The Human Face of Crowdsourcing: A Citizen-led Crowdsourcing Case Study

Sheryl Grant
School of Information and Library Science (SILS)
University of North Carolina Chapel Hill
Chapel Hill, NC, USA
sherylgrant@gmail.com

Kristan E. Shawgo
HASTAC
Duke University
Durham, NC, USA
kristan.shawgo@duke.edu

Richard Marciano
School of Information and Library Science (SILS)
University of North Carolina Chapel Hill
Chapel Hill, NC, USA
richard_marciano@unc.edu

Jeff Heard
Renaissance Computing Institute (RENCI)
University of North Carolina, Chapel Hill
Chapel Hill, NC, USA
jeff@renci.org

Priscilla Ndiaye
Southside Advisory Community Board
Asheville, NC, USA
priscillapower2@gmail.com

Abstract— The Cyber-Infrastructure for Billions of Electronic Records (CI-BER) project is a collaborative big data management project based on the integration of heterogeneous datasets and multi-source historical and digital collections, including a place-based citizen-led crowdsourcing case study of the Southside neighborhood in Asheville, North Carolina. The project is funded by the National Science Foundation and the National Archives and Records Administration (NARA) agencies. A test-bed collection containing nearly 100 million files and 50TB of data was developed, with content representing electronic Federal Government records from 150 federal agencies. The CI-BER project advances the state of the art in generalizable and extensible ultra-highly scalable data management architectures, potentially enabling robust technical preservation of, and access to, electronic records and digital data in the context of emerging national scale cyber-environments. A first-generation open source collaborative mapping environment prototype is currently being developed to support novel "citizen-led crowdsourcing" possibilities for archival material.

Keywords: Crowdsourcing, citizen-sourcing, Southside, Asheville, Housing Authority of the City of Asheville, Asheville NC, urban renewal, African-Americans

I. INTRODUCTION

The Cyber-Infrastructure for Billions of Electronic Records (CI-BER) project is a collaborative big data management project based on the integration of heterogeneous datasets and multi-source historical and digital collections, including a place-based citizen-led crowdsourcing case study of the Southside neighborhood in Asheville, North Carolina. The project is funded by the National Science Foundation and the National Archives and Records Administration (NARA) agencies. A test-bed collection containing nearly 100 million files and 50TB of data was developed, with content representing electronic Federal Government records from 150 federal agencies. The CI-BER project advances the state of the art in generalizable and extensible ultra-highly scalable data management architectures, potentially enabling robust technical preservation of, and access to, electronic records and digital data in the context of emerging national scale cyber-environments. A first-generation open source collaborative mapping environment prototype is currently being developed to support novel "citizen-led crowdsourcing" possibilities for archival material. Concurrent with the development of this prototype was the co-creation of a citizen-led crowdsourcing protocol to solicit design input from community members, a process that represents a viable technique to deal with "big data" processing and management. The connection between these two prototypes lies in the ability to expose relevant CI-BER datasets (1930 Census, 1937 redlining data, 1940 Census, etc.) to the "crowds" and allow them to tag and augment this content with additional annotations and knowledge.

II. CITIZEN-LED CROWDSOURCING

Crowdsourcing is "the process of leveraging public participation in or contributions to projects and activities" [1] and is carried out in different ways depending on the community involved, the content to be crowdsourced, and the technology available. What sets this project apart from a straightforward digitization or visualization project is its deployment of citizen-led crowdsourcing, in which the community assumes an essential role in the design and implementation process. The CI-BER study followed a four-phase crowdsourcing process designed with and for the Southside Asheville community: 1. crowdsourcing framework, 2. community history, 3. remapping process, and 4. citizen-led crowdsourcing process.

A. Crowdsourcing Framework

Drawing on *what citizens know* and *want to know* is instrumental to CI-BER's implementation of crowdsourcing. The involvement of Asheville's Southside community in co-creating and designing the process was of paramount importance to its success. Dunn and Hedges [1] describe three types of crowdsourcing approaches: contributory, collaborative, and co-created. CI-BER proposed a process of co-creation, in which the community was actively involved in most or all steps of crowdsourcing, which framed their engagement as citizen-led crowdsourcing. Energetic and vital participation in an earlier North Carolina Humanities Council (NCHC) project, Twilight of a Neighborhood, [2] indicated that the community would willingly participate in a citizen-led crowdsourcing process.

