NYPL Labs: Hacking the Library

BEN VERSHBOW
NYPL Labs, New York Public Library, New York, NY, USA

ABSTRACT. Over the past couple of years, The New York Public Library has made significant strides in jumpstarting a digital humanities research program through the establishment of a curator-focused, R&D technology unit called NYPL Labs. This article will examine the first three projects produced by NYPL Labs, a trilogy of online experiments exploring methods of mining new humanities data sets from special collections material, and of translating the Library’s public mission to the Web via user collaboration and crowdsourcing. In just two years, NYPL Labs has demonstrated how much can be accomplished when technical resources are focused and procedural constraints loosened, though the challenges of sustaining this work are already becoming apparent as the team must continue to innovate while maintaining a growing portfolio of projects.

KEYWORDS digital humanities, research libraries, crowdsourcing, geographic information systems, skunkworks, New York Public Library

The New York Public Library (NYPL) is a strange platypus: a sprawling branch library system, with 88 facilities serving neighborhoods across three of the city’s five boroughs (Brooklyn and Queens, formerly separate towns, have independent systems), crowned by four research centers of international stature, with collections rivaling those of the world’s great academic libraries. In contrast to their university counterparts, NYPL’s research libraries have no resident faculty or student body. They are open and free to the public, serving a diverse array of independent researchers, artists, writers and miscellaneous autodidacts drawn to the exhaustive collections, deep archives, free electronic resources and lofty spaces. In this respect, NYPL
more closely resembles the Library of Congress, the British Library, or any number of other national libraries, yet it is also a quintessentially New York institution, conditioned by the dense urbanism and intellectual energies of the metropolis. And its collections, with notable concentrations in Anglo-American literature, the performing arts and black culture, have been shaped indelibly by the literary, political, artistic and ethnic strands of the city’s history.

The NYPL’s century-old flagship central branch, renamed the Stephen A. Schwarzman Building in 2008, sits at the epicenter of midtown Manhattan (5th Avenue between 42nd and 40th Streets), on a site that at various points in the layer cake of New York’s history held bodies (a potter’s field), 20 million gallons of fresh water (the Croton Distributing Reservoir), a crystal palace (Exhibition of the Industry of all Nations, 1853) and since 1910, one of the world’s great repositories of human knowledge, now totaling 4.5 million books and over 60,000 linear feet of archival material.

NYPL Labs (Labs), an unlikely crew of artists, hackers and liberal arts refugees, works in the south courtyard, in a decade-old architectural insertion dug directly into the building’s foundation—like a bit of extra RAM slotted into the aging library motherboard. Labs is a new group, several years in the making but only acquiring full-time staff in the past year. Like many cultural organizations still reeling from the digital shift, NYPL has undergone frequent reconfigurations of its technology staff in recent years, establishing large new teams with broad mandates only to dismantle and redistribute them after a couple of years as leaders came and went.

Labs emerged amidst one such transition, as part of a strategic planning process that sought to better position the NYPL vis-à-vis the burgeoning field of Digital Humanities. Focusing on NYPL’s public mission and deep collections, the program was sketched with a heavy emphasis on user collaboration and open data, envisioning a kind of in-house technology startup that would venture proactively into the library in search of curatorial collaborations. The work was envisioned as inherently inter-disciplinary, empowering curators to think more like technologists and interaction designers, and vice versa.

Labs’ first round of investigations built directly on the foundations laid by a now shuttered unit called the Digital Library Program (DLP), which had served from roughly 2000–2008 as a soup-to-nuts digital production and curation team, covering everything from scanning to metadata creation to software development. Through the DLP’s efforts, NYPL digitized sizeable portions of its visual research material, making many of its most prized collections accessible on the open Web, along with lesser known curiosities. The great public legacy of this work is the NYPL Digital Gallery, which launched in 2005 and remains heavily used to this day.

