

A) Known-Bad Insertion

The data verification technique that we will use is called *Known-Bad Insertion or Error Insertion*. The crowdsourcing literature describes Known-Bad Insertion as a method in which system-generated ‘bad’ entries are offered to players/users in order to assess the accuracy of their contributions. Implementing crowdsourcing games to control for quality of transcription data is always a challenging task. Fortunately, of the tested methods that exist, Known-Bad Insertion can be modified to fit BHL’s needs exceptionally well: the method ensures high player accuracy with a low possibility for player abuse. In order to implement the Known-Bad Insertion method for this application, the game system must have relatively high confidence that one of a provided number of possible transcriptions (either from OCR or transcribers) is correct. Starting with these possibly-correct transcriptions, the game system:

1. Analyzes the differences between possibly-correct transcriptions.
2. Based on these differences, generates believable but incorrect known-bad transcriptions. Better (more believable) known-bad insertions can be generated using error matrices.
3. Presents the possibly-correct transcriptions mixed with the known-bad insertions to the player in the game context. If the player chooses a known-bad insertion, she is penalized in gameplay and her data is not saved. If the player chooses a possibly-correct transcription, that transcription’s weighting is increased in the database and the player is rewarded. (Note: higher ratios of known-bad insertions to possibly-correct transcriptions increase confidence of correct answers)



In addition, if players frequently choose known-bad insertions, or OCR or transcription outputs that can be algorithmically flagged based on low player accuracy (frequent choice of known-bad insertions) or through use of a limited 'skip' function which could flag potentially problematic examples.

B) Identifying and Capitalizing on Player Motivation

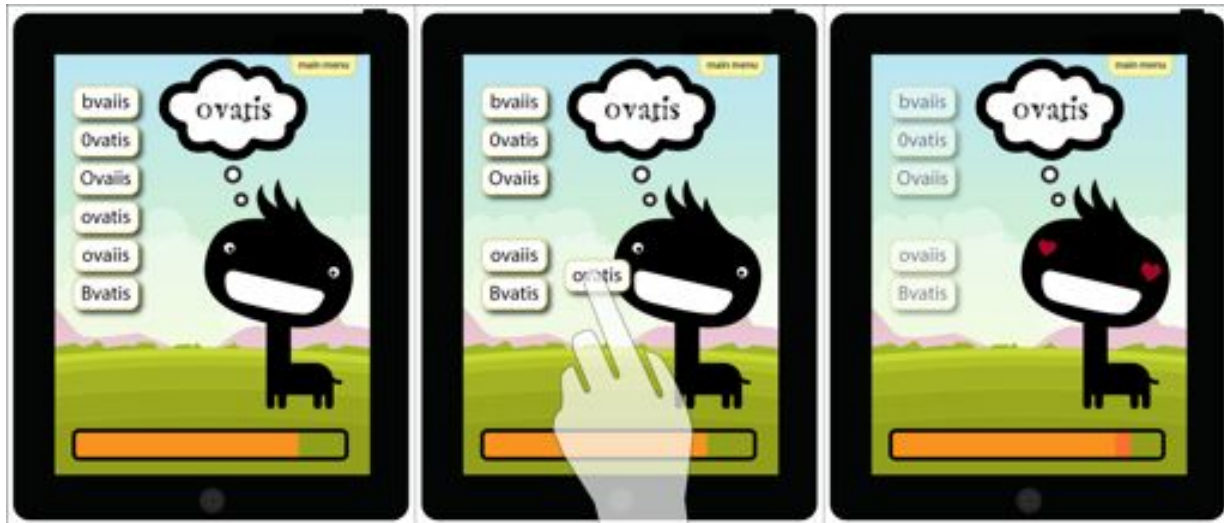
Digital crowdsourcing games can be designed to match participatory motivations and game play options in ways that most effectively address the idiosyncratic needs of players. Our past research has led us to analyze "gaming" and "non-gaming" target audiences through a different lens - not by their gaming proclivities, but by their core motivations in participating in crowdsourcing activities. *Altruism-motivated* players are primarily playing in order to contribute to BHL's collections. These players want minimal frills and distractions, but as with any activity, poignant feedback and clean interfaces can help streamline player experiences and lead to longer player retention. *Game-motivated* players are primarily playing for the game experience itself. These players want authentic, engaging gameplay, with developed narratives and progression. To this end, we propose developing two games - one primarily for altruism-motivated players and one for game-motivated players.

Preliminary Game Concepts for BHL

Using Known-Bad Insertion, we have brainstormed three initial game concepts as examples of what games for this project *could* look like, as well as an additional transcription tool that could be developed for the project. These draft game pitches are intended for a mix of audiences across altruism-motivated and game-motivated spectrum. The game concepts should be seen as examples, and may bear no relation to eventual game designs.

1. Game One: Booklouse

Intended for mostly 'altruism-motivated' audience, *Booklouse* is a simple know-bad insertion game with a light narrative and an empathy-inspiring character. Players get to know the Booklouse, a strange yet adorable creature who likes to eat words. The Booklouse has a challenge, however: *poisonous* words often look very similar to *delicious* words. On each level, the Booklouse wants to eat a specific word, shown as an image in a thought bubble that is in fact drawn from a digitized manuscript. The player can feed him one of a number of words, where the number of words increases throughout the game. Two of the words shown are possibly-correct transcriptions, and the rest are known-bad insertions. If the player feeds the Booklouse a possibly-correct word, the booklouse becomes happy and its progress bar fills a small amount. If the player feeds the Booklouse becomes sick and its progress bar decreases a moderate amount. Either way, after feeding the Booklouse it comes up with a new word it desires, and new possibilities appear on screen.



