



The Green-Wood Historic Fund
2019 Implementation Proposal to the National Endowment for the Humanities
Humanities Collections and Reference Resources
Providing Access to the Unexpectedly Rich Records of Brooklyn's Green-Wood Cemetery
Compiled by Julie I. May
Contributors: Lisa Alpert, Stacy Locke, Bill McKinney, Colin McDonald

TABLE OF CONTENTS

Introduction	3
Step 1: Transcribing 60 Volumes of Burial & Vital Records	4
Step 2: Building the Database	6
Step 3: Website Integration	8
Next Steps	9
Lessons Learned & Conclusion	11
Appendices	12

Introduction

The Green-Wood Cemetery, one of America's oldest rural cemeteries, was established in 1838 on 478 acres in what is now known as the Sunset Park neighborhood of Brooklyn in the Greater City of New York. It is the resting place for 570,000 permanent residents, including Leonard Bernstein, Boss Tweed, Charles Ebbets, Jean-Michel Basquiat, Louis Comfort Tiffany, Horace Greeley, Civil War generals, baseball legends, politicians, artists, entertainers, and inventors. In addition, it is a Revolutionary War historic site (the Battle of Long Island was fought in 1776 across what is now its grounds), a designated site on the Civil War Discovery Trail, and a registered member of the Audubon Cooperative Sanctuary System. Green-Wood was designated a National Historic Landmark by the United States Department of the Interior, which recognized its national significance in art, architecture, landscaping, and history, in 2006. In 1999, it established The Green-Wood Historic Fund, a 501c3 organization whose mission is to preserve and maintain the cemetery, including its monuments of historical and cultural significance; to advance the public knowledge of Green-Wood; to conduct, sponsor, and host education programs in the community; and to safeguard the natural habitat and parklands of one of New York City's first green spaces.

In recent years, the Fund came to appreciate the importance of its vast collection of institutional records, which contains the documentation for every interment since the cemetery's first burial in 1840. These records were the focus of Green-Wood's 2015 National Endowment for the Humanities Foundations grant, "*Gone but Not Forgotten: Digitizing the 177 Year-Old Legacy of New York City's Green-Wood Cemetery.*" With engagement from an advisory board, the Fund sought to identify high priorities to focus its energies upon in order to advance its mission. Among all of the archival documents at Green-Wood — 3,000 cubic feet in total — the advisory board zeroed in on 60 handwritten, cloth-bound volumes, which recorded all interments at Green-Wood in its first century, known as the Burial & Vital Records. The scholars were particularly excited about the unique combination of demographic data they contained. Each record includes: name; age at death in years, months, and days; cause of death; place of birth; street address; marital status; lot number (location in the cemetery); and name of the undertaker. In total, there are 438,180 individual records in these ledgers. Significantly, the data pertains primarily to individuals who lived in either New York City or Brooklyn — two cities whose populations experienced explosive growth in the nineteenth century — during which New York City established itself as the largest, most prosperous and most industrious city in the nation. While parts of the data available in these unique records are discoverable by other means, this is a comprehensive set of data which has the potential to be very consequential to the work of researchers of history, anthropology, epidemiology and public health, immigration statistics, and more. The *Gone But Not Forgotten* project, completed in 2016 was the first step toward making this rich dataset available by digitizing every page in all 60 volumes. The ledgers were shipped to the digitization vendor, dismantled, scanned at high resolution, and swiftly returned to Green-Wood where they continue to be utilized by staff and researchers.

The second part of making this data accessible to the wider public is the 2019 implementation grant funded by the National Endowment for the Humanities, *Providing Access to the Unexpectedly Rich Records of Brooklyn's Green-Wood Cemetery*. The project team, made up of Lisa Alpert, Grant Administrator; Stacy Locke, Project Technician; Julie May, Project Director, Bill McKinney, Developer; Colin McDonald, Webmaster; and Tommy Alpert, Technical Consultant worked together to transcribe the ledgers, build the database and search engine, upload the records to the database, and build a web page to make the information accessible. The following pages, define, describe and disseminate the project's details so that other cemeteries may replicate the program, others may harvest the data, and the public may search for the many residents in Green-Wood Cemetery. By making its records publicly accessible through this online database, Green-Wood provides access to almost 100 years of data never before available outside Green-Wood; welcomes new historical analysis and scholarship in many fields of study, including urban studies, public health history, population studies, social research, urban planning, and more; and paves the way for an unlimited amount of additional records from other cemeteries to be made accessible to researchers who employ any combination of the

software, data interpretation workflows, and programming language exemplified by this project and also made freely available.

The Burial & Vital Records is comprised of 60 cloth-bound ledger books, which record all interments at Green-Wood from the first burial in 1840 through 1937¹. The Burial & Vital Records were maintained by cemetery record keepers on a daily basis. Each interment had an entry for name of the deceased; burial date; age at death in years, months, and days; place of birth; marital status; street and house number of residence; cause of death; name of the undertaker; lot number; occasional ancillary notations; and small sketches or diagrams of the lot itself. Their format is similar to the United States census records; names are listed down the left hand margin of the page with details in vertical columns that span two pages. In communicating with other historic cemeteries around the country, such as Bellefontaine Cemetery in St. Louis, Forest Lawn Cemetery in Buffalo, Hollywood Cemetery in Richmond, and Graceland Cemetery in Chicago, Green-Wood learned that their burial ledgers contain essentially the same information fields as Green-Wood's. This means that the path to make Green-Wood's records accessible and searchable through this project has the ability to also be utilized by hundreds if not thousands of historic cemeteries in the United States.

Transcribing 60 Volumes of Burial & Vital Records

As mentioned previously, the burial registries were already scanned by Innovative Document Imaging (IDI) in East Brunswick, NJ. The scan quality of these images allowed Green-Wood to utilize them in a variety of ways: the high resolution files that are stored on the server can be investigated closely by zooming in; the screen resolution files are small enough to load into a website but high quality enough for a detailed view; and the PDFs may stand in for the book experience in page-by-page browsing. These files were also used to send to the remote transcription team based in India.² The transcription team employed the double keying method, by which two data typists enter the same data into different files to generate two outputs. The two data outputs were compared using an IDI proprietary quality control software that highlighted any quality issues between the two keying outputs. This data capture process, along with the software, promised a guaranteed accuracy of 99.5%. In addition, the transcription plan was to capture and reproduce any and all misspellings and/or other errors, so that the database is a true primary source that accurately reflects each word as written. IDI provided the transcribed records to Green-Wood in the latest version of Microsoft Excel that facilitated its incorporation into the database, but also allowed the project team to review the data and make global updates to the records when necessary. The transcription part of the project can be described in the following three stages, having different and increasingly better outcomes.

Stage 1

The first method employed was to take advantage of the transcribers who were looking at every line and every field in the registry. Then, the QA/QC team, made up of volunteers, would spot-check and fix glaring errors. Therefore, Green-Wood's first set of instructions for the transcription team included interpretation of the data rather than an as-written, word-for-word transcription. For example, the first set of instructions were to transcribe the single field, "place of death," into two fields: one for City and another for State. An interment record with the entry "Flatbush, Long Island" was interpreted by the transcribers as City: "New

¹ In 1937, Green-Wood stopped recording interments in the Burial & Vital Records and moved to burial cards instead (index cards that captured similar information.) However, the cause of death was no longer recorded by the cemetery. It is unclear exactly why the change happened in that year. Fear of epidemics had subsided, eliminating the need for daily interment lists, and the cause of death had come to be viewed as a private matter. Some speculate that because the cemetery had reached the century mark, it was time to modernize and move away from 19th century-style ledgers.

² Green-Wood considered options for transcription, including crowd-sourcing. Ultimately, the advisory board advised against crowdsourcing due to the burden on Green-Wood staff, including at least one individual to actively manage and oversee the project to ensure quality control. The advisors strongly advocated for working with an established company, such as IDI, to realize the most accurate and efficient results.