The Digital Gallery marked the NYPL’s first major effort toward large-scale public access to digitized content. Despite a few redesigns, and a near quadrupling of the amount of material it provides, the site’s architecture
and user interface remain largely the same as when it launched seven years ago. The site provides free, straightforward access to approximately 800,000 digital images: one of the largest open cultural heritage repositories of its kind. There you can digitally peruse a wide range of visual marvels ranging the full history of photography, and a dazzling range of prints, illustrations and other visual documents.

Given a strong directive to experiment, but with minimal access to the NYPL’s digital infrastructure (and without any remit to digitize new collections), NYPL Labs looked to the impressive legacy left by the DLP and imagined ways of going deeper with certain collections already available via the Digital Gallery. These projects looked at digitization as a beginning, not an end, the first step in a process of remediation and user interaction that might lead to entirely new derivative collections and data sets native to the digital medium. And, in the process, new experiences for the publics who engage with them.

THE MAP WARPER: HOW TO BUILD A TIME MACHINE

The sine qua non of Labs’ work is the Map Warper (2010), a project that predates the formation of the team and whose existence was crucial in making the case for establishing the curator-focused digital skunkworks that Labs has become today. The Warper project is helmed by a geospatial librarian and like all but one of Labs’ six members, he comes from a non-library background, having studied art and cartography, but his hybrid sensibility makes him an ideal curator for a collection so uniquely poised for digital transformation. The art and science of mapping have undergone a revolution in the digital era that can hardly be overstated. The most quotidian interactions with maps today draw upon startling convergences of data that allow us to query a vast range of information within the frame of a specific locale.

In Google Maps, for instance, or any comparable digital map service, one can search for “Soho NYC” and be dropped directly into this neighborhood on the western part of lower Manhattan between Houston and Canal Streets. From there one can view up-to-the-minute traffic conditions, overlay public transit routes, bring up user-generated photos of the area, even check the local weather conditions. One can also search for restaurants, banks, shops, theaters, etc., and read reviews and ratings of these places—even do a virtual walk of the storefronts. And all of this is via an intuitive interface in which the Earth is rendered as a seamless spatial canvas through which one pans, zooms and queries. Today this startling ability is commonplace. It is how we expect maps to work.

These advances in geospatial technology led to an extended thought experiment for the Labs team: could a historical equivalent be built (at least for New York City) of Google Maps (or more properly, its open source
competitor, Open Street Map)? Essentially, a digital map interface with a “go back in time” function, allowing you to compare changes in the urban landscape over the years. Imagine if the Library’s vast quantities of New York City historical source material could be retrievable through such an interface.

The mind boggles when one extrapolates outward because what is being imagined is a kind of time machine: detailed, drillable layers of urban history down to the individual address or landmark. And when the lens expands outward to include other library collections with a geographical dimension (both at NYPL and beyond)—residential and business directories, local newspapers and periodicals, literary archives, corporate records, photographs, prints, menus, playbills, church registries, the list goes on—one begins to see an intricate needlework of inter-stitched data, cross-referencing through time and space.

Within the NYPL Digital Gallery today one can explore thousands of digitized New York City maps, originally published by fire insurance or real estate companies, whose businesses depended on fastidiously detailed records of the City’s layout and architecture down to the individual structures. Taken in the aggregate, these maps comprise a dazzlingly granular data set of the built environment of New York City going back hundreds of years. But scanning the maps does not automatically give you access to this data.

A simple geographic query in the Digital Gallery, say for “Wall Street,” illustrates the limitations of the available bibliographic records. Atlas sheets of lower Manhattan with Wall Street running right down the middle of the page often, frustratingly, will fail to turn up in the results. The metadata tends to contain only the bounding streets of the handful of square blocks depicted on a single sheet: the data on the edges. Information on such edges is in relatively good supply since it is by these edges that the traditional user of a bound atlas volume has to mentally stitch together geography that has been sliced and diced into paged denominations. But maps, in their essence, are not naturally codex-shaped. They contain many nested edges, an infinite number of possible frames of inquiry, and the names of interior streets are just the tip of the massive iceberg of submerged data—street addresses, named landmarks and building footprints with all their related attributes (e.g., wood structure, three stories, commercial use, etc.)—that can be extracted only after the maps have been digitized. To get at this data, Labs created the Map Warper (2010).

