

**“What’s on the Menu?”- From Software to Funware  
at The New York Public Library  
IMLS Project Number: LG-46-11-0080-11**

**Part I: Administrative Information**

**Institution:** The New York Public Library/NYPL Labs

**Project Title:** *What’s on the Menu? – From Software to Funware*

**Award Amount and Total Project Cost:** \$25,000

**Grant Period:** June 1, 2011 – May 31, 2012

**Project Directors:** Rebecca Federman (Culinary Collections Librarian), Michael Inman (Curator, Rare Book Division), Ben Vershbow (Manager, NYPL Labs)

**Part II: Project Summary**

**Overview**

*What’s On The Menu?* was designed to transform roughly 9,000 digitized images from NYPL's famed Buttolph Menu Collection into a searchable database of historical culinary and economic trends. Because of difficulties in mechanically extracting quality text from the menus, and from a desire to build a *structured* data set of discrete dishes and prices, it was determined that manual transcription would be the best method for creating the database. As this task was far too large and time-consuming to accomplish with internal resources, we (the project directors) saw this situation as an opportunity to explore user collaboration, or ‘crowdsourcing’, as a means of accomplishing the work. In other words: build a simple web application for transcribing the menus and see if members of the public would be willing to volunteer their time.

Our original application operated on the hypothesis that in order to sustain participation in an ambitious collections crowdsourcing project with a multi-year time horizon, it was essential to be creative with user engagement, drawing upon game design principles to devise ways of keeping participants interested, rewarded, and motivated. The initial plan was to launch the user-interface application as a no-frills beta, learn from the initial response, and then to work with a game design studio to revamp the site along more imaginative lines. But things turned out rather differently. Due to an overwhelming public response, the initial batch of menus was dispatched within about three months of launch, requiring us to seek additional resources, ramp up our digitization efforts, and re-factor other library workflows involved in keeping fresh content flowing into the site. Most significant, after settling down after the initial publicity spike, use of the site didn’t diminish. Instead it relaxed into a steady engagement pattern that continues to this day.

Contrary to our original intent, we have not applied any "gamification" approaches, focusing instead on improving core tools, user interface/design, and improving access to the data set. The experience has led us to question our assumptions about what motivates people to participate in collaborative projects, and has built our confidence that public organizations such as libraries and museums are uniquely positioned to tap into incentives and rationales for participation that are different from commercial enterprises.

***Bigger Picture: So You’ve Digitized a Collection...Now What?***

In recent decades, libraries, archives, and museums have been digitizing large quantities of rare and unique materials, moving many of them out of the institution and onto the Internet. While these efforts have greatly expanded basic access to certain collections, additional work is often required to fully realize their usefulness in the digital environment. Where descriptive metadata exists, it often requires modification and normalization against existing standards in order to better surface the materials in federated searches and other data-driven discovery mechanisms.

In some cases, metadata must be created for the first time. In other instances, as with the menus project, materials must be intensively processed following digitization in order to make their richest contents accessible to web browsers, search engines and application developers. Historical maps must be “georectified” in order to be searched and queried like today’s digital maps<sup>1</sup>. Books and newspapers must

---

<sup>1</sup> See Old Maps Online directory of georeferencing tools - <http://help.oldmapsonline.org/georeference>

be optically processed to extract searchable text, the output of which process often requires subsequent manual cleanup in order to become useful<sup>2</sup>. Still other texts (such as old city and business directories, church registries, playbills, ship's manifests, etc.) are in essence databases in printed form, and their contents must be laboriously transformed into structured information in order to be accessed and referenced against other data sets<sup>3</sup>.

Migrating the sum of human knowledge to the Internet is daunting, especially for resource-strapped cultural organizations. But these institutions, while lacking the financial and engineering assets of big technology and media companies, do have one ace up their sleeve: their public mission. The web contains harbingers of new kinds of public libraries and museums, such as the Internet Archive, Project Gutenberg, Freebase, Wikipedia, Flickr Commons, the Creative Commons; their purpose is to provide knowledge and tools for the civic realm, to serve as free resources for a lifetime of learning.