CI-BER's employment of "citizen-led crowdsourcing" is inspired by the Obama administration's Open Government Initiative [3], which encourages public participation and collaboration. It is a derivative of the term citizen sourcing, defined as the "government adoption of crowdsourcing techniques for the purposes of (1) enlisting citizens in the design and execution of government services and (2) tapping into the citizenry's collective intelligence." [4] Vivek Kundra, Chief Information Officer of the United States from March 2009 to August 2011 under President Obama, described citizen sourcing as a way of "tapping into the ingenuity of the American people to solve those problems that are too big for government to solve on its own." [5]. Citizen sourcing is derived from the term crowdsourcing and emphasizes the type of civic engagement typically enabled through Web 2.0 participatory technologies, over a more impersonal crowd-based distributed problem-solving and production model.

Hilgers and Ihl [6] suggest that "external collaboration and innovation between citizens and public administrations can offer new ways of citizen integration and participation, enhancing public value creation and even the political decision-making process." This type of collaboration extends to what the Archivist of the United States, David Ferriero, called "citizen archivists"[7] in reference to citizen scientists, as a means to increase public engagement in the archives given the National Archives and Records Administration's over-abundance of paper records and the need to digitize and transcribe them. A participatory archive is "an organization, site or collection in which people other than archives professionals contribute "knowledge or resources, resulting in increased understanding about archival materials, usually in an online environment."[8]

This definition relates to the notion of "citizen-led sourcing" implemented in the CI-BER project, putting civically engaged community members at the forefront so that the focus is on their collective engagement with the archive.
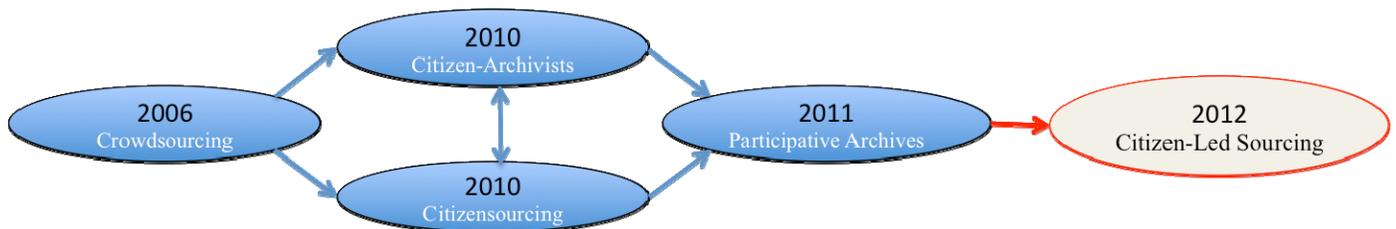

Fig. 1: From crowdsourcing to citizen-led sourcing (dates indicate when concepts were first introduced)

B. Community History

The University of North Carolina-Asheville (UNC-A) Ramsey Library's Special Collections & University Archives acquired the Housing Authority of the City of Asheville records in July of 2007. The nearly intact and complete collection of urban renewal documents [9] includes approximately 130 linear feet and 129 cartons that document a number of "significant redevelopment projects undertaken from the early 1960s to the mid-1980s." CI-BER focuses on Asheville's Southside area, which at over 400 acres was the largest urban renewal area in the southeastern United States. In 1966, the Southside community

represented about 4,000 people, 98% of whom were African American, living in nearly 1,300 households, and comprising 7% of Asheville's population. In 2008, the North Carolina Humanities Council (NCHC) sponsored a project that documented the community's life prior to urban renewal [2].