The Map Warper toolkit, developed by an external geospatial technology firm called Topomancy, is an open source, Web-based alternative to commercial software such as ArcGIS designed to create, compile and analyze geographical data. The Map Warper is used heavily by Map Division staff, and also by members of the public, who are invited to create an account and learn to work with our maps through a video tutorial. Or, if based in New York, users can learn directly from staff in our Citizen Cartography workshops, held on a roughly monthly basis in the Map Division reading room.
With the help of a generous grant from the National Endowment for the Humanities, tens of thousands of NYC atlas sheets have been digitized in recent years and fed not only into the Digital Gallery, but into the Map Warper, where, through a series of manipulations, they are enriched with spatial reference information.

Here is how it works: once maps have been scanned and converted to high resolution digital images, users (both Map Division staff and interested members of the public) walk the maps through a process of “georectification.” Also known as “rubbersheeting” or “warping,” georectifying is the process of aligning pixels on an old map to precise latitude/longitude points on a contemporary virtual map. Given the high precision of the insurance atlases, the process works remarkably well, resulting in near-perfect synced historical maps that can be viewed against today’s geography. (See Figure 1).

After warping, users then crop away page borders and any other non-map information and proceed to tie the atlas sheets into comprehensive layers. This is when the maps truly begin to transcend their original format: the books are unbound, their pages laid across the geography they describe. For example, an 1854 New York street atlas created for insurance companies by the civil engineer and surveyor William Perris, becomes Layer 861. (See Figure 2.) This layer appears as a seamless canvas of streets, blocks and buildings that lies across Manhattan Island like a quilt running from the Battery up to 42nd Street (calling to mind Borges’ famous parable of the map so exhaustively accurate that it literally draped itself across the territory at a scale of one to one). This layer is a new type of library resource: a

**FIGURE 1** Rectified Map of Several Square Blocks in Lower Manhattan, 1857. (Color figure available online).
skein of historical geodata that can be freely downloaded and explored in programs such as Google Earth or Quantum GIS, and potentially mashed up with other data sets.

But warping, cropping and layering are just, as it were, the groundwork. Next comes the real time-eater: capturing by hand (via polygon tracing tools and text transcription) the vast quantities of information contained in the maps which are readily viewable to the human eye, but invisible to machines. In GIS parlance, this is called digitizing “features” from the maps—rendering them in machine-readable forms. In the case of the NYC atlases, the features include ward boundaries, landmarks and points of interest, hydrography, building footprints, addresses, and anything else the map may convey via writing, color coding or other markings. Great strides have been made in harvesting these sorts of data from the afore-mentioned 1854 William Perris atlas (Layer 861). As of this writing, over 58,000 building footprints, with their corresponding attributes, have been traced from the georectified map mosaic. This represents yet another derivative resource, another skein of data that can be visualized, analyzed and referenced in the digital datascape (see Figures 3 and 4).
NYPL Labs and Topomancy are now at work on a spin-off project: a geographical index, or gazetteer, of historical New York City. This database and Web service, provisionally titled “NYC Chronology of Place,” will contain place names and geospatially bounded locations going back to the beginnings of the city’s recorded history. It will aggregate both existing public data sets such as GeoNames, Open Street Map, and historical county and ward boundaries, as well as the fine-grained, hand-gathered features from the Map Warper, tracking changes to the built environment over time. The gazetteer will function both as a Web-based directory of New York’s geographical past, and as a historical georefencing service: a tool for placing pins on maps, not only in specific places, but specific times.
However, the vast majority of feature digitization has been within map layer 861, covering approximately eight square miles of Manhattan Island in 1854. To continue harvesting data at this level of detail and to populate the gazetteer, Labs will have to ramp up the crowdsourcing efforts around the project, which at this point are relatively modest. Though there have been notable successes (one user, for instance, single-handedly georectified nearly an entire Brooklyn street atlas, well over 200 sheets), public participation is currently more of a supplement to work carried out steadily by Map Division staff. The Labs team is currently doing some initial work on making these currently expert-level tools a little more user-friendly, and even contemplating more game-like experiences to engage a broader audience.