Most of these initiatives were built collaboratively, often for no pay or material reward, by a motivated corps of well-informed, diligent enthusiasts looking to devote their time to something bigger than themselves. Through networked collaboration on the Internet and an inspiring mission, Wikipedia built an open access encyclopedia by tapping into the same civic impulses and irrepressible curiosity that have propelled people through the doors of libraries and museums for generations. Now traditional institutions must engage this Internet citizenry in new ways—and many already are.

For NYPL, this suggests a new kind of public library, one built not only *for* but in collaboration *with* its public. The buzzword for this is “crowdsourcing”: breaking up large tasks into small pieces for a dispersed pool of workers. Increasingly, and in large part through the lessons learned in *What's on the Menu?*, we are coming to see crowdsourcing not only as way to accomplish work that might not otherwise be possible, but as an extension of our core mission. As Trevor Owens, Digital Archivist at the Library of Congress, puts it: “*it is about offering your users the opportunity to participate in public memory.*”<sup>4</sup>

### **Part III: Process**

#### ***Prior Investigations***

Founded as a repository for predominantly print and paper-based materials, NYPL is now exploring how to convert an analog knowledge base into a digital resource of comparable import. Mass digitization of the Library's printed book corpus began nearly a decade ago with its involvement in the Google Library Project and the Open Content Alliance (now absorbed by the Internet Archive). Our internal digitization efforts have focused on special collections: prints, photographs, manuscripts, and other rare book and archival materials. This also has been the locus of NYPL's experiments in digital collection building and data processing, which lately have been spearheaded by a recently-formed group, NYPL Labs<sup>5</sup>.

The Buttolph Menu Collection was identified early on as a promising test bed for experimentation: an ephemeral “edge” collection that has historically been difficult to catalog and preserve, yet which contains, in the aggregate, vast quantities of cultural data, a potential goldmine to historians, journalists, chefs, novelists and educators. Embedded in ink and print (and more recently, pixels) are millions of interrelated data points that tell social histories as vast as the population of oysters in New York City and as particular as the price of a cup of coffee at Child's Lunch Room on 130 Broadway in 1901 (5 cents). How best to extract this data? The menu collection has the added advantage of popular appeal. Featured in two major exhibitions at NYPL over the past decade and a major source for various scholarly works, the collection has been a consistent draw to enthusiasts, educators, and researchers, tapping into broad popular interest in food and the origins of dining culture.

The Library's first major venture into crowdsourcing was an NEH-funded project, the *Map Warper*<sup>6</sup>: a web-based platform where users are invited to assist with the georectification of historical atlases, producing

---

<sup>2</sup> See Google reCAPTCHA - <http://www.google.com/recaptcha> ; also Trove Australian Newspapers - <http://trove.nla.gov.au/ndp/del/home>

<sup>3</sup> See Pastmapper - <http://www.pastmapper.com/>

<sup>4</sup> Trevor Owens, “Crowdsourcing Cultural Heritage: The Objectives Are Upside Down” - <http://www.trevorowens.org/2012/03/crowdsourcing-cultural-heritage-the-objectives-are-upside-down/>

<sup>5</sup> <http://www.nypl.org/collections/labs>

<sup>6</sup> NYPL Map Warper - <http://maps.nypl.org>

exportable historical geodata. While the *Map Warper* project has enjoyed great success, the intensive task of transforming maps works best when accompanied by hands-on training or situated residencies—for instance, with a geography class from a local university—that occur over multiple weeks. Participation, therefore, while consistent, has been more modest and confined to a specialist audience.



A 1854 William Perris insurance atlas georectified against Open Street Map

The aim with *What's on the Menu?* (WOTM) was to create an experience that was simpler and less intimidating than the *Map Warper*, and to capitalize on popular interest in food to elicit greater participation. In order to attract the broadest possible audience, we strove to keep the task basic and clear (capture only dishes and prices); build the most stripped-down tool possible (click on a region of a menu image, type what you see); and keep barriers to participation low (no registration or login required). We decided to focus on these basics before diving into gamification.