Urban renewal as a federal government program was a 24-year (1949-1973) initiative started under the Housing Act of 1949, and modified under the Housing Act of 1954. It used the 1930's Home Owners' Loan Corporation (HOLC) redlining terminology of "blight" and "slums" to launch an ambitious redevelopment and eminent domain process that led to the bulldozing of some 2,500 neighborhoods in 993 American cities. It is estimated that one million people were dispossessed in the process. [10]

C. Remapping Process

Maps are a natural visualization method for presenting the changes across neighborhoods. However, GIS applications focus primarily on structured data, analytics, and thematic mapping, and require thorough grounding in geographic principles to be used effectively. A platform that supports community mapping must be able to handle structured data and unstructured data equally, and must allow for collaborative editing of content by non GIS-experts. These requirements are driven by the project itself and the nature of the data collection.

The goal of CI-BER was to recreate an entire collection in the very first iteration, to capture key elements (owner, renter, parcel number, street address) per scanned documents (initial appraisal sheet, and house photo) from the "Housing Authority of the City of Asheville Records" at UNC-A [11] and rapidly remap the entire Southside neighborhood. This information was combined with fragments of maps found in the collection to create a digital spatial canvas of the entire neighborhood in 1965 prior to urban renewal. Each digitized parcel is clickable and linked to its key elements. Being able to reproduce this process online and identify and tag collection items is a key output of the citizen-led sourcing of the archive. Through iterative and incremental passes over the files, the CI-BER team gradually digitized strategic content, transcribed it through citizen-led crowdsourcing, visualized and mapped the content, enhanced the collection, and developed a working content model, and added functionality to the software user interface being built.

The software supporting this process is Big Board, [12] part of RENCI's Geoanalytics [13] geospatial cyberinfrastructure. Big Board is a web-based collaborative mapping interface that enables its users to juxtapose traditional GIS data with unstructured data via a set of drawing tools and an underlying content management system (CMS). Originally envisioned as a decision support tool for emergency managers [14]-who need more flexibility and timeliness than traditional GIS affords, Big Board was repurposed to support the cooperative work of digital humanities scholars and Asheville community members [15]. In the Big Board, users collaborate in a virtual data room that allows GIS maps to be layered onto web-maps and users are able to draw on and edit content, seeing changes in real time.

CI-BER collaborators began by pulling in an historical 1960s map of the Asheville area, georeferencing it in a GIS system, and adding it to a Big Board room. Modern layers from the Asheville city GIS system such as street centerlines, tax parcels, and addresses were then added to support and provide context to the historical content. From this basis, collaborators drew individual polygons using the Big Board's drawing tools that represent tax parcels from the historical map.

Big Board software associates an editable web page from the CMS with each polygon. Web pages were edited in the Big Board's WSIWYG editor adding the aforementioned key elements: owner, renter, parcel number, street address and collaborators can append scanned pages, photographs, audio clippings, or additional descriptive text to each page. In the Big Board map interface, the polygons are now clickable links on the map that take a user to the created content (and allow editing if the user has sufficient permissions).

This process allows incremental, "schema-less" editing of content by the CI-BER collaborative team whereas traditional GIS setups would enforce a schema to start with and force collaborators to email documents back and forth and work from a "master copy," changing the process and leaving it more error-prone.

D. Citizen-led Crowdsourcing

Critical to citizen-led crowdsourcing is the collaboration of a key community member who contributed to the design of Big Board on behalf of the Southside residents. The key member of the CI-BER project team is the chair of the Southside Community Advisory Board in Asheville; as a resident of the Southside neighborhood during the time of urban renewal and a leader still residing in the community today, her feedback provided an invaluable connection to the community and its history. Building and maintaining trust with the community cannot be understated in terms of its significance to the credibility and success of a project like CI-BER. Having a leader from the Southside Community Advisory Board represent the project at community events, discuss it locally via word of mouth, give interviews to local news [16], and introduce CI-BER at a "Remapping Community" information session on August 3rd, 2013 ensures that the project is citizen-led.

"Remapping Community" was the CI-BER project's initial introduction to the community and included information about the process of urban renewal in Asheville, an introduction of the CI-BER team to the community and a demonstration of Big Board's capabilities and possibilities for preserving and displaying the neighborhood's African American history through varied data streams (archival data, text, image, audio, and video). The information session concluded with the Southside neighborhood's civically engaged community members contributing to the design and input for execution of the next steps of data collection to be appended to Big Board, including the collection of photographs, written documents and oral histories. Central to this process is the importance of understating what citizens' motivations are and how to sustain their interest. We hope this project will also lead to insights in this area.