York”; State: “New York”. Those who are familiar with the geography and history of New York City know that Flatbush is a neighborhood in the borough of Brooklyn in what is now the consolidated Greater City of New York and New York, New York refers to the borough of Manhattan, also in the Greater City of New York. In other words, these are not the same geographic areas and the goal of this project is to provide more granular, local information about these interments. In another example, when transcribers encountered a “+” next to someone’s name, they were instructed to add “lot owner” to the “Notes/Remarks” column. Another example regarding “burial location” instructed the following: “Note: R.T. is "Receiving Tomb; S.V. is "Strangers Vault", if there is a range use "170-176" for ex.; Use the most current Lot and grave location listed (that which has NOT been crossed out). Use tags for "Stranger's Vault" and "Single Grave" and "Receiving Tomb." This series of if/then variables then adding tags to facilitate website search functions posed some challenges for the transcribers. Finally, the columns on the transcription template spreadsheet didn’t always match the columns on the ledger pages so that across multiple volumes, each transcriber made different decisions on how they handled information that fell outside of the provided template. In short, it was unreasonable to request a team unfamiliar with the ledgers' information to make these types of interpretive decisions. It yielded an increase in the opportunity to make errors with each data point that may have strayed from consistency (of which there are many instances); and checking the accuracy of this data input would require the same review time as re-transcribing the ledger data. Volumes one to 15 were transcribed and delivered before the errors were discovered and course-corrections could be made and were without any consistency that could be immediately rectified. Therefore, the project team temporarily left Volumes one to 15 as transcribed in order to create a new set of instructions that would improve the outcome.

Stage 2

The second stage included a revised set of instructions and a formal workflow: 1) an enhanced excel template that represented the exact column headers a transcriber would encounter. For instance, there was one version of the template/column headers for volumes 16-20, another for volumes 21-28, and a third for volumes 29-60; and 2) the instructions for each field were embedded in the spreadsheet alongside a screen-shot of the original page with a specific symbol like a “+” or ditto marks and then the examples of how that field should be transcribed using the aforementioned representations and instruction³; 3) no interpretation was included in the instructions; rather, instructions for a transcribe what you see method; and 4) the team created a detailed workflow and environment within which to track when the transcription was delivered, QA/QC’d, uploaded to github for processing into the database; and approved for inclusion in the online dataset using a Google Sheet embedded in the Microsoft Teams group exclusively for this project team. This was utilized to immediately address any repetitive mistakes or course-corrections before getting too far along with the transcription. To complete the QA/QC review, the project team engaged a group of Green-Wood volunteers, held training sessions, created a Microsoft Teams environment containing the tracking sheet and instructions to facilitate a comprehensive and remote but collaborative working environment.⁴ The transcription IDI produced through this method was drastically improved, so much so that they were able to finish out the remaining volumes using this method and only required a few additional variations on the excel templates for guidance. While the transcription improved, the volunteers dropped out one by one because the QA/QC instructions required eight hours or more to complete and despite the elaborate instructions, they felt overwhelmed with questions and lack of confidence when identifying errors and making changes to the spreadsheet data knowing it would make up the data in the future website.

Stage 3

In the third and final stage of transcription, QA/QC duties were transferred exclusively to Stacy Locke, a tenured Green-Wood staff member familiar with the burial records and their utilization by staff and external

³ See Appendix A | Transcription Directives Templates

⁴ See Appendix A | QA/QC Environment

researchers. The QA/QC steps were also reduced⁵ in order to reach the grant's project goals and deadline, but also employed developer tools like name parsing, date parsing, and Python to convert symbols or abbreviations into data. For example, ditto marks and "Do" in the transcription were converted into the value of the previous entry.⁶ This reduced the review time from more than eight hours to four hours or less. Nevertheless, QA/QC revealed many data variations and quandaries that needed further research in order to resolve. For example, in two volumes, several lines of margin notes were not transcribed into the remarks column. Since the remarks are additional notes and don't compromise discovery of that interment, the spreadsheets were uploaded to the database, but this information will be appended later on. In another volume the "Place of Death" column was copied from the "Late Residence" column instead of transcribed as a column with unique data. This volume was set aside and not uploaded until it can be transcribed correctly. Finally, another volume is missing the first seven pages of transcriptions because the ledger pages were torn, the data taped over, and illegible. In this case, the data was uploaded without the first seven pages and will also be appended at a later date. The notes for all volumes are recorded in a QA/QC Tracking Sheet.⁷ This third and final transcription and QA/QC method proved the most successful, though it was still time-consuming and required more time allocation than available.

In Summary

While the transcription part of this project is complete, the QA/QC and data normalization to facilitate search remains a significant challenge but long predicted as an ongoing process described in the grant proposal and reflected in the setup of the database. Green-Wood expects significant input and commentary when the data is made public and then searched, investigated, questioned, and cross-checked. Green-Wood staffers will be able to make updates, either to correct or append data, as mentioned in the above scenario. In retrospect, time could have been saved and transcription issues highlighted with a different staffing model from the beginning. Institutional knowledge and subject familiarity cannot be underestimated or valued here as evidenced by the improvement in the QA/QC process and quality of the data once Stacy Locke adopted the task. As a long-time staffer, she had extensive familiarity with these records and could identify data variations and uniqueness. Despite her skill, the project would have benefitted from a larger, compensated, trained, and dedicated team to perform the work with the benefit of her supervision rather than volunteers and placing responsibility for the accurate of 60 volumes' worth of data integrity upon Ms. Locke.

Building the Database/Building the Search Index

The search experience is made possible by converting the registry data from the standard spreadsheets provided by the transcribers and QA/QC'd by the project team to a format called JSON (JavaScript Object Notation)⁸ which the database and search engine can interpret. In this case, the database selected was Elasticsearch hosted by ReactiveSearch (formerly Appbase.io). Elasticsearch is "a distributed, free and open search and analytics engine for all types of data, including textual, numerical, geospatial, structured, and unstructured."⁹ Elasticsearch ingests this JSON data and creates an inverted index that enables very fast searches for users and researchers. In addition to simply ingesting it, a schema is necessary to describe all of the fields in an interment record and how to handle those fields. The schema required by the Elasticsearch search index consists of a mapping file that describes all of the fields in each JSON entry along with their data type. It also describes how each field should be indexed under the hood based on its data type. In this case,

⁵ See Appendix A | Revised Instructions for QA/QC Duties

⁶ The transformation logic and rules were coded in a Python program available here: https://github.com/Green-Wood-Cemetery/burial-registry-search/blob/master/data/process_spreadsheet.py and the output for this example is here:

<https://github.com/Green-Wood-Cemetery/burial-registry-search/tree/master/data/excel/output>

⁷ See Appendix A | Review Log Sheet

⁸ See <https://www.json.org/json-en.html>

⁹ See <https://www.elastic.co/what-is/elasticsearch>

fields or properties in the search index simply map to each column in the spreadsheets.¹⁰ It's important to note that the fields were broken down to a more granular level in order to make the search interface possible with the components the project team believed should be available. For example, in order to provide a year range slider as in the Date of Interment and Age at death fields: the year field that contained values like "August 8, 1873" had to have a new year assigned with the value type "long" so that it would be treated as a number that could be sorted and ordered.

Hosting

Since Elasticsearch is an open source search engine, it can be hosted internally or with a third-party company that provides external hosting. Green-Wood evaluated whether their existing web hosting provider had the expertise and willingness to host Elasticsearch internally and, after discussion, decided to go with a third-party provider, appbase.io. This company provides a web dashboard that makes it easy to update and add new data going forward. The dashboard allows non-technical staff to interact with the data with minimal training. For example, new registry data can be uploaded via a simple drag and drop interface. Existing data can also be corrected or modified by simply searching for the interment ID and editing the fields in place.

Full-text Search Capabilities

A search engine like Elasticsearch uses an inverted index¹¹ and has a number of advantages over searches that rely only on relational databases. In particular, very powerful and useful search operations become available. Examples of some of those features are:

Text Searches

When a search term is entered into the main search box it searches every field in the database that is mapped to the transcribed spreadsheet but includes both the transcribed data and the displayed data.¹²

Weighted Searches

The first five name fields¹³ have been assigned a larger field weight than the other fields. We assigned these fields a weight value of 3 out of a range of 0 to 10, 10 being the most substantial weight. This means that they will appear higher in the results list. For example, a search of "ireland" will return results first where that term appears as the last name compared to where it appears as part of the birth place field.

