

# The United States Children's Bureau Papers Project: a short introduction

Shane Landrum, Brandeis University

June 8, 2012

## Introduction and Goals

The United States Children's Bureau Papers Project is an experiment in collaborative digitization, description, transcription, and annotation for a major federal collection in United States women's history. Built on the the NEH-funded Scripto tool for crowdsourced transcription (<http://scripto.org>), it will enable researchers, teachers, and students to collaborate in making important primary sources available to a wider audience.

Researchers are increasingly using digital photography to create our own reference-quality collections of archival materials. The USCBPP seeks to enable us to share those collections with one another, add metadata to them, and create open-access digital versions for research and teaching purposes.

## About the Collection

The Records of the United States Children's Bureau is a minimally-processed collection housed in Record Group 103 at the National Archives and Records Administration at College Park, Maryland ("Archives II"). The United States Children's Bureau, founded in 1912, was the first federal agency specifically to focus on the health of children and mothers. Between its founding and the early 1940s, the bureau pioneered many of the programs which became part of the federal welfare state. The scope of this project extends to 1947, when a major federal reorganization changed the Bureau's work significantly.

The Children's Bureau papers are a major research collection for the history of United States women and the US welfare state. They are one of the largest single collections of manuscript correspondence about the experiences of early 20th century American mothers, especially rural women. The collection is also valuable for studying grassroots public health movements, transnational women's activism for maternal and child health, and early 20th-century family law reform.

The bulk of the collection from 1912-1947 is textual, though photographic and film materials are also available. Correspondence from constituents and from Children's Bureau employees working in the field is often handwritten; replies from the Washington office are usually typed.<sup>1</sup>

This collection is open to the public, contains no classified material, and is largely in the public domain.<sup>2</sup> Unlike other NARA collections which have more potential for digitization by commercial partnership (e.g., military pension records, manuscript census records), RG 103 is unlikely to be a financially attractive target for commercial digitization anytime soon. As a result, grassroots digitization and transcription methods are a promising alternate approach.

## Previous Work and Significance

The existing 290-reel microfilm edition of the Children's Bureau papers, published by Chadwyck-Healey in 1993, is exceedingly partial. Its selection illuminates the institutional history of the agency and its major initiatives but omits much of the texture of ordinary women's letters to the Bureau. It also contains no subject index.

Because this project is an outgrowth of my research, it's designed to require minimal funding and to be volunteer-staffed by anyone who's interested in helping.<sup>3</sup> Historians have not engaged deeply enough with open-source models of software development and knowledge creation, which allow contribution by anyone and rely on the labor of interested, skilled volunteers.

---

<sup>1</sup>This fact is significant because it enables a certain amount of rough-and-ready indexing via optical character recognition. Shane Landrum, "OCR'ing archival research photos with DevonThink Pro Office," October 11, 2001, <http://cliotropic.org/blog/2011/10/ocring-archival-research-photos-with-devonthink/>.

<sup>2</sup>There's some copyright gray area about the constituent correspondence, since many letter-writers may not have been dead for the 72 years technically necessary for copyright to expire. I plan to treat them as orphan works and to implement a takedown-upon-request policy.

<sup>3</sup>Individuals and organizations contributing to the project will have their names and contributions enumerated on the site's credits page. Bethany Nowvickie, "Where Credit is Due: Evaluating Collaborative Digital Scholarship," May 31, 2011, <http://nowvickie.org/2011/where-credit-is-due/>; Tanya Clement and Doug Reside, "Off the Tracks: Laying New Lines for Digital Humanities Scholars," [http://mith.umd.edu/wp-content/uploads/whitepaper\\_offthetracks.pdf](http://mith.umd.edu/wp-content/uploads/whitepaper_offthetracks.pdf).

Most existing subscription-based digital collections reflect the fact that historians value formally-trained expertise in matters of interpretation. However, using software-enabled crowdsourcing methods, significant portions of the labor of archival processing can be broken down into tasks that can be done by interested amateurs, including students.<sup>4</sup> Deploying such a system has potential for reducing the repetitive organizational labor currently done by skilled archivists and making significant portions of the Children's Bureau papers accessible to a wider audience.

## **Major Components**

I propose to build a website where scholars who do in-person research with the Children's Bureau papers can upload their images and metadata in a structured way, describing the locations of the original objects and the subjects they were researching when they shot the images. Visitors to the site will be able to participate in crowdsourcing transcriptions, additional metadata, and interpretive comments.

Such a system will be useful for other scholars, even if admittedly partial, and it can be used when we teach our students about archives to help them understand transcription and other skills of reading handwritten documents. For the general public, the project will offer the possibility of finding their grandmothers' and great-grandmothers' letters to the Bureau, which were previously unindexed and unavailable.

As the project grows, it should develop mechanisms for peer review to verify transcriptions and mark them as approved by experts. The scholarly rigor of the collections will evolve by user participation rather than by restricting access to materials until they're verified.

## **Big Data for Women's History**

Over time, the ultimate goal of this project is to enable the collaborative creation of data sets to be used for algorithmic methods of analysis. Examples might include:

- adding geocoded metadata to correspondence, which will enable researchers to find all letters written by women from a particular state, or to create GIS-enabled visualizations of the rural travels of particular Children's Bureau fieldworkers.<sup>5</sup>
- an authority-controlled list of correspondents, which will enable social network analysis of the grassroots women's organizations that the Children's Bureau used to build support for its policy initiatives.

This data should be freely-downloadable for use by scholars; it should be versioned and preserved much like scientific data, so that it can be cited by researchers unaffiliated with the project itself.

## **Status**

This project is a spinoff from my dissertation research on the history of birth certificates and birth registration in the United States. While revising and completing my dissertation in 2012-13, I'll begin writing the framing explanatory documents targeted at archivists and researchers, and I'll build a proof-of-concept site. The initial image contributions will begin with the roughly 20GB of reference images I've collected during my research trips to College Park. I'm hoping to have an early version ready for use by students in my spring 2013 course on US women's history.

I'm actively seeking scholars, including archivists, who want to participate in the framing of this project so that it can grow into a sustainable, multi-researcher collaborative endeavor. Technologists with skills in bulk handling of images and metadata, repository software, and virtual research environments would be particularly welcome.

In my talk about this project, I'll highlight some of the most exciting documents I've found in the Children's Bureau papers, explain some exciting possibilities as well as some concerns, and share more details about its current status. I'm especially interested in hearing from workshop participants about how to plan a project of this scale and build a workable all-volunteer team for the project's early phases.

---

<sup>4</sup>For a few existing systems, see the papers given at "Crowdsourcing History: Collaborative Online Transcription and Archives," session at American Historical Association 2012 annual meeting, <http://crowdsourcinghistory.wordpress.com/>.

<sup>5</sup>I've pursued these approaches at a much smaller scale using existing tools; they've enabled me to answer important questions for my dissertation research.