Because *Booklouse* is a simple game, its details are of utmost importance. Special care will be taken with fine-tuning animation and audio feedback. During the development of *Booklouse*, factors including narrative, scoring incentives, time pressures, and text framing will be experimented with to determine the optimal arrangement for approachability, player retention, and both quality and quantity of data generated.

2. Game 2: Cipher

Designed for 'game-motivated' players, *Cipher* is a game that challenges the player to get as far as she can in one fast-paced level by adeptly avoiding obstacles. *Cipher* is inspired by the highly successful "infini-runner" genre. Infini-runner games, such *Temple Run* (mobile), *Flight* (mobile, browser), and *Canabalt* (browser, mobile) are known for being approachable by many different kinds of players, engaging, and addictive.

In *Cipher*, the player is a hacker controlling her wireframe, a digital avatar, as she moves through the 'net. The player works to avoid firewalls and outrun the ever-present antivirus program giving chase, all while collecting packets of data and decrypting them for in-game currency.

When the player collects a data packet, the data is delivered to her in the form of a known-bad insertion puzzle: the 'encrypted' data is a scan of a word, and the possible decryptions are the possibly-correct transcriptions mixed with the known-bad insertions. The player has a limited amount of time to choose the correct decryption in order to get paid, all the while dodging obstacles. Like many games in the genre, each level of *Cipher* is procedurally generated, and increases in speed and difficulty throughout its course until players inevitably lose.



Cipher's gameplay is built on a compulsion loop that encourages repeated play, motivating players to verify as many words as possible. Once the player is caught by the antivirus program in the game, the level ends, and the player can utilize all the currency she accumulated by verifying data on upgrades so that she can do better next run: faster speeds for her wireframe, abilities to ignore some obstacles, more valuable data packets, etc. This creates a drive to continue playing, as the player pushes forward to collect more currency (verify more words) so that she can unlock more upgrades, which then let her collect more currency.

3. *WordFinder*

WordFinder is a game-inspired tool to enable players to place words in transcriptions to their locations on a scanned manuscript page. Designing transcription verification games that can handle both single-word located OCR output *and* full-page transcription data is a difficult challenge. *WordFinder's* goal is to break up full-page transcription data into located words, so that it can be handled by games in the exact same manner as located OCR output.

Using the tool is simple: the user is shown a manuscript page and given three lines of the transcription, for context. One word in those three lines is highlighted, and the user is asked to click and drag (in browsers) or touch and drag (on tablets) a box around the highlighted word. Then they move on to the next word.



Contract Terms and Conditions

Contractor (Tiltfactor) shall be paid no later than thirty (30) calendar days after submitting to Company (BHL) completed work and applicable time-sheets and itemized invoice for fees and any actual expenses pre-approved by Company provided that Consultant's services meet Company's satisfaction. Contractor will submit completed work through electronic download, remote installation, submission into source control systems, updates to issue tracking systems and E-mail as appropriate. Typically this will be electronic document or computer source-code. Contractor agrees that the fees and charges for any follow-on or additional work not covered in the scope of work described in any Work Order shall be performed at the lesser of (1) Contractor's then-current rates for such work or (2) the hourly rates applicable to the scope of work fixed by the initial Work Order. Company shall pay all undisputed fees and expenses owing to Consultant hereunder within thirty (30) calendar days after Consultant has submitted to Company an itemized invoice.

FEES, EXPENSES, AND PAYMENT

For any future maintenance services and support, such as code updates, maintenance, or support beyond the 3 months or \$6,000 allocated, the Contractor will provide an estimate of the work and time required. The first 10 hours of updates will be complimentary.

Payment Terms

- 25% upon delivery of game design documents
- 25% upon delivery of two Prototypes
- 25% upon delivery of two Beta games
- 25% upon release of games and close of the project May 1, 2015



References

Flanagan, M., & Carini, P. (2012). How games can help us access and understand archival images. *American Archivist* 75: 514-537.

Flanagan, M., Punjasthitkul, S., & Kaufman, G. 2014 (submitted). A proposed “OED Model” for parsing and incorporating crowdsourced metadata into digital collections. Paper submitted to the *2014 International Conference on Digital Libraries*.

Flanagan, M., Punjasthitkul, S., Seidman, M., Kaufman, G., & Carini, P. 2013. Citizen archivists at play: Game design for gathering metadata for cultural heritage institutions. In *Proc. of the 2013 DiGRA Conference* (pp. 1-13). Atlanta, GA: DiGRA.

Manzo, C., Kaufman, G., Punjasthitkul, S., and Flanagan, M. (2014). “Folksonomies to Libraries and Back Again: Assessing The Value of Crowdsourced, User-Generated Metadata.” *Proceedings of Dublin Core 2014* (submitted).

Placeholder Resources

“[Giraffe](#)” designed by Pirog tetyana under [CC BY 3.0](#)

“[Boy](#)” designed by Housin Aziz under [CC BY 3.0](#)

“[Pointing](#)” designed by Anthony Lui under [CC BY 3.0](#)

“[Cursor](#)” designed by Brandon Keelean under [CC BY 3.0](#)

“[Base Human Models \(Low Poly\) for Blender 2.5x](#)” by FreeLikeGNU under [CC BY SA 3.0](#)