Boolean Searches

By default, search terms are joined by the boolean OR operator. This means a search of: 'Ireland England' will find all interments that have either of those terms. You can change this default by adding AND between the terms to limit results to interments that include both. You can also combine boolean operators using parentheses to indicate precedence:

```
(ireland OR england) NOT manchester  
(ny OR ("new york") OR ("n y"))  
cause_of_death_display:suicide AND gender_guess:M
```

Phrase Searches

Use double-quotes to enclose phrase searches: "287 Lexington Ave"

Wildcard Searches

Searching a name with an asterisk (john*) will return results containing Johnson, Johnston, Johnsten, etc.

¹⁰ See Appendix B | Figure 1

¹¹ See https://en.wikipedia.org/wiki/Inverted_index

¹² See Appendix B | Figure 2

¹³ name_last, name_first, name_transcribed, name_display, and name_middle

Fuzzy Searches

Searching a term in combination with a tilde (Hemorrhage~) will yield results containing “Hemorrhage” and “Hemorrhagia”

You can specify a number after the tilde symbol to indicate the “edit distance” or the number of one character changes that need to be made to one string to make it the same as another string. The default value of 3 is normally sufficient.

Fielded Searches

You can narrow your search to a specific field using this syntax:

```
burial_location_lot_current:194
interment_date_iso:"1897-02-23"
interment_date_display:"July 4, 1880"
cause_of_death_display:heart
```

You can perform fielded searches on many fields, whether they are displayed in the results list or not:

```
is_lot_owner:true
has_diagram:true
gender_guess:F
registry_page:024
registry_volume:39
interment_id: 282624
age_years:56
```

Range Searches

```
age_display:35-40
(interment_date_year_transcribed:>=1885) AND (interment_date_year_transcribed:<=1886)
```

Filtered Searches

The final method is a visual search filter component displayed in the user interface rather than typing them directly in the search box. This provides suggestions on search terms and also refines the search as filter items are selected.¹⁴

Website Integration

Building the site is accomplished in multiple parts: the data transformation that is uploaded to Github; the communication between the web page and the search engine containing the database; and the aesthetic integration of the search engine into the existing user experience. As mentioned previously, the project team reviewed the spreadsheets provided by the transcribers and, once approved, uploaded those in the latest version of Excel to the GitHub repository where they are considered the definitive or “single source of truth” that populate the database and where they are permanently stored and versioned in the data/excel/reviewed directory.¹⁵ Any subsequent additions, corrections, and their version histories can be found here.¹⁶ Once uploaded to the Github environment, they were then transformed into JSON by writing a Python programming language¹⁷ to read the Excel spreadsheets and convert them into the format required by

¹⁴ See Appendix B | Figure 3

¹⁵ See <https://github.com/Green-Wood-Cemetery/burial-registry-search/tree/master/data/excel/reviewed>

¹⁶ Spreadsheets are a particularly good format for data transformation and reuse. There are many publicly available tools to export them to other formats. Specifically, comma or tab separated value exports make it possible for developers and other technically minded researchers to easily convert and reuse the data.

¹⁷ Python modules like `openpyxl` and `pandas` made it a good language choice for working with spreadsheet or tabular data. Green-Wood's transformation code can be found at <https://github.com/Green-Wood-Cemetery/burial-registry-search/tree/master/data>

Elasticsearch. Any time data needs to be updated or corrected, it should be done in the relevant spreadsheet and that spreadsheet uploaded to GitHub. Following that, it should be re-converted to JSON and re-imported into Elasticsearch. Remember that a benefit to using Elasticsearch is that data may be updated directly within its interface and will yield the same result as this multi-step process, an updated database. While either method is acceptable, best practice mandates that the spreadsheets and the database match, especially when deploying the data to other venues.

Once the data is in the database, it must speak to a website utilizing a prepackaged build directory containing HTML, CSS and Javascript files, available through Green-Wood's github page.¹⁸ The directory enabled our website designer (or anyone else) to embed the search application in a website. In this case, Green-Wood created a new page, Burial and Vital Records: 1840-1937¹⁹ to embed the search. These files do not require external dependencies or any particular web framework or programming language to be in place. Green-Wood's search application is deployed to a WordPress website, written using PHP as the scripting language. In addition to Elasticsearch providing the data, the hosting service, appbase.io, provided open source search UI components used to construct Green-Wood's search page called ReactiveSearch²⁰. A variety of user interface components may be implemented into a website from an open source library of components that communicates directly with the Elasticsearch index including a date slider bar, the birthplace drop down search ordered alphabetically, and an open search field.²¹ This component is easy to configure using existing documentation and doesn't require particularly specialized programming knowledge or expertise. In addition to search UI, Green-Wood also employed Ant Design, another open source UI software library that enabled us to arrange the Elasticsearch components like inserting a pagination button at the bottom of the search application, provides callout tooltips, asserts the row and column layout, and programs collapsible sections like the "show more" arrow on the search results.²² Finally, a fully mobile-compatible version of the search application will be completed in the next iteration of the site due to the time and resource constraints of the project. The software tools and libraries currently in use provide a means to reach the goal of mobile compatibility by leveraging the core user interface components provided by Ant Design and Ant Design Mobile. In addition, appbase.io provides open source user interface packages that can produce a native version of the search application for Android and iOS phones.

Next Steps

The most important step to complete is to finish QA/QC'ing the remaining spreadsheets from the transcribers. This will be completed by Green-Wood staff and flowed into the processes outlined previously. Bill McKinney, who implemented all of the above programming languages and scripts to make the search engine work, will provide the necessary training to append the rest of the Burial & Vital Records data, but also to allow for edits to the existing data set. In addition, details of this project with its challenges, lessons learned, and outcomes are in service to other cemeteries who want to make similar data available to researchers. As such, the data, code, bug reports, enhancement requests, search interface, and documentation is open and publicly available to anyone through the institutional Green-Wood GitHub account where the repository for the project lives.²³ There, the project team attempted to track issues so that every team member could create, assign, search, locate and comment on issues rather than tracking bugs, questions, and enhancement requests in numerous email threads where they can be easily lost and forgotten and difficult to share with other interested parties. Further, team members could track issues and see if they have been

¹⁸ See <https://github.com/Green-Wood-Cemetery/burial-registry-search/tree/master/build>

¹⁹ See <https://www.green-wood.com/burial-and-vital-records/>

²⁰ See <https://opensource.appbase.io/reactivesearch/>

²¹ Appendix C | Figure 4

²² <https://github.com/ant-design/ant-design/>

²³ <https://github.com/Green-Wood-Cemetery/burial-registry-search>

addressed or not in the latest code. In addition to a basic thread of errors or issues, the communications are also tagged by categories: bug, data error, dependencies, documentation, duplicate, enhancement, good first issue, help wanted, invalid, question, task, and wontfix.²⁴ This information might also be useful to review prior to embarking on a similar project as it could flag variations in data before they are encountered as an error in the database. Contributing the data to search aggregators like Ancestry.org is also a future goal. When all spreadsheets are reviewed by Green-Wood staff for accuracy, and uploaded to the search engine, the spreadsheets will also be deposited to the Dataverse Project for preservation and dissemination.²⁵ This will provide an independent and non-commercial location to store, preserve, version, and share the data with researchers around the world. GitHub will also store a copy of this data, but since it is owned by a commercial entity (Microsoft Corporation) whose main purpose is to aid ongoing development of code and data, it is a supplementary data repository to the Dataverse Project due to its primary purpose as a long-term preservation venue and a source for academic data citation. Finally, future aspirations include future iterations of the data and search engine including:

Gender Guessing

The archival registry pages record quite a bit of information but unfortunately they do not include gender and this was an extra project goal to infer by first name. It was made possible by purchasing a name-gender database, saved locally in a single SQLite database, in order to use a Python module to query as part of the spreadsheet processing. The new field named “gender_guess” contains one of the following possible values: M; F; Unisex; and Unknown. Because this data is only a guess and is in no way authoritative, it is not visible in the user interface by default. However, researchers who are willing to accept this provisional data can perform advanced search queries using the general search field and entering: “gender_guess: [value]”, e.g. “gender_guess:M”.