**WHAT'S ON THE MENU?: DEEPER INTO CROWDSOURCING**

The next project that took shape, and the first to be launched formally under the Labs rubric, put user collaboration more squarely in its sights.

The menu archive, begun by a donation from Miss Frank E. Buttolph, is one of those testaments to the wildly omnivorous collecting carried out by The New York Public Library in its more than a century of existence.
Comprising more than 45,000 items going back to the 1840s and the origins of public dining culture, it is the largest collection of its kind, with particular strengths in New York City in the late 19th and early 20th centuries.

The collection has always been popular, playing a prominent role in two major NYPL exhibitions in the past decade alone: “New York Eats Out,” curated by former New York Times food critic William Grimes, whose book Appetite City, on the origins of New York restaurant culture, is based heavily in research carried out in the NYPL’s menu archive; and the currently running “Lunch Hour NYC.”

Culinary and social historians consult the collection frequently, as do novelists looking for period detail (e.g., the price of a pint of beer on the Bowery in 1870), and of course chefs seeking gastronomic inspiration. For most of the collection’s history menus were arranged in boxes by date, and for years this was the only way researchers could request them. A decade or so ago, on-site volunteers created bibliographic records for approximately two thirds of the menus, resulting in an online database where users could query by keyword, enabling them to search for particular restaurants, social organizations or other peculiarities, and to request specific menus by call number.

Anyone wanting to investigate the actual food in the menus, however, had to pore through them by hand and see what they stumbled upon. There was no way to ask questions of the collection as a whole, for example, to request all menus with “abalone” or “macaroni,” or to observe the changing price of coffee over the decades, or to uncover the first appearance of instant coffee or diet soda. A few researchers had made such queries the old-fashioned way: going through thousands of menus individually by hand. In 2005, an oceanographer from Texas A&M University meticulously traced the fluctuating prices of seafood over decades of menus, making inferences regarding the health of local fisheries over time. A culinary historian from Yale University conducted a close reading of thousands of late 19th century New York hotel menus to paint a more nuanced picture of high-end dining in that era.

But to all but these brave few, the archive remained opaque, its data riches still largely untapped, despite the fact that nearly a quarter of the collection was freely available online via the NYPL Digital Gallery. A librarian, who specializes in culinary collections, along with the curator of the Rare Book Division³ (where the menu collection is housed), began to look into what it would take to get full-text access to this already digitized set of menus.

OCR (optical character recognition) scanning would undoubtedly get usable output from a portion of the menus. But many items are handwritten, or printed in idiosyncratic type fonts or fanciful layouts. Moreover, as initial conversations progressed in Labs, we agreed that what was of most interest was building a structured data set out of the menus, where each individual
dish had an instance (and, frequently, a relationship to a price), and that these instances could be tracked, aggregated and visualized. Given these aims, manual transcription into a predefined schema seemed the best way to produce a high quality data set. But who would carry out this immense work? Drawing on our modest user collaboration through the Map Warper, and considering the persistent public interest in the menus (and in all things culinary), this seemed like the ideal project to push the crowdsourcing experiment further. And so What’s on the Menu? (WOTM) was born (2011a).

