

## 7. Data Management Plan

### i. Roles and responsibilities

CWBC data management and maintenance will be handled by co-PDs Pitti and Martin (the co-directors of IATH), in conjunction with the IATH staff. IATH has a full-time system administrator who provides additional support for all IATH project data and for maintenance and publishing software. Martin and Pitti will oversee data management and will monitor preservation and dissemination efforts for the grant period. CWBC is part of PD Booth's Collective Biography of Women IATH Fellowship project, so IATH will continue to maintain and provide access to both CBW and CWBC data for the foreseeable future. IATH's general policy is to migrate and update project data, to the degree that technological evolution allows, so that projects remain viable well beyond the original project window. Long term decisions about data management are the responsibility of the IATH director and technical staff.

SNAC is an IATH project and is currently housed on an IATH-maintained server that is maintained by IATH and University of Virginia staff. SNAC is being transformed to a sustainable cooperative project hosted by the National Archives and Records Administration (NARA) in Washington, DC, within the next few years. The technical infrastructure will be hosted by the California Digital Library (CDL).

The CBW BESS database is populated by CBW's technical and editorial staff and is maintained as part of the project's data resources.

### ii. Expected data

CWBC will generate textual and visual data. Textual data includes XML files, EAC-CPF records, metadata, analysis, and the white paper. Visual data include graphs and timelines. We will also generate URIs as unique identifiers and have descriptive and administrative metadata associated with both SNAC and CWBC records.

Textual data associated with the Advisory Board and classroom users (e.g., notes and e-mails messages) will and will not be preserved or made accessible beyond the lifespan of the grant period.

### iii. Period of data retention

The data that CWBC is drawing from is not restricted or under copyright. Data generated by CWBC will be made available to the public as soon as it is intellectually and technically stable.

### iv. Data formats and dissemination

*Data Formats:* Data will be distributed in widely used open standards. The SNAC match/merge process will generate EAC-CPF records, an XML-Schema, and administrative and descriptive metadata stored in PostgreSQL (an open source database). Other data related to the cohorts will be also be stored in PostgreSQL.

We will use open source standards (e.g., HTML, CSS, and XML) for all textual data. Graphs and other visualizations will also be published in open source standards (e.g., GraphML).

*Data Access:* CBWC data analysis and code will be also freely available to the public, via either the CBW blog or the CWBC web interface, under a CC BY license.<sup>1</sup> While planning for the emerging SNAC cooperative is still underway, the policy with respect to SNAC data will insure that it is freely available to the public under CC BY.

*Dissemination:* CBWC data will be made publicly available via a web-based interface from the CBW web site when it has reached a point of technical and intellectual stability. SNAC data is available via the SNAC Prototype History Research Tool. The SNAC cooperative will continue to provide free access via a web interface.

### v. Data storage and preservation of access

CWBC's data resources will be stored on one of IATH's four servers, which are located at the University of Virginia's Information Technology Services (ITS) data center. SNAC data is on a dedicated server

---

<sup>1</sup> <https://creativecommons.org/licenses/>

maintained by IATH, also in the ITS data center. ITS provides 24x7x365 maintenance of the servers as well as backup and security services. The IATH system administrator also maintains redundant copies of IATH server data on a set of external hard drives.

CWBC data will be stored and accessible on an IATH server indefinitely, beyond the grant period. IATH's backup plan provides for data redundancy and off-site storage.

SNAC will remain an IATH project and IATH, NARA, and CDL are committed to maintaining the data and provide free public access to it for the foreseeable future.