### **The Idea**

When NYPL applied for Sparks! Ignition funding, the Library's web team had already begun designing and coding the WOTM beta website as a side project. Our plan was to launch the application in Spring 2011 with some modest outreach and to monitor the public's response. We were confident that the project would attract attention, but less certain that participation would persist. The plan was to assess the basic viability of the tool (was the task sufficiently clear? the tools usable? would people participate?), make any needed refinements and fixes, and then begin to devise interface and engagement elements to make the project "sticky" over time. From our original proposal:

*Menu transcription will be the central activity, embedded in a thoughtfully constructed array of game dynamics that tap into the thematic and historical richness of the material, celebrate users' identity, reward participation, recognize achievement, and give participants opportunities to share, collaborate, and compete around what they find. For this project, NYPL proposes to work with leading game designers to conceptualize and implement this layer of play, investigating a new frontier for libraries and other institutions looking to work with archival materials in the networked, participatory age.*

Our thinking was heavily influenced by two important new media thinkers: Clay Shirky and Jane McGonigal. Shirky has written on the opportunities and challenges faced by institutions seeking to leverage the internet to do work traditionally handled within an organization's walls.<sup>7</sup> Shirky cautions that as the organizing costs of participation networks fall, the difficulty of attracting participation (and then sustaining it when you have it) will rise precipitously. Along similar lines, Jane McGonigal asserts that for organizations to stay competitive in "the economy of engagement," they should think of their projects along the lines of games.<sup>8</sup> Games, according to McGonigal, have tremendous world-improving potential if harnessed for public-interest projects. What if the hours spent outrunning the cops in *Grand Theft Auto* or blasting Nazis in *Call of Duty* were channeled into figuring out peace in the Middle East, folding protein puzzles to aid scientific research, or gathering historical weather data sets to better predict the climate's future?<sup>9</sup> We saw the Library as being a perfect context for building "games for good."

<sup>7</sup> Shirky, Clay. Here comes everybody: The power of organizing without organizations. New York: Penguin, 2008.

<sup>8</sup> McGonigal, Jane. Engagement Economy. Palo Alto: Institute for the Future, 2008.

<sup>9</sup> PeaceMaker - <http://www.peacemakergame.com/> ; Foldit - <http://fold.it/portal/> ; Old Weather - <http://oldweather.org/>

We had some rough ideas of what the menu “game layer” might entail: recognition/reward structures and achievement-oriented narratives that might encourage users to keep up their productivity and to remain committed to the project over time. The specifics would come to light once our games collaborator/consultant was selected and the redesign process got underway.

### What Actually Happened

A basic first version of the participatory transcription application launched on April 18, 2011 and was an instant hit. In the first ten days, 100,000 dishes were transcribed. The project was featured in *The New York Times* 'Dining Section', National Public Radio, and a number of food and library blogs.



What's on the Menu? home page, two and a half weeks after launch

In the following months, New York chefs such as Mario Batali (*Babbo*, *Del Posto*, *Eataly*) and Brooklyn's Doug Crowell (*Buttermilk Channel*) publicly endorsed the project, and others (such as Rich Torrisi of Little Italy's *Torrisi Italian Specialties*, and David Chang, creator of the *Momofuku* chain) created special menus derived from items in the collections. As a testament to the ease-of-use of the tools, a class of 4th graders in San Antonio, Texas used the site to practice typing skills while learning about historical cuisine.



Transcription interface

Clearly, the project was having no trouble keeping participants engaged. In fact, we ended up having a new problem: keeping up with demand! WOTM had begun with a portion of the collection (a little under a quarter of the total ~45,000 items) that had been digitized some years back for the launch of the NYPL Digital Gallery<sup>10</sup>. With the public making quick work of that initial batch, we found ourselves re-prioritizing existing digitization queues to ensure a steady flow of new menus. New resources had to be secured to pay a collections assistant in the Rare Book Division (where the menu collection resides) to process the menus by hand, creating high-level item records and insurance estimates box by box before shipping the assets off to our Digital Imaging Unit in Queens.

<sup>10</sup> NYPL Digital Gallery – <http://digitalgallery.nypl.org>

But that wasn't the only hurdle. Effectively re-starting the cataloging of the menus exposed deficiencies in workflows and data models. Rudimentary records (created, in an interesting anticipatory echo of WOTM by a group of on-site volunteers a decade earlier) already existed for many of the un-digitized menus, but they were inconsistently formatted, and in some cases simply non-existent; even to remediate the existing data would require sorting through each individual menu by hand. A meeting was held with the project directors, the application developer, and the Library's metadata team and it was decided that it ultimately made more sense to retire the legacy menu database and to re-catalog the menus according to a new, lighter-weight schema. This process was put into effect in mid-summer of 2011, around the time that our Sparks! funding took effect. We were deep in the complexities of creating a sustainable digital collections project with real-time public demand and the game design ideas felt very much in the background. The public-facing tools were working well. We were approaching half a million dishes transcribed.