Geocoding Places

Early in the project the project team experimented with converting places (late residence, place of death, birth place) into standard geographic coordinates (latitude and longitude) and individual location fields (street name, city, state, country, etc). As an example, if you use the MapBox Geocoding API²⁶ to search for a late residence address of “204 Prince St Brooklyn” as entered in the registry page, you will get complete geographic details in the response. This data would help enormously to standardize all place names and avoid numerous typographical variations of the same place names in the search interface, enable researchers to link to maps and perform queries using geographic proximity, and generate maps of disease spread in Brooklyn like Cholera or the Spanish Flu. Unfortunately, in the process of exploring and testing this data enhancement several challenges rendered this a feature for the next iteration of the search engine and interface.

Place Names Change Over Time

Street names, numbers, building names can all change over time. There is no guarantee that “204 Prince St Brooklyn” is at the exact same place today as it was in 1897.²⁷ This would require extensive and time consuming review by people with specialized knowledge of historic Brooklyn addresses and New York

²⁴ See Appendix D | Figure 5

²⁵ “The Dataverse Project is an open source web application to share, preserve, cite, explore, and analyze research data. It facilitates making data available to others, and allows you to replicate others' work more easily. Researchers, journals, data authors, publishers, data distributors, and affiliated institutions all receive academic credit and web visibility.”

²⁶ See <https://docs.mapbox.com/api/search/geocoding/>

²⁷ See 150 Years Ago Brooklyn Renumbered All Its Streets. It Was a Disaster at <https://www.nytimes.com/interactive/2021/01/27/nyregion/brooklyn-streets-numbers-renaming.html>

geography or by cross-referencing against a resource like Center for Brooklyn History's Map Collections portal.²⁸

Synonyms

Elasticsearch has a feature that provides for the ability to manage synonyms. This might be very useful to handle common archaic words or archaic spellings (e.g.: "diarrhoea" and "diarrhea"). It could also come in handy to map historic medical terms to their modern equivalents (e.g.: "scarlatina" and "scarlet fever"). In both cases, the user would benefit by being able to search using modern terms and still get relevant results from the search index. The current Elasticsearch hosting provider includes the synonym feature in their dashboard with a premium plan so it could be implemented in the future if time and resources allow.

Connection to legacy Green-Wood databases

A future project might be to connect or link the legacy burial search database to this new full-text search when identical records exist in both places. It might also be another method to perform additional quality assurance on the transcription accuracy.

Cemetery Mapping

A future project might include adding grave geolocation data to the search index so that visitors could more easily locate graves using the search application, similar to Mount Auburn Cemetery in Cambridge Massachusetts who offers this feature as part of their search application.

Lessons Learned & Conclusion

The Green-Wood Cemetery, one of America's oldest rural cemeteries: 478 acres, 570,000 permanent residents, and nearly 100 years of interment records handwritten in 60 ledgers known as the Burial & Vital Records. The Green-Wood Historic Fund's goal was to make these records available to the public to serve its mission to advance the public knowledge of Green-Wood, and in doing so added a unique and comprehensive data set to benefit research in urban studies, public health history, population studies, social research, urban planning, and family genealogy while also smoothing the way for other cemeteries seeking to accomplish similar accessibility to their records. The first project accomplished by Green-Wood digitized every single page in each ledger which then facilitated a second grant to transcribe each interment record into a searchable online database by name; age at death in years, months, and days; cause of death; place of birth; street address; marital status; lot number (location in the cemetery); and name of the undertaker. The team encountered several challenges and learned many lessons that are summarized above and available in greater detail through Green-Wood's GitHub account. Institutional knowledge and familiarity with the subject data is an essential element to ensure the data is transcribed well; the QA/QC task of a data project like this is important and requires many hours of person power due to its complexity and variations across 100 years of human, hand-written data entry. This project and others like it would benefit from resourcing with an experienced or extensively trained team. Also, working with a Developer who can design a workflow that can be executed and maintained by the existing cemetery staff was a request of this project, but is highly recommended especially since Green-Wood remains an active cemetery and the dataset in this case will be further enhanced and corrected. This is an incredible accomplishment for which Green-Wood is immensely proud and excited for the research that is now possible by making these records available. Future iterations of the site are expected in the months and years to come to fully realize the strength of the database within which the data is served, but also to respond to the research areas that Green-Wood has yet to encounter by making it publicly accessible.

²⁸ See <https://mapcollections.brooklynhistory.org/>

Appendix A | Transcription Directives Templates

Stage 1 Directives

Directives - Version 3 ☆ 📄 ☁

File Edit View Insert Format Data Tools Extensions Help [Last edit was on July 17](#)

100% | \$ % .0 .00 123 | Arial | 10 | **B** *I* ~~ABC~~ A | 📏 🗺 📄

	A	B	C	D	E	F
1	Image No.	Interment Number	Date of Interment			Name of Deceased
2	Image No.	Interment Number	Month	Day	Year	
3	This corresponds to the scanned image of the Chron book page on which this record appears		Note: Spell out full month	Note: Use only cardinal digit (not ordinal number)	Note: This is not a column, but we would like to include it as one	Entire name as it appears in this column, including saltuation, first name, middle name, last name, and infant or child of, if listed
4						
5						
6	General Notes: Repeat the same information in fields where "ditto marks" or "same" OR "DO" is written (except in case of Age--see notes); Horizontal or squiggle lines indicate no value and should be treated as blank fields. Leave blank fields blank; Aim to transcribe exactly as originally written.					

Stage 1 Directives (continued)

G		H		I	J	K
Burial Location						Place of Birth (known in earlier vols. as "Nativity")
Current Lot (last recorded lot)	Current Grave (Last recorded grave)		Original/Previous Lot	Original/Previous Grave		
Use the most current Lot and grave location listed (that which has NOT been crossed out).	Note: Grave is often indicated as "No." after the lot number in early volumes without a grave field (ex. 769 no. 19); There is not always a grave number. (This field becomes more complicated in later volumes, transcribe exactly what is written)		Use if the original lot and/or grave have been crossed out. The crossed out number goes here. ; If there are multiple previous lots/graves, separate entries with a semicolon (ex. 1243; 6788)	Use if the original lot and/or grave have been crossed out. The crossed out number goes here.		List exactly as written
OR, list the crossed out item in the same field with a strikethrough, or brackets or something						

L	M	N	O	P	Q
Age			Marital Status	Late Residence	
Years	Months	Days	M/S/W/D	Late Residence	Street & Number
If a ditto mark ("") or line appears instead of a number, this should be translated as NO DATA. Do not copy the number from the value above.	Same as "R"	Same as "R" May also have hours if lived less than one day. In this case, include the word hours if in the column.	Note: Married, Single, Widowed, Divorced (Spell out full word)	Transcribe exactly as written	Note: Include Street number and name (when present) (ex. 17 Smith); This field will also include other places of death (like hospitals, asylums, ships, etc)

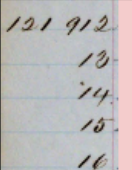
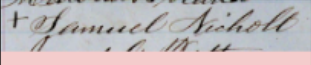
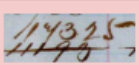
Stage 1 Directives (continued)

R	S	T	U	V
Place of Death (Doesn't appear in early volumes)	Date of Death (Does not appear in early volumes)			Cause of Death (known in some vols. as "Disease"); Also includes instances of removal TO Green-Wood
	Month	Day	Year	Disease/Cause of Death
Transcribe exactly as written				Note: Do not correct misspellings that were made by the original writer. Diseases should be transcribed exactly as they are written. Remonals from other cemeteries can be left in this column if they appear there.