A no frills beta version launched in late April of 2011 and was an immediate success. From the get-go, Labs had made usability of the core transcription tool the top priority (see Figure 5). We were banking on the task being not only simple and easy, but ideally fun (and maybe even a little bit addictive). Technical resources were tightly limited (the WOTM menu was built entirely as a spare time project by a few developers on NYPL’s Web team), and there was no official project staff (as in the Map Division) to keep the work going if public participation waned. So getting the tool right, and broadcasting a clear motivational message (“Help The New York Public Library improve a unique collection”), were essential. Taking the pulse of the Twitter feeds in the days and weeks following launch, and observing how quickly users were making their way through the menus, Labs knew it had hit the mark.

By mid-summer, Labs was nearly finished with the initial batch of digitized menus and were scrambling to begin moving the rest of the collection in the direction of the scanners. The sudden real-time public demand for fresh digital content was something NYPL had never experienced in quite
this way, where there was something very specific users wanted to do with the collection, and right now. The site continued to draw attention from press, blogs and social media. A class of deaf fourth graders in San Antonio were even using the site to practice typing skills. It was no small feat to re-prioritize digitization queues, re-assign staff, and establish a reliable new process for maintaining steady stream of new menu content to the site amidst competing priorities. And although these efforts were largely successful, thanks in part to generous start-up assistance from the National Endowment for the Humanities and the Institute of Museum and Library Services, Labs still periodically had to endure content droughts, sometimes a week or more, during which we would be assailed by benevolently impatient emails and tweets from the hard core users.

Unlike the Map Warper, which requires users to create an account in order to take part in the georectification and feature digitization tasks, WOTM does not require (or even offer) registration. This was a deliberate choice at the outset, to keep the barriers to participation as low as possible. Labs built in basic security measures using browser session tracking to prevent abuse or malicious scripts. To this day, we have seen virtually no instances of vandalism and the general quality of transcription is exceptionally high. Menus go through a basic workflow. “New” menus are fresh from the scanners and ready to be transcribed. When a user finishes a menu, they promote it to the “Under Review” category, where it remains editable and project interns and other users can look it over for mistakes or missing data. From there, it gets nudged into the “Done” pile, where editing is turned off. Users can always email the site administrators if they spot further errors, and the menu’s status can be changed.

Despite the success of the low-barrier approach, the staff in Labs sometimes lament that we were not better able to recognize the contributions of the top transcribers, or to develop a hierarchy of tasks where more challenging work might be offered to the more experienced participants. Beyond raw analytics, all of the user stories are anecdotal, but we do nevertheless have a sense that the project tends to attract people with a strong culinary or library science interest, that our most vocal users (the ones who most frequently email us with software bugs, bibliographic errors or clamoring for new menus) are women, and that judging by the traffic logs, transcription happens all throughout the day. We also know through Web analytics that the project, while drawing predominantly from the English-speaking world (about 75% United States), has an audience that far transcends New York. A little less than a quarter of overall visits originate within the state.

Building on the Map Warper’s success, WOTM has undoubtedly impacted the internal conversation at NYPL around digital strategy, user engagement and collections policy. It has helped shift the attention, at least in part, away from static online exhibitions, which notoriously struggle to retain users’ attention, toward participation-oriented Web sites with longer life
cycles, involving the public in the grand (if painstaking) work of building the digital library. It has also jumpstarted policy discussions around user-contributed content and its relation to Library-authored metadata.

In the spring of 2012, with the Labs team established, the new developers began a total revamp of the WOTM site. The functionality of the transcription tool in it was good, but the beta version lacked features that would enable users to browse, search and visualize the collection, and the enormous heap of data that had been harvested from it. There was also a certain amount of “technical debt,” incurred from the quick-sketched coding of the pilot phase, that had to be paid down in order to get the site on a more sustainable track.