III. CONCLUSION

Nationwide, thousands of cities followed the same urban renewal development patterns that the Southside neighborhood in Asheville, North Carolina experienced, and this case study naturally leads to scaling at a very large national level. By focusing on the Southside neighborhood in Asheville, North Carolina, CI-BER demonstrates the potential for automation and integration of temporal and spatial datasets involving census, economic, historic, planning, insurance, scientific, and financial content, with an eye on scalability and the goal of making a national impact on future research.

REFERENCES

[1] S. Dunn and M. Hedges. (2012). "Crowd-sourcing scoping study: Engaging the crowd with humanities research." Centre for e-Research, Department of Digital Humanities King's College London. Arts & Humanities Research Council. A project of the AHRC Connected Communities Theme. Retrieved from http://humanitiescrowds.org/2012/12/21/final-report/

[2] S. Judson (2010). Twilight of a neighborhood: Asheville's East End, 1970. Crossroads: A Publication of the North Carolina Humanities Council. Summer/Fall 2010. Retrieved from http://nchumanities.org/sites/default/files/documents/Crossroads%20Summer%202010%20for%20web.pdf

[3] Open Government Initiative of The White House. http://www.whitehouse.gov/open

[4] T. Nam. (2012) "Suggesting frameworks for citizen-sourcing via Government 2.0." Government Information Quarterly, vol. 29(1), pp.12-20, doi:10.1016/j.giq.2011.07.005

[5] V. Kundra. "How 'Open Gov' datasets affect parents and consumers." White House Open Government Blog. January 23, 2010. Retrieved from: http://www.whitehouse.gov/blog/2010/01/23/why-open-gov-matters-you.

[6] D. Hilgers and C. Ihl (2010). "Citizensourcing: Applying the concept of open innovation to the public sector." The International Journal of Public Participation, vol. 4(1), pp. 67-88. Retrieved from http://search.ebscohost.com.proxy.lib.duke.edu/login.aspx?direct=true&db=poh&AN=48757187&site=ehost-live&scope=site

[7] D. Ferriero (2013). "Volunteers help NARA do its job, support professional archivists." Archival Outlook. January/February 2013, p. 16. Society of American Archivists.

[8] K. Theimer, "Exploring the Participatory Archives", on ArchivesNext blog, August 30, 2011, http://www.archivesnext.com/?p=2319.

[9] City of Asheville Donates Urban Renewal Files to UNC Asheville. October 23, 2007. http://www2.unca.edu/news/releases/2007/renewal.html.

[10] M. Fullilove. Root Shock: How Tearing Up City Neighborhoods Hurts America, and What We Can Do About It. New York: One World/Ballantine Books, 2004.

[11] "Housing Authority Records of the City of Asheville." Acquired by University of North Carolina Asheville. July 2007. http://toto.lib.unca.edu/findingaids/mss/housing_authority_city_asheville/default_haca.htm

[12] J. Heard, S. Thakur, J. Losego, K. Galluppi. "Big Board: Teleconferencing Over Maps for Shared Situational Awareness." Journal of Computer Supported Cooperative Work (JCSCW). Springer, July 2013. doi:10.1007/s10606-013-9191-9

[13] J. Heard. "Geoanalytics." Tech Report TR-11-03. http://www.renci.org/publications/technical-reports

[14] K. Galluppi, University of North Carolina, Chapel Hill, NC; and J. Heard, J. L. Losego, B. E. Montz, C. F. Smith, and S. S. Schotz "Weather Service and Emergency Management Collaboration Through An Interactive Web-Based Map Conferencing Infrastructure." AMS 2012.

[15] J. Heard, R. Marciano. "The Big Board: A Collaborative Environment for Geography." Virtual School on Computational Science and Engineering (VSCSE), Data Intensive Summer School, workshop taught on July 8, 2013, http://www.vscse.org/summerschool/2013/bigdata.html

[16] R. Neufield. "Project Seeks to Recreate History of Asheville's Southside," August 2, 2013. http://www.citizen-times.com/article/20130804/LIVING/308040026/Project-seeks-recreate-history-Asheville-s-Southside.