W	X
Undertaker	Notes (Also called "Remarks")
Undertaker	Notes
Note: Full name in single field; No periods after Initials, spaces between initials (i.e. D M Bennett).	Transcribe content exactly as written. (**Notes made going across several other fields [usually removals] can be moved here.) Use semi-colons to separate multiple statements. If there is a diagram, use text: "Diagram available: See digital image".

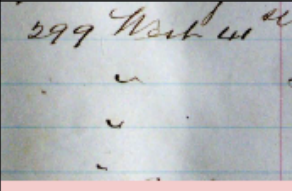
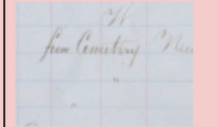
Appendix A | Transcription Directives Templates

Stage 2 & 3 Directives

	A	B	C	D	E	F
1	Image No.	NUMBER.	DATE		NAME	NO. OF LOT
2						
3	Insert the filename of the image you are transcribing	1) Populate with 6-digit number; example = 121912; 2) 2) When the field includes a 2-digit number, populate including the previous field's first 4 digits + 2-digit number in field; example = row 10 contains "15" and previous 6-digit number is "121912", then populate row 10 with "121915"	1) Populate with the month; 2) When field is blank, leave blank.	1) Populate with the day; 2) When field is blank, leave blank.	1) Enter the name exactly as it appears; no changes; 2) When there is a "+" preceding the name, enter "+"	1) Enter the number exactly as it appears; 2) When the field also contains a strike-through number, place that within brackets []
4						
5						
6	Volume 01_004	121912			+ Samuel Nicholl	114325 [11193]
7		121913				
8		121914				
9		121915				
10		121916				
11						

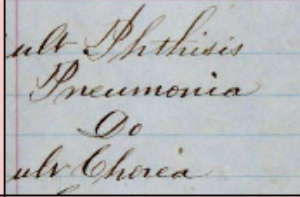
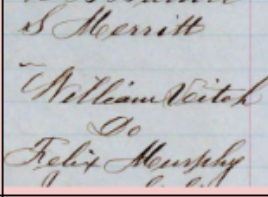
Providing Access to the Unexpectedly Rich Records of Brooklyn's Green-Wood Cemetery

Stage 2 & 3 Directives (continued)


L	M	N	O	P
MARRIED	SINGLE	LATE RESIDENCE	STREET & NUMBER	PLACE OF DEATH
1) Enter exactly as it appears; 2) When field is blank, leave blank; 3) When field contains a hashmark, populate with a hashmark (""); 4) When it contains number, enter exactly as it appears.	1) Enter exactly as it appears; 2) When field is blank, leave blank; 3) When field contains a hashmark, populate with "	1) Enter exactly as it appears; 2) When field is blank, leave blank; 3) When field contains a hashmark, populate with "	1) Enter exactly as it appears; no changes. 2) When field contains a hashmark ("), populate with a hashmark (""); 3) When field is blank, leave blank; 4) When field contains a "~" (horizontal or wavy line), enter a dash "-". Example: 175 Smith	1) Enter exactly as it appears; no changes. Example: "Brooklyn" is entered as "Brooklyn" 2) When field contains a hashmark ("), populate with a hashmark "
				
				
M		New York	299 West 41st	New York
"		"	-	"
		Brooklyn	-	Brooklyn
M		"	-	"
3 1/2 hours		"		
W				
from Cemetery				
"				

Providing Access to the Unexpectedly Rich Records of Brooklyn's Green-Wood Cemetery

Stage 2 & 3 Directives (continued)

Q	R	S	T	U
TIME OF DEATH		DISEASE	UNDERTAKER	REMARKS
1) Enter exactly as it appears; 2) When field is blank, leave blank.	1) Enter exactly as it appears; 2) When field is blank, leave blank.	1) Enter exactly as it appears; 2) When field contains a "~" (horizontal or wavy line), enter a dash "-"; 3) When field contains "DO", enter exactly as it appears; 4) when it contains "ult", enter exactly as it appears.	1) Enter exactly as it appears; 2) When field contains a "~" (horizontal or wavy line), enter a dash "-"; 3) When field is blank, leave blank. 4) When field contains "DO", enter exactly as it appears	1) Enter exactly as it appears; 2) Insert notes that appear outside of rows, columns, and in margins
				
		ult Phthisis	S Merritt	
		Pneumonia	-	
		Do	William Veitch	
		ult Chorea	Do	
			Felix Murphy	

Stage 2 & 3 Directives (continued)

V	W
Diagram	Needs Review
Where a diagram appears, please enter "diagram" in the first row in which that diagram is drawn.	Please "review" where the row contains characters that are unusual, confusing, or difficult to transcribe so that Green-Wood staff can take an additional look at it.
	
	<i>review</i>
<i>diagram</i>	

Providing Access to the Unexpectedly Rich Records of Brooklyn's Green-Wood Cemetery

Appendix A | QA/QC Environment
Tracking Sheet

	A	B	C	D	E	F	G	H	I	J
1	Volume #	Transcribed?	Approved for Processing + uploaded to Bill	Processed by Bill + saved/assigned in Google drive	Assigned to Volunteer	Approved for Website?	Uploaded to Github for website incorporation	Comments/Feedback	Scanning assessment	Requires special attention from G-W team
17	16	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Erin	<input type="checkbox"/>	<input type="checkbox"/>	In progress		<input type="checkbox"/>
18	17	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Aubrey	<input type="checkbox"/>	<input type="checkbox"/>	In progress		<input type="checkbox"/>
19			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	6/23/21: (1) Pages 114 through 121 appear to be exact duplicates (in scan and transcription) of pages 105-112. (Int #143351-143806). Thus we are missing pages covering Int #143807 to 144319 (8 pages); (2) Ends with half page, not matched. Does not appear that the corresponding page at the start of volume 19 was transcribed.		<input type="checkbox"/>
20	18	<input checked="" type="checkbox"/>			Stacy					
	19	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Stacy	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Re-reviewed by St. on 6.15.22 to correct Death place		<input type="checkbox"/>
21		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input type="checkbox"/>	8/30/2021: Stacy Removed half page from end, no longer overlaps with volume 21; Julie/Stacy need to merge with other volume so entire last/first page is present; 6/8/2022: First 7 pages removed because data incomplete (to be reviewed at a later date)		<input type="checkbox"/>
	20				Stacy					
22		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	8/30/2021: Uploaded revised file to Bill (Julie); 8/30/2021: Resolved by Stacy, see above; 8/29/2021: As per Bill - Volume 21 starts with 157712, which is included in Volume 20. Sent to Stacy to fix, may require reprocessing Volume 20 also (Julie); 8/11/2021: Stacy corrected this (Julie); Julie/Stacy need to merge with other volume so entire last/first page is present; corrected version uploaded to Bill 8/30/2021; 6/23/2021: Second half of first page (listed in transcription as p.2) is incorrect. (Not the one that appears at the start of volume 21, different data)		<input type="checkbox"/>
	21				Stacy					
23	22	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Stacy	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Returned for transcription correction; revised delivered		<input type="checkbox"/>
24	23	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	On hold for re-scanning: Insert right half of 2nd page of Volume 24 to last page of Volume 23 when it's ready.	Needs to be rescanned in full	<input type="checkbox"/>
								12/5/2021: Removed 2nd page lines -- they should be inserted to the end of Volume 23 when that is ready (Julie); 8/31/2021: Revised received (Julie); 8/30/2021: Revised file expected on 8/31/2021 (Julie); 8/29/2021: Reviewed file and page 2 has not been transcribed; returned to IDI (Julie); 8/12/2021: Received revised volume 24 (Julie); Volume 24 from Darlene ends with data transcribed from scanned image page 122, omitted pages 123-133 (with		<input type="checkbox"/>
25		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input type="checkbox"/>

Appendix A: Revised Instructions for QA/QC Duties

Unhighlighted text indicates the original instructions; highlighted text indicates the revised, reduced instructions

Instructions for Volunteer Reviewers of Burial & Vital Records Transcriptions

The goal of your volunteer work is to

- 1) Identify and correct incorrect personal names, street names, etc. in cases where they were mistranscribed.
- 2) Confirm the interpretation of the transcription data is successful

The workflow is as follows:

- Stacy and Julie completed a cursory review of all volumes and prepared Transcription Directives for IDI Transcribers
- IDI Transcribers completed transcription by inserting information into spreadsheet provided by Julie and Stacy
- IDI returned transcription; Julie and Stacy reviewed the document to make sure there were no errors so egregious or extensive that it needed to be sent back
- Saved original transcription to Google Drive: **_RAW IDI TRANSCRIPTION FILES**
- Julie/Stacy uploaded duplicate of raw file to Github for Bill
- Bill applied programming to duplicate file; returns it to Stacy and Julie; saved to Google Drive: **_POST-PROGRAMMATIC NORMALIZATION**
- Volunteers Reviewers quality check the transcription + programmatic normalization according to below instructions.