The new WOTM site (see Figure 6), redesigned and re-engineered, also offered users access to the data set in a variety of ways. Bi-weekly exports (containing all dishes, prices, page coordinates, and bibliographic data for each menu) are available for download. Labs also created NYPL’s first publicly promoted API (application programming interface), providing more technically advanced users programmatic access into the data set. In the months since release, dozens of requests have come in for API access,
representing initial interest from a variety of constituencies ranging from food-focused tech startups, to computational linguistics researchers, to journalists, to library and museum hackers.

The future applications of the WOTM data remain to be seen. Undoubtedly, the crowdsourcing effort has raised the profile of the collection many times over, landing it frequently in the press over the past two years, and consistently generating small ripples through the culinary and techie social media subcultures. It also has radically enhanced the accessibility of the collection. A perusal of keyword-driven traffic to the site reveals a plethora of fascinatingly obscure “long-tail” searches that have brought users serendipitously to our menus:

- “eggs argenteuil”—a scrambled egg preparation found in 1910, reappearing in 1961 (88 visits);
- “remus fizz”—a citrusy gin cocktail with egg whites, mid-century (40 visits);
- “moskowitz and lupowitz”—a vanished Romanian-Jewish eatery from the Lower East Side (23 visits);
- “ss Homeric”—an ocean liner, originally German, relinquished to Britain as war reparations in 1919 (16 visits).

By these and other measures, we can witness the collection’s weaving into the fabric of the Web.

THE STEREOGRANIMATOR: REMIXING COLLECTIONS

Completing the trilogy of projects diving deeper into existing digital collections, the Stereogranimator (2012) stands out of the NYPL Labs portfolio in that it was inspired not by a curator but by a user. Joshua Heinemann, a San Francisco-based artist and writer, had caught the attention of NYPL Labs several years ago with a delightful art project on his Web site, *Cursive Buildings*. Rolled out gradually over several years, “Reaching for the Out of Reach” (2009) is a series of eerie, flickering animations, rolled out gradually over several years, created out of the NYPL’s vast collection of stereographic photography. Stereographs are pairs of images taken from slightly different angles, recreating binocular vision to produce an arresting 3D effect. Without a handheld stereoscope device, however, the images appear flat and two-dimensional, and it is in this diminished artifactual form that they are presented on the NYPL Digital Gallery—over 40,000 in total. Like the map and menu collections, with their valuable but difficult-to-extract data, the digitized stereographs, despite now being more accessible via the Web, were as remote as ever from their original intended user experience.

Heinemann’s idea was to reacquaint the viewer with the original immersive nature of the images in a Web-native way. To do this, he employed
one of the oldest and hardest forms of Internet multimedia: the animated gif. Animated gifs are nothing more than a sequence of two or more images, typically run as an infinite loop. In the early days of the Web, long before Flash, HTML 5 or streaming video, animated gifs were wildly popular with the stewards of hand-crafted HTML home pages, adding a splash of visual dynamics to otherwise static, text-heavy screens. Heinemann’s gifs toggle between the stereo images, tweaked to just the right degree of parallax, producing twitchy cinematic fragments with alluring depth in which the subjects (or the scenes around them, depending on the settings), seem to tremble like mysterious ejecta from the past.

So-called ‘wiggle gifs’ have been employed before to convey stereoscopic effects on digital screens (they are a popular byproduct of NASA’s Mars rover photography, for example). But Heinemann’s images ingeniously remind us of the original impact of this once popular art form, and also that the Web too has been around long enough to have its own trail of forgotten artifacts. Vernaculars a century apart unite in the glitchy aesthetics of Heinemann’s series of 60 animated stereo-gifs.

“Reaching for the Out of Reach” is an example of the sorts of wondrous remixes and derivative works that can arise when collections are made accessible on the Web, simultaneously reinterpreting them and drawing public attention back to the originals. The NYPL has always looked kindly on these sorts of activities, but what if it were to actively encourage them, building tools and services expressly for creative reuse? When the Labs team convened, we set about investigating whether a simple, browser-based tool could be built that would allow even a novice user to quickly create gifs from our stereographs and instantly share them via the Web.