All the while, we were focused heavily on managing our new community of volunteers. As part of the "keep things simple" strategy, the site did not have a user registration system, so all contacts with users were initiated voluntarily through email and social media. We quickly set about recruiting a small team of interns in Fall 2011, who were trained to manage the day-to-day content administration duties: moving menus to the public transcription queue, checking contributions in the site's "under review" section (the intermediate stage in the menu transcription workflow), and responding to user queries via email.

In spite of the "on hold" nature of the original gamification concept, we did have active some basic engagement strategies. Social media was and continues to be our main outreach tool, particularly Twitter and Facebook (as well as good old-fashioned email). These channels help sustain public interest by providing staff with an outlet for sharing discoveries and curiosities from the archive (peculiar dishes, linked recipes, menus with a connection to current events, etc.) as the project chugs along. Interns are primarily responsible for maintaining these accounts (under the supervision of the project directors), and this allows them to get their feet wet as representatives of a collection and, at times, providers of a kind of web-based reference service.

Around the same time that the intern team was coming online, significant organizational changes were taking place. WOTM had been conceived as a pilot project of the Library's central web team, but a major strategic planning process for the research libraries in the spring and summer of 2010 laid the blueprint for a new experimental technology and design unit that would be fully devoted to projects such as these. Just as WOTM was making its explosive debut, resources were being allocated to assemble this new team, NYPL Labs.

This had several implications. First, it meant shifting away from working with outside contractors to building in-house expertise. Establishing a dedicated framework for collections innovation also led us to think more deeply about the implications of the project. Developing new user experiences would remain a strong focus, but so would data management and technical sustainability. Tackling the internal workflow issues that WOTM had exposed became higher priority—especially as public participation remained vigorous without the addition of any new website features or frills. What's more, the project's success sparked discussion at the highest levels of the institution about how to treat and preserve user-contributed metadata, so it became all the more incumbent upon the WOTM team to create a viable data management model and process.

Additionally, although the WOTM beta site was effective from a user standpoint, it had been built as a spare time project. The point was to launch quickly and to test hypotheses before investing any further resources. The beta did a good job with the transcription tool, but it lacked functionality that would enable users to browse, search and explore the collection (and the data emerging out of it). Moreover, many core functions needed to be re-written in the interest of sustainability and performance. In other words, in order to get off the ground, the project incurred a certain amount of "technical debt" that eventually had to be repaid by NYPL Labs in order to get the site into a more sustainable, fully featured condition.

The new Labs team began examining the technical platform and user interface in Fall 2011. We also continued to explore the idea of gamification, having several meetings throughout the fall, including one with Mary Flanagan of Dartmouth University, whose emerging "Metadata Games"<sup>11</sup> platform shows much

---

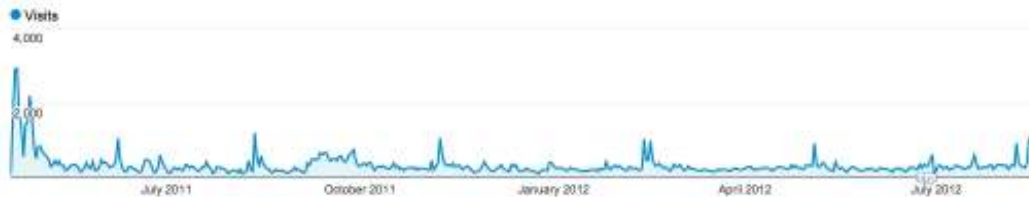
<sup>11</sup> <http://www.tiltfactor.org/metadata-games>

promise. We also explored the possibility that the Labs developers (some of whom had experience designing games) would implement the game layer internally. Several deadline-driven projects took precedence until Spring 2012 and then a substantial redesign and re-engineering of WOTM began in earnest.