File Management:

- 1) Navigate to [POST-PROGRAMMATIC NORMALIZATION](#) folder and open the file you're assigned by locating the file that contains your initials
E.g. VOLUME_17 Final Data_processed_JM.xlsx
- 2) Select the Image Number hyperlink to view the original page/data in order to compare against transcription; download file to open in Adobe Reader if desired
- 3) When you are finished reviewing your Volume/spreadsheet, rename file with "completed" to the end
E.g. VOLUME_17 Final Data_processed_JM completed.xlsx

Quality Check Instructions:

When you discover that the [TRANSCRIBED] column is incorrect due to an error with the transcription, edit the [TRANSCRIBED] column to match the original volume

When you discover that the [DISPLAYED} column is incorrect due to an error with the transcription OR misinterpretation of the programming, edit the [DISPLAYED] column so that it is correct. (The displayed column is what database users will search and see in online burial records)

When the Geocoding is faulty, enter FALSE in the Faulty Geocode [ADMIN] column to flag it.

1. Column A | Interment ID:

- Confirm #s are populated with the same # of digits and in sequential order

- No chunks of rows missing

2. Column C-G | Interment Dates:

- Confirm date is populated accurately in the Interment Date [DISPLAYED] column

In this column, a blank row should be populated with the last field that was populated.

E.g. October followed by blank rows means every column following should be displayed as October until a row contains November; then displayed as November until a row contains December and so on.

- The Day column is treated the same as the month.

- Exceptions: When an interment date is out of sequential order

- Confirm that the Interment Date - ISO [SEARCH] field is generating the correct date in YYYY-MM-DD format; if not, edit this column so that it is correct.

- Scan for any errors (any months/days out of sequence? Any "s? Anything missing?)

3. Column H-J | Name:

- Scroll through entire column and look for mis-transcriptions and lack of capitalization

E.g Theodore F Vanbrunt → Theodore F Van Brunt

- Where a + precedes or follows the name, the Lot Owner in column O should be populated as TRUE

- When you discover that the [TRANSCRIBED] column is incorrect due to an error with the transcription, edit the [TRANSCRIBED] column to match the original volume; When you discover that the [DISPLAYED] column is incorrect due to an error with the transcription OR misinterpretation of the programming, edit the [DISPLAYED] column so that it is correct.

- Try to catch whatever errors you can along the way (Qui and Qua good to search, Mand for Maud, etc)

- Possible "a" issue in some volumes (ex. George A Sharp reads as Georgea Sharp), fix

- Correct Eliza/Elizabeth (not necessary after vol 41)

4. Column Q-S | Lot:

- A number in brackets indicates the Previous lot; the naked number indicates the current lot; confirm the Current and Previous columns are populated correctly

- Check for any "s carried over into display columns and populate correctly

- Locate any deviances, like periods and fix where you can (periods should be removed from lot numbers, typically used as commas in numbers. Not necessary in Grave numbers typically)

- Check for "From Cemetery," (or similar) remove from display column, make sure this text is populated in the Remarks column (if not, add it)

- Search for any brackets not properly carried over

5. Column T-V | Grave:

- A number in brackets indicates the Previous grave; the naked number indicates the current grave; confirm the Current and Previous columns are populated correctly in the relevant columns

- Same as above

- Any "See R column"? Check. If so, make sure the Diagram column is set to TRUE and remove from any incorrect records. Add to remarks (only done on earlier volumes).

6. Column W-Z | Birth Place

-Review for obvious misspellings by transcription team (Common Errors: The initials "SI" (Staten Island) and "LI" (Long Island) appear very similar in 19th-century script and may be mistranscribed)

- When you discover that the [TRANSCRIBED] column is incorrect due to an error with the

transcription, edit the [TRANSCRIBED] column to match the original volume; When you discover that the [DISPLAYED] column is incorrect due to an error with the transcription OR misinterpretation of the programming, edit the [DISPLAYED] column so that it is correct.

- Chat to us if you discover a value that is frequently picked up as incorrect in the Geocoded value and could be fixed programmatically (ex. "Conn" has been regularized to "Connecticut")

- Check that Geocoded [DISPLAYED] value is a logical match for [TRANSCRIBED]

(Common Errors: (1) In all cases where a numbered street address is populated in the Geocoded column, this is incorrect as the system is identifying a business on today's map with the same name. (You may see this with hospitals, but also with location names that no longer exist) (2) Geocoded [DISPLAYED] state is often misidentified so look out for that)

- In the case of an error, please add FALSE in the [Faulty Geocode [ADMIN] column

- When you see "removed from" leave it in that column

- Review for obvious misspellings using spell check and/or manually

- Use search/global updates to update abbreviations like "Pa" or "SI"

- Flag in yellow anything to be reviewed manually

- Search for any "s that were not resolved and fix manually

- (In earlier volumes) When a body has been moved from another cemetery it will say "From" or "From (place name)." Leave this here, but make sure the remarks column is populated with the original burial location. (Just copy From [wherever] over). Often the place or cemetery may be in the late residence/place of death column as well. It can stay in this display column, if logically populated by the original writer in the book. If not, if it seems they have strangely added the cemetery name, it may be deleted from displayed.

- Locate where New York State had erroneously been carried down because of the duplication intended just from New York and fix

- NORMALIZE N Y, N J, U S etc to NY, NJ, US

7. Column AO-AS: Age

- Review years, days, hours column and confirm the Age: Display [DISPLAYED] column is populated accurately.

- Where the age is hours, that appears in the Marital Status column. Make sure it is populated in the Age: Display [DISPLAYED] column

- Where the age is minutes, enter the number of minutes in the hours column including "minutes" E.g. 21 minutes

- When you discover that the [TRANSCRIBED] column is incorrect due to an error with the transcription, edit the [TRANSCRIBED] column to match the original volume; When you discover that the [DISPLAYED] column is incorrect due to an error with the transcription OR misinterpretation of the programming, edit the [DISPLAYED] column so that it is correct.

- Search the transcribed age columns for "hours, minutes, weeks, about, etc" and update columns appropriately. [Leave it where it is, but update misspellings; duplicate in display column where it says unknown; add to the hours column as a single digit if hours. If weeks or minutes, include word weeks or minutes. If about, omit word in search and include just number]

- Also look for "/" for fractions. Copy over to display column and format. Enter into the search for years/months/days/hours as a decimal (.5 for 1/2) or as hours if a decimal is unreasonable (1/24 day as 1 hour)

8. Column AX-AZ: Marital Status

- Confirm Marital Status is populated accurately in the Marital Status: Merged [DISPLAYED] column

- In these columns, a hashmark (") indicates the row should be populated with the content of the

previous field

- Rows that are blank should be displayed as “Not recorded”

- When you discover that the [TRANSCRIBED] column is incorrect due to an error with the transcription, edit the [TRANSCRIBED] column to match the original volume; When you discover that the [DISPLAYED] column is incorrect due to an error with the transcription OR misinterpretation of the programming, edit the [DISPLAYED] column so that it is correct.

- In some cases (earlier) search to see if “from cemetery” or other notes are there and that they went into remarks (if they did not, copy them).