Faster than expected, an intuitive interface was developed, and to it Labs added the option of creating (via the same interaction) red-cyan anaglyphs (glasses required), for the 3D connoisseurs (see Figure 7). In the lead-up to launch, Labs contacted Heinemann to let him know that his work had inspired our latest project and invited him to write a brief blurb or artist’s statement to accompany the site. He responded with a lengthy essay on his lifelong relationship with stereographs (his father had employed aerial stereoscopy in his work as a state forester in Minnesota) and his accidental discovery of their potential effects as gifs in Web browsers. Heinemann’s essay became the introduction to our site, and a longer version was published in The Huffington Post the day the site launched.

The Stereogranimator enjoyed a bigger response in its initial days than any other NYPL micro-site. In one month, it racked up more visits than all of the NYPL’s static online exhibitions put together over the previous year. Media coverage was extensive and spanned arts, design, science and pop culture domains. Admittedly, this attention was of a more transient nature, fitting the viral meme-ish nature of the project, and traffic has slowed to a small but persistent trickle in the months since. But the
Stereogranimator powerfully demonstrated a new disposition toward the NYPL’s patrons, explicitly inviting them to make new works out of old ones via tools that were themselves inspired by user creativity. It is also suggestive of the sorts of inventive apps that might be built when NYPL eventually opens up all of its digital collections through APIs.

In the months since launch, the site has steadily accumulated new user creations, now totaling over 34,000. In the spring of 2012, the Labs team built a connector to the Flickr API, allowing the inclusion of over 600 stereographs from the Boston Public Library’s collections. In this modest way, Labs modeled how one library might build a tool that many collections can benefit from. Early discussions are currently underway with other great photography archives, exploring similar collaborations.

In the interest of working fast with minimal constraints, NYPL Labs’ apps have mainly been built outside of the NYPL’s infrastructure, served from the cloud on “platform-as-service” hosts such as Heroku and Amazon Web Services, which are better suited to a more iterative development style. This has helped us to build a portfolio quickly, sidestepping some of the usual institutional inertia, but it is not a sustainable strategy. Working outside NYPL’s
infrastructure enables the developers to work fast, but with each successive project, it adds administrative overhead to the managers, who must keep track of a range of systems duties usually handled by dedicated specialists. Fortunately, these pilots have helped to accelerate discussions already underway in the IT department about moving systems architecture to the cloud, and embracing a more agile approach to software development. In this way, Labs has also served a technical R&D function for the Library, alongside its more visible experiments in user collaboration and data transformation.

Audience outreach has also been largely do-it-yourself, or DIY. Currently, Labs projects are not visibly tied to the Digital Gallery, whose contents they draw so heavily upon, or any core NYPL platform beyond a basic information page on the main NYPL Web site (2011b). Each project therefore develops its own pocket community, rather than drawing from and expanding upon the broader NYPL user base. Our sites have attracted vigorous participation in spite of these obstacles, but it is hard not to wonder how the impact might multiply if these projects were placed more at the heart of the NYPL's Web experience.

We may find out before too long. NYPL is now in the early stages of implementing a new “Virtual Library” plan, which among other things will support the development of a crowdsourcing platform, where projects like the Map Warper, What’s on the Menu? and the Stereogranimator will be more prominently featured and tracked in a kind of public participation dashboard. Inspired by initiatives such as the Citizen Science Alliance's Zooniverse (2007), the National Archives' Citizen Archivist Dashboard (2011), and the University of Iowa Libraries' recently launched DIY History site (2012), the still-to-be-named NYPL crowdsourcing platform will weave participatory projects and tasks into a cohesive user experience, promoting a culture of online participation at NYPL. A high visibility hub of this nature not only can attract more volunteers, but better exposes the resulting data.