## Part IV: Project Results

### What We Delivered, Where We'd Still Like to Go

By Spring 2012, the gamification idea had faded away. We were at the one year anniversary of the project and it was still going strong. The site had won various accolades, including the prestigious Roy Rosenzweig Award for Innovation in Digital History from the American Historical Association and was widely touted as a model in the cultural heritage sphere. Since the initial frenzy around launch, participation had tapered but then settled into a steady pattern of deep engagement, with occasional attention spikes whenever significant new press or social media coverage occurred.



Web analytics provide compelling clues as to the *depth* of engagement the site has fostered. In the 16 months following launch (April 18, 2011 to August 18, 2012), page views per visitor have averaged at nearly 23 (compare to 2.27 for visits to NYPL's main website in the same period). That has led 163,690 visits to produce nearly 4 million page views. Similarly, time spent on WOTM has averaged at 6:36 minutes (compared to 2:38 on NYPL.org). This suggests a story of deep immersion in the transcription activity and in exploration of the steadily growing trove of menu-derived data.

We decided to focus our new development efforts on the following areas: 1) overhaul of the application code (paying down our technical debt); 2) new visual design and user interface improvements; 3) new search engine and browsing features; and 4) a public API (applications programming interface) to provide other application developers or digital researcher access to real-time data from the project.



The new *What's on the Menu?* home page

The revamped WOTM site is a vast improvement over the original beta application. The release of the menus data API is a milestone for NYPL, bringing the Library a step closer to the time when the bulk of our digital content and data (at least the portion in the public domain) is available programmatically for other technologists, designers, and researchers to build upon. We have received a number of requests already from researcher and cultural heritage hacker types to start playing around with it, and have been in discussions with a few publications about highlighting stories told by the data. And in accord with modern web engineering best practice, we are using the same API to run some of the key features of the WOTM site.



New dish page interface, with data visualization tools

From a user engagement standpoint, the site was almost the same as it was. We still do not require participants to create an account, and do not, in fact, even offer the option. All contributions are anonymous and communications with participants are all user-initiated via email and social media. Eventually we'd like to offer a non-mandatory registration option so that we can begin to offer some recognition to our core participant base on a "leader board" or "hall of fame" type honor roll<sup>12</sup>. More than that, we'd like to someday create another tier of participation for more serious, dedicated users, entrusting them with more difficult tasks (such as defining menu sections or harvesting geographical data) and perhaps giving them other special privileges on the site or means of guiding and helping other users.

We have not yet created a user account system in part because NYPL is still developing aspects of its digital infrastructure that will make implementation simpler and more sustainable, such as a "single sign-on" account system that would traverse all NYPL web properties. As we have learned from this project, crowdsourcing at scale across the Library should not be attempted in the context of production-intensive, standalone apps like WOTM. For highly specialized tools such as the *Map Warper*, there may continue to be a rationale for freestanding projects, but document transcription and other manual data entry, such as tagging or item-level descriptions, could easily become a core function of an evolving digital collections platform. NYPL Labs would like to move in this direction, and we are paying close attention to platform-scale efforts such as the Citizen Archivist Dashboard<sup>13</sup> at the National Archives, or some of the things that may eventually emerge from the Digital Public Library of America. NYPL's Digital Gallery could very well evolve in this direction, continuing to provide online access to an ever wider array of digitized material alongside a suite of crowdsourcing projects and tools for those who want to get more deeply involved.

On the collection side, we are in early discussions with other leading menu repositories such as Cornell University, the Los Angeles Public Library and the University of Nevada, Las Vegas, about federating collections through our website and sharing the crowdsourced data with our respective repositories.