- Same above with hours/minutes in the age fields

- Make sure that any ditto marks also in that column are translate correctly as married/single/widowed (Transcribers have treated this differently across volumes, look at original for how to handle)

9. Column BA-BF: Late Residence

- Review for obvious misspellings by transcription team [Common Errors: (1), the initials “SI” (Staten Island) and “LI” (Long Island) appear very similar in 19th-century script and may be mistranscribed)]

- Check that Geocoded value is a logical match for [Transcribed]. If a numbered street address is provided in the geocoded column that DOES NOT match the address written by the original secretaries, this is an error as the system is identifying a business on today’s map with the same name. (You may see this with hospitals, but also with location names that no longer exist).

- In the case of error in Geocoded column, please add FALSE in the Faulty Geocode [ADMIN] column

- [Added 9/13/2021 by Stacy L] In some volumes, ditto marks (“) are used after digits in the “Late Residence: Street Address” column to signify ordinal numbers. (Ex. 267 14” Street means 267 14th Street; 20 41” Street means 20 41st Street) In these cases, you will see the Admin note “Unable to resolve ditto in death place, Unable to resolve ditto in late residence city” in the last column. Please change the display column to show the ordinal number, as described in the above example.

- Use search/global updates to update abbreviations like “Pa” or “SI”

- Look out for issues from transcribers global updates (ex. BlackwellSIIsland)

- Flag in yellow anything to be reviewed manually

- Search for any “s that were not resolved and fix manually

- In street address, search for [“] and replace with [St] where appropriate (or Ave, place, where appropriate)

- Locate where Brooklyn ED had erroneously been carried down because of the duplication intended just from brooklyn and fix, and New York State and East New York

- NORMALIZE N Y, N J, U S etc to NY, NJ, US

- ABBREVS: L I, S I, Brooklyn E D, E N Y, So Brooklyn, Pa, Penn, B I, W I, Cal, Fla, G P, Ill, Bklyn, J C. (Misc states, Va, Me, Miss, Mi, Minn, Mo, La, etc.)E N Y

10. Column BU-BX: Place of Death

- Review for obvious misspellings by transcription team

- Check that Geocoded [GEOCODE] value is a logical match for [Transcribed]. If a numbered street address is provided in the geocoded column that DOES NOT match the address written by the original secretaries, this is an error as the system is identifying a business on today’s map with the same name. (You may see this with hospitals, but also with location names that no longer exist).

- In the case of error in Geocoded column, please add FALSE in the Place of Death: Faulty Geocode [ADMIN] column

- Use search/global updates to update abbreviations like “Pa” or “SI”

- Look out for issues from transcribers global updates (ex. BlackwellSIland)
- Flag in yellow anything to be reviewed manually
- Search for any “s that were not resolved and fix manually
- Locate where Brooklyn ED had erroneously been carried down because of the duplication intended just from brooklyn and fix
- Look out for where the city from Late residence street address has been carried over (Like Bklyn or NY)

For above two fields (Late Residence and Place of Death) there may be some data that was originally written across columns (or overrunning columns) and has thus been transcribed in a way that does not read logically. When you locate issues, fix in the most logical way. (Ex. [Asylum for the] [Insane Blackwell's Island] can be changed to appear in all one column) Sometimes this means duplicating the content in two fields.

- NORMALIZE N Y, N J, U S etc to NY, NJ, US

11. Column CM-CR: Death Date

- Confirm date is populated accurately in the Death Date [DISPLAYED\ column
- In the Month column, a blank row should be populated with the last field that was populated. E.g. October followed by blank rows means every column following should be displayed as October until a row contains November; then displayed as November until a row contains December and so on.
- In the Date column, a hashmark (“) indicates the row should be populated with the content of the previous field
- Where “ULT” appears in the Cause of Death column, it indicates that the month should be the one previous to the current month in which it was written (“ULT” was used by the original writers as short for “ultimate,” meaning “previous” or “last”). In these cases there is a hiccup in the chronological order of the death dates from top to bottom. Therefore, change November to October, for example.

10	1866	October 10, 1866	1866-10-10	REQUI RED	Ult Diarrhoe a
----	------	------------------	------------	--------------	----------------------

E.g. Date should be September 10, 1866

- Similarly, if a two-digit or four-digit number appears in the Case of Death Column, this indicates that the Year of death is different from the current year. In these cases, replace the year for the Date of Death with the one indicated here. (Ex. you may see “1856” or just “56”)

- Scroll through the original volume scan next to the spreadsheet and compare for errors.
- In most cases, the transcriber has not accounted for any years out of the norm (written either above the month, day, or in the cause of death column) Correct using Ctrl+H or Command+Shift+H to update three appropriate fields with correct year
- If the year has been transcribed in the Cause of death column, remove only from the displayed (same with “)
- Check for instances with only year OR only month and year, and make sure all three fields are properly populated
- Look out for carried over years or months that should not be carried over and correct
- Look out/search for [“]s in month field and correct
- Note: If large chunks of the display and search dates have errors (such as with “), you can correct the date display, copy that data into the search field, and reformat the date using Format->Number->2008-09-26
- In early volumes, rather than duplicate the name of the month previous to the current, the writer may have written “Ult” after the day (appearing in the cause of death field). Bill's program has populated the column CR with “Required” in these cases. Update month as appropriate and take note of instances where (“) has

been included to mean Ult but not picked up by program

- If there is NO date of death, scroll over to see if this was a removal and if so, it is properly accounted for in the remarks column

12. Column CS-CT: Cause of Death

- Confirm the displayed column is populated correctly

- In this column, a hashmark (“”) indicates the row should be populated with the content of the previous field

- When you discover that the [TRANSCRIBED] column is incorrect due to an error with the transcription, edit the [TRANSCRIBED] column to match the original volume; When you discover that the [DISPLAYED] column is incorrect due to an error with the transcription OR misinterpretation of the programming, edit the [DISPLAYED] column so that it is correct.

- If the word “ult” appears or a two-digit or four-digit year (as indicated in instruction #11) correct as indicated above and remove said text from the Cause of Death column.

- Review for obvious misspellings using spell check and/or manually

- Look out for issues from transcribers global updates (ex. BlackwellSIland)

- Flag in yellow anything to be reviewed manually

- Search for any “s that were not resolved and fix manually, also “Do”

- If a column where ult is used, remove “ from display (or year, if carried over)

- When the word “difficult” appears (ex. Difficult labor), programmatic changes have treated this as “Ult” and removed the word “difficult” from displayed. Correct.

13. Column CU-CV: Undertaker

- Confirm the displayed column is populated correctly

- In this column, a hashmark (“”) indicates the row should be populated with the content of the previous field

- When you discover that the [TRANSCRIBED] column is incorrect due to an error with the transcription, edit the [TRANSCRIBED] column to match the original volume; When you discover that the [DISPLAYED] column is incorrect due to an error with the transcription OR misinterpretation of the programming, edit the [DISPLAYED] column so that it is correct.

- Look out for issues from transcribers global updates (ex. BlackwellSIland)

14. Column CW-CX: Remarks

- Confirm the displayed column is populated correctly

- Review for obvious misspellings using spell check and/or manually

- Look out for issues from transcribers global updates (ex. BlackwellSIland)

- Scan for any other extraneous errors

- In some cases, records with [“] in place of ordinal numbers cause the value to not be carried over into the displayed column. In that case, easiest thing to do is replicate the Transcribed column and then manually fix all of the [“]s meant to indicate duplicates

- In some cases the interment number has been duplicated in this field, remove it

15. Column CZ: Has Diagram

- No action needs to be taken regarding this column

- This should only be populated if the Grave column says “see R column.” (Or sometimes See register) (Should only be in vols 39+)(*Correction—appears earlier)

- Should be fixed from earlier actions, but do a search for all instances of true and make sure they align with “See R column” or the like in remarks

16. Column DB-DC: Needs Review

- When either/both column is populated with "TRUE" go to the original and check the transcription of every column in this row; make corrections as necessary

- Ignore

Providing Access to the Unexpectedly Rich Records of Brooklyn's Green-Wood Cemetery