The data too will soon have a new, more centralized home. As the Digital Library Program was winding down in 2008–2009, a massive Digital Repository project was begun to lay down infrastructure for the long-term preservation of NYPL’s digital assets. Now nearing completion, the repository will provide not only trusted, redundant storage of all digital collections and data, it will serve content into all NYPL Labs applications, as well as whatever ends up taking the place of the Digital Gallery. Our hope is that eventually the Repository will serve this content to external developers and researchers via a suite of open APIs, turning NYPL into a full-fledged technical platform for the public to interpret, improve and build upon: a data clearinghouse for digital humanists.4

Labs' initial success has strengthened our experimental mandate, but we have also inherited some larger digital access projects which had been previously stalled by departmental reorganizations. As we prepare a new round of participatory apps, the team (now expanded to four developers), is currently
at work on an archival data platform that will power a new Web-based finding aid interface and digital asset viewer. Labs is also building a new video platform, piloted around deep audiovisual holdings in the Library’s Dance Division, that will replace antiquated analog playback consoles with high definition streaming Web video.

Managing these two work streams—developing small, innovation-focused apps alongside larger access platforms—can be tricky, but ultimately the two worlds reinforce one another. The lessons learned in our more imaginative crowdsourcing sites infuse the core projects with a more user-centered philosophy, improving access to the collections while building tools and methods for their enhancement. The dance video site, for example, will include tools for user annotation, juxtaposition and mashup of time-based content. Similarly, the new archives platform, following an initial beta release in early 2013 with several signature U.S. history collections, will eventually include tools by which users can augment collections with item-level metadata, georeferencing, document transcriptions and other value-adding information.

Digital collections allow Labs to be more bold with materials—at times even promiscuous and playful, as with the Stereogranimator. But as countless institutional and personal data losses testify, digital assets can be alarmingly ephemeral. We speak of “trusted” repositories and “long-term” preservation of digital content, but the difficulties encountered in accessing even decade-old formats make us duly cautious. The track records of such systems just are not long enough to shake off the scare quotes. But the more that digital collections are used—the more data they accrue, the more they are copied, linked, repurposed, remixed and remediated—the better their chances of survival in the volatile digital medium. As Bethany Nowviskie elegantly and succinctly puts it: “public access is tactical preservation” (2012).

For NYPL, this suggests a powerful rediscovery of its mission. We are coming to see crowdsourcing not only as a way to accomplish work that might not otherwise have been possible, but as an extension of our core institutional duties. It is, in Library of Congress digital archivist Trevor Owens’ words, about “offering our users an opportunity to participate in public memory.” At a time when libraries have been massively disrupted by new information technologies, when we are only beginning to get a handle on the new mechanisms of memory, deepening our bonds with the public through networked collaboration literally helps us to remember our purpose—and our stuff.

NOTES

1. The term “skunkworks” originates from a secretive subdivision of Lockheed Martin producing experimental aircraft designs beginning in World War II. The name itself is a now-obscure reference to the *Li’l Abner* cartoons popular at the time. It is generally understood today to refer to small innovation units within an organization operating outside normal procedural constraints, and generating new products or
methodologies to be absorbed into general practice. In this issue, Bethany Nowviskie elaborates on the idea of digital skunkworks in a library context.

2. Matthew Knutzen is NYPL’s geospatial librarian and director of the Map Warper project.

3. Rebecca Federman, NYPL Culinary Collections Librarian; Michael Inman, Curator of NYPL Rare Book Division.

4. David Weinberger of Harvard’s Library Innovation Lab has elegantly sketched the notion of the “library as platform,” modeling in part on technology companies like Facebook who have opened their data to outside developers, but also delineating more library-specific aspects (geographically defined communities, more emphasis on end-users as opposed to software developers). See: Weinberger, David. (2012). Library as Platform. Library Journal. Retrieved from http://lj.libraryjournal.com/2012/09/future-of-libraries/by-david-weinberger/

REFERENCES