### **What We Learned: 'Gamification' isn't Necessarily Necessary**

There may be scenarios where creating a full game-like experience is warranted, especially when inviting users to contribute their own original content. (See the Smithsonian Museum of American Art's *Ghosts of a Chance*, or NYPL's own *Find the Future* —created with Jane McGonigal—as examples.<sup>14</sup>) But in the case of crowdsourced cultural heritage projects, where individuals are taking part in actual work carried out on collections, the incentives reside in the materials themselves and in the proposition of working in partnership with a public trust. This is also the reward. Involving the "crowd" isn't just a means to an end, it is an end in itself. That is the engagement we're looking for, and it turned out that WOTM had it all along. Returning to Trevor Owens (Library of Congress)<sup>15</sup>:

<sup>12</sup> See Trove Australian Newspapers, Text-Correctors Hall of Fame - <http://trove.nla.gov.au/ndp/del/hallOfFame>

<sup>13</sup> <http://www.archives.gov/citizen-archivist/>

<sup>14</sup> <http://ghostsofchance.com/> ; <http://game.nypl.org/>

<sup>15</sup> Trevor Owens, "Crowdsourcing Cultural Heritage: The Objectives Are Upside Down" - <http://www.trevorowens.org/2012/03/crowdsourcing-cultural-heritage-the-objectives-are-upside-down/>

*What crowdsourcing does, that most digital collection platforms fail to do, is offer an opportunity for someone to do something more than consume information. When done well, crowdsourcing offers us an opportunity to provide meaningful ways for individuals to engage with and contribute to public memory. Far from being an instrument which enables us to ultimately better deliver content to end users, crowdsourcing is the best way to actually engage our users in the fundamental reason that these digital collections exist in the first place.*

This is in the case of the public sector. In commercial scenarios, where there is on some level an act of persuasion taking place, a sometimes slippery exhortation (or seduction) to consume, an entirely different set of motivators and incentive structures are at play. Cultural heritage crowdsourcing is about contribution, not consumption. It is less persuasion, more call to action. It resonates on a frequency instantly picked up by a portion of the public that is already well convinced of the value of memory organizations and will go to lengths to bolster them. And by offering new, Internet-native ways of getting involved in the monumental task of migrating our heritage to the digital medium, we open ourselves to new audiences and collaborators who share these basic values, and who may begin to see libraries, archives, and museums as peers of Wikipedia et al.—true organs of the open web.

Should we recognize and reward participation? Absolutely. As collections crowdsourcing gains traction in cultural heritage organizations, it will be important to develop online community architectures not only to recognize, but also manage and coordinate a growing participant base. But doing so doesn't make our projects games, and we would do well to cut through the confusion on this subject. Margaret Robertson, of London-based game design studio Hide&Seek, has written lucidly on the topic, making the distinction between gamification and what she calls "pointsification"<sup>16</sup>:

*That problem being that gamification isn't gamification at all. What we're currently terming gamification is in fact the process of taking **the thing that is least essential to games** [her emphasis] and representing it as the core of the experience. Points and badges have no closer a relationship to games than they do to websites and fitness apps and loyalty cards. They're great tools for communicating progress and acknowledging effort, but neither points nor badges in any way constitute a game. Games just use them – as primary school teachers, military hierarchies and coffee shops have for centuries – to help people visualise things they might otherwise lose track of. They are the least important bit of a game, the bit that has the least to do with all of the rich cognitive, emotional and social drivers which gamifiers are intending to connect with.*

We in the library and museum trade should remember that the "rich cognitive, emotional and social drivers" Robertson describes come pre-woven into cultural heritage crowdsourcing projects. And not just for the nonprofits. Hugely successful participatory initiatives launched by genealogy websites such as FamilySearch and Ancestry have demonstrated that this can hold true in the commercial sphere as well.

In the case of WOTM, there are people for whom deep interaction with the menu collection is a reward in itself. They have stayed with the project because they believe in its overall objectives: to build a new kind of research tool for better understanding these menus in historical and cultural context, and for revealing the many stories they tell.

Our advice to another library or archive contemplating such an experiment is to take a step back and look at their collections. What items are unique to your institution? Of those items, which have demonstrated (or in your estimation potential) public appeal? Is there a problem that the public can help you solve to improve access to this collection? It does not matter if your idea involves a relatively marginal collection. NYPL's menu archive, though a treasure, was hardly at the top of the institution's collection development priorities, but it provided an excellent test bed for a new kind of library work, and its influence continues to ripple through subsequent efforts. If you have an idea clearly in mind, look around for available tools. And don't get hung up on gamification. The appeal of your collections and the call of your mission are your most valuable assets.

---

<sup>16</sup> Margaret Robertson, "Can't play, won't play" - <http://www.hideandseek.net/2010/10/06/cant-play-wont-play/>