Appendix A | Review Log Sheet

NEH Review Log Stacy																		
File Edit View Insert Format Data Tools Extensions Help																		
100% View only																		
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
VOLUME	Completed	CATEGORIES																Volume Notes
		Interment ID	Interment Dates	Name	Lot	Grave	Birth Place	Age	Marital Status	Late Residence (City)	Late Residence (Street)	Place of Death	Death Date	Cause of Death	Undertaker	Remarks		
VOL 22	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		remarks should be cw and ca
VOL 26	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 27	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 32	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 38	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 33	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 40	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		"see r column" fixed; periods in lot numbers removed; those in grave
VOL 28	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		Seem to be large chunks where notes were not recorded in remarks column. This had the "A" name issue
VOL 34	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 30	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 29	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		Seem to be large chunks where notes were not recorded in remarks column
VOL 39	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 31	?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 19	5/3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		Place of Death column appears to be fully copied from late residence city
VOL 35	5/12	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 36	5/22	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
VOL 20	5/25	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		First 7 pages are incomplete data (due to damage in volume). Removed and not sent to Bill for accession. Remarks column had interment number copied over
VOL 37	5/31	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		Removals located; Diagrams fixed
VOL 24	6/3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		

Appendix B | Figure 1

- 1) Here are a few examples of fields defined in the mapping file:

```
...
"name_last": {
  "type": "text",
  "fields": {
    "keyword": {
      "type": "keyword",
      "ignore_above": 256
    }
  },
},
"is_lot_owner": {
  "type": "boolean"
},
"interment_date_year_transcribed": {
  "type": "long"
},
...
```

- 2) Here is an example of a single registry interment entry encoded using this mapping:

```
{
  "interment_id": 164900,
  "registry_image": "Volume 22_002",
  "interment_date_month_transcribed": "August",
  "interment_date_day_transcribed": 10,
  "interment_date_year_transcribed": 1873,
  "interment_date_display": "August 10, 1873",
  "interment_date_iso": "1873-08-10",
  "name_transcribed": "Ellen Owen",
  "name_display": "Ellen Owen",
  "name_last": "Owen",
  "name_first": "Ellen",
  "name_middle": "",
  "name_salutation": "",
  "name_suffix": "",
  "is_lot_owner": false,
  "gender_guess": "F",
  "burial_location_lot_transcribed": "8259",
  "burial_location_lot_current": "8259",
  "burial_location_lot_previous": "",
  "burial_location_grave_transcribed": "131",
  "burial_location_grave_current": "131",
  "burial_location_grave_previous": "",
  "birth_place_transcribed": "Ireland",
  "birth_place_displayed": "Ireland",
  "age_years_transcribed": "67",
}
```

```
"age_months_transcribed": "",
"age_days_transcribed": "",
"age_hours_transcribed": "",
"age_display": "67 years",
"age_years": 67,
"age_months": 0,
"age_days": 0,
"age_hours": 0,
"marital_status_married_transcribed": "W",
"marital_status_single_transcribed": "",
"marital_status": "Widow",
"residence_place_city_transcribed": "NY",
"residence_place_city_display": "New York",
"residence_place_street_transcribed": "St Lukes Hospital",
"residence_place_street_display": "St Lukes Hospital",
"death_place_transcribed": "New York",
"death_place_display": "New York",
"death_date_month_transcribed": "August",
"death_date_day_transcribed": 8,
"death_date_year_transcribed": 1873,
"death_date_display": "August 8, 1873",
"death_date_iso": "1873-08-08",
"death_date_ult_month": "",
"cause_of_death_transcribed": "Phthisis",
"cause_of_death_display": "Phthisis",
"undertaker_transcribed": "S Merritt",
"undertaker_display": "S Merritt",
"remarks_transcribed": "",
"remarks_display": "",
"burial_origin": "",
"has_diagram": false,
"cemetery": "Green-Wood Cemetery, Brooklyn, NY, USA",
"registry_volume": 22,
"registry_page": "002"
}
```

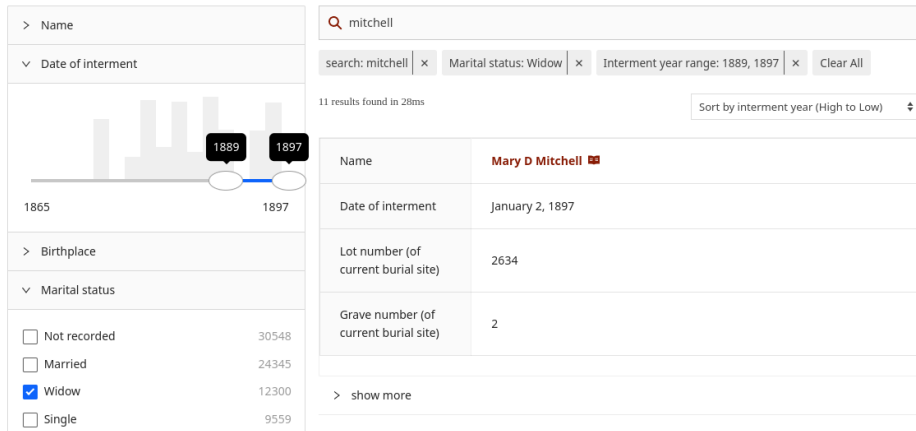
Appendix B | Figure 2

Text Searches

When you type search terms in the main search box you are currently searching the following fields:

Technical Data Field Name in Spreadsheet	Examples
name_last	Currier
name_first	John
name_transcribed	John A Currier
name_display	John A Currier
name_middle	A
burial_location_lot_current	736 &C
burial_location_lot_previous	8839
burial_location_grave_current	3
burial_location_grave_previous	6
birth_place_displayed	England
birth_place_transcribed	U S
residence_place_city_display	Brooklyn
residence_place_city_transcribed	New York
residence_place_street_display	217 - 7th Ave
residence_place_street_transcribed	62 Hoyt St
death_place_display	Brooklyn
death_place_transcribed	St Vincents Hosp N Y
death_date_display	February 21, 1897
death_date_iso	1897-02-21
interment_date_display	February 23, 1897
interment_date_iso	1897-02-23
cause_of_death_display	Phthisis
undertaker_display	McLean & Dougherty
undertaker_transcribed	John F Fagan
remarks_display	Removal From New York
marital_status	Not Recorded, Single, Married, Widow

Appendix B | Figure 3



Appendix C | Figure 4

1) Date of Interment Slider

The screenshot shows a search interface with a 'Date of interment' filter. The filter includes a histogram and a slider set to the range 1865-1897. Below the slider is a list of birthplaces with checkboxes and counts. To the right, search results for 'Jacob E Young' and 'John Walton' are displayed, both with interment dates of January 1, 1897.

Name	Interment Date	Lot Number	Grave Number
Jacob E Young	January 1, 1897	14964	507
John Walton	January 1, 1897	14964	711

2) Date of Interment Source code:


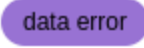
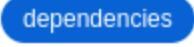

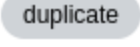
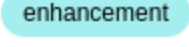

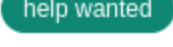
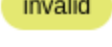
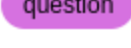

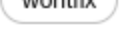
```

<Panel header="Date of interment" key="8">
  <DynamicRangeSlider
    componentId="interment_year_facet"
    dataField="interment_date_year_transcribed"
    rangeLabels={(min, max) => ({
      start: min,
      end: max,
    })}
    stepValue={1}
    tooltipTrigger={"always"}
    showHistogram={true}
    showFilter={true}
    interval={2}
    react={{
      and: ["CategoryFilter", "SearchFilter"]
    }}
    loader="Loading ..."
    filterLabel="Interment year range"
    includeNullValues
    URLParams={true}
  />
</Panel>

```


Appendix D | Figure 5
Github Categories

In addition, we used 12 labels for our GitHub issues in order to group and organize them into common categories:

	Something isn't working
	Something in the data doesn't seem right
	Pull requests that update a dependency file
	Improvements or additions to documentation
	This issue or pull request already exists
	New feature or request
	Good for newcomers
	Extra attention is needed
	This doesn't seem right
	Further information is requested
	task
	This will not be worked on