THE AMERICAN SOLDIER COLLABORATIVE DIGITAL ARCHIVE

WHITE PAPER

National Endowment for the Humanities
Division of Preservation and Access
Humanities Collections and Reference Resources

Grant PW-253766-17
Project Director: Edward J. K. Gitre

Lead White Paper Authors:
Edward J. K. Gitre
Kurt Luther

Virginia Tech
July 31, 2018

Table of Contents

Table of Contents 1

Project Activities and Accomplishments 2

- Project Collaboration 3
- Digitization of Verbatims 4
- Transcriptions of Handwritten Free-comment Responses 4
  - Initial Deployment and Testing of Incite 4
  - Transition to Zooniverse 5
- Assessment of ARB Survey Data and Planning for Data Conversion 7
  - The Roper Center for Public Opinion Research 8
  - Social and Decision Analytics Lab at Virginia Tech 9
- Historical Context for Digital Archive 11
- Additional Partnerships and Agreements 12

Audiences 13

Evaluation 14

Project Continuation 16

Long-Term Impact 19

Grant Products 20

Appendix A: Project Accomplishments 22

Appendix B: Zooniverse Site 25

Appendix C: Feedback from Student-Testers & Sample Feedback 30

- Feedback on Process 30
- Feedback on Tutorial 30
- Feedback on Field Guide 30

Appendix D: Summary of Feedback from Zooniverse Review 32

Appendix E: Press Related to the Public Launch of Zooniverse Site 33

Project Activities and Accomplishments

With funding provided by an NEH Foundation HCRR grant, PW-253776, our team finished the first phase of The American Soldier Collaborative Digital Archive. The goal of this project is to create a free, public website to house and disseminate an incomparable collection of historical sources that capture the experience of Americans who served in the U.S. Army during the Second World War.

During that global conflict, the U.S. Army administered social surveyed to approximately half a million service personnel to better understand their adjustment to military service and, in conjunction, to help improve the Army's performance using feedback from these personnel. The surveys administered by the U.S. Army Research Branch (ARB) consisted of more than multiple-choice questions, however. Many provided space for the soldiers to record their opinions and personal experiences. Over the course of the conflict, ARB collected over 65,000 pages of handwritten "free comments." After the war, these pages were photographed separately from the rest of the survey pages that contained multiple-choice answers. Our aim is to make these valuable orphaned free-comment documents internet accessible in a searchable format as well as to reunite them with their parent surveys. By doing so, we hope to reconstitute the most comprehensive portrait of the largest and most diverse "citizen-soldier" Army in U.S. history.

Our interdisciplinary team opted to take a phased approach to the project, owing to its complexity. We needed, first, to digitally scan these handwritten documents from microfilm rolls. Afterward, we would need to have these digitized sources transcribed to render them text searchable for accessibility purposes. Only after these two steps were more or less completed could we then reunite the transcribed documents with corresponding ARB survey datasets. Complicating matters further, these extent ARB datasets are also relatively inaccessible to non-specialists, for they only exist as ASCII-formatted data files. On top of this, the surveys were designed to be anonymous. Indeed, the majority of photographed pages have nothing but a soldier's handwritten response. This, then, was our challenge: could we reconstituting the relationship that once existed across these datasets and humanities sources and then present the resulting information in a compelling way to the broadest possible audience through an open-access website? This intellectual and technical challenge, in service to our greater goal, has driven our efforts.

We focused on the following objectives during our year of planning: (1) digitize the microfilmed handwritten documents that were composed by soldiers, anonymously, during the war; (2) assess a crowdsourcing web-based platform for transcribing these documents and, after selecting

Zooniverse, create a project website; (3) ascertain the required resources and protocols to convert the social surveys accompanying these handwritten, verbatim responses to a modern statistical format; (4) determine the best methods for reconnecting these humanities and social science sources; and (5) create a full implementation plan. We have completed the majority of these activities and have made considerable progress in other areas, as we describe below. (See Appendix A for a comparative list of proposed activities and project accomplishments.)

Project Collaboration

Project planning under the NEH grant commenced in late spring 2017 with the hiring of a Virginia Tech graduate student in computer science, Nai-Ching Wang. Much of the initial work during the first semester of planning was conducted through individual or expertise-clustered meetings, as well as through phone calls, via email, and with Google Meet. These consultations included colleagues located at Virginia Tech's metropolitan D.C. campus; Dr. Amanda French, now located in Durham, North Carolina; Zooniverse co-Investigator Dr. Laura Trouille; and with other external partners and interested parties. We held our first all-stakeholders meeting in November 2017 in Blacksburg, Virginia. Dr. Peter Enns, Executive Director of the Roper Center; Kathleen Weldon, Director of Data Operations and Communications; Dr. French; and the project's Advisory Board participated remotely through video teleconferencing.

After the November meeting, we interspersed ongoing, informal collaboration with more expertise-clustered meetings. In January 2018 PI Gitre also travelled to New Orleans to discuss the project with staff and leadership of the National World War II Museum. Early in 2018, PI Gitre and our technical team started to meet weekly to design our project's Zooniverse site and prepare it for formal review. We then held biweekly meetings as we collectively wrote this White Paper, our Implementation Plan, and planned the launch of our project on Zooniverse with a commemorative daylong transcribe-a-thon corresponding with the celebration of VE (Victory in Europe) Day. After we had drafted and circulated both the White Paper and Implementation Plan, we convene a final stakeholders' meeting to receive feedback, especially on the latter, intending to submit a follow-up proposal for implementation funding from NEH. Scheduled meetings were typically managed by the NEH-funded graduate student, Mr. Wang.

This flexible approach worked reasonably well, especially as staffing changes became necessary (as we explain below) and as more members were added to the team. Enthusiasm for the project has helped to drive momentum throughout our year of planning, building a strong sense of communal investment and instilling a strong desire to create an outstanding digital project for the public's benefit.

Digitization of Verbatims

To lay a foundation for The American Soldier, Gitre used Virginia Tech research funds to have the National Archives digitize two of forty-four ARB microfilm rolls containing the free-comment responses in May 2015. In our original NEH grant application, we proposed using internal Virginia Tech funding to digitize the remaining rolls. In January 2017 Virginia Tech initiated the process with Ms. Denise Henderson, the Internal Digitization Coordinator for the Office of Innovation at the National Archives. Given our team's interest in the collection and the inherent public value of the records, NARA's Office of Innovation offered to digitize the collection at no cost.

In the summer of 2017 NARA scanned all 65,000-plus images as 400-DPI JPEGs. In September NARA transferred the entire collection to an external hard drive and shipped it to Virginia Tech. Upon receipt our team made multiple copies of the scanned rolls for preservation purposes on local hardware and on Google Drive with 2-factor security authentication. An inspection of the scanned documents found that images from a quarter of the rolls had been cut off and therefore required rescanning. NARA reprocessed the defective scans, reviewed the new images for quality control purposes, and sent us a copy. NARA is preparing to upload all the images into its public catalog. Meanwhile, we have uploaded documents sets to the Zooniverse crowdsourcing platform, where they are already being transcribed.

Transcriptions of Handwritten Free-comment Responses

Initial Deployment and Testing of Incite

We originally proposed using Incite, a crowdsourcing plug-in for the Omeka content management system developed by Dr. Kurt Luther, a project co-PI, to support the transcription and annotation of the handwritten survey responses. Gitre had already incorporated Incite and Omeka into course-related activities in six of his World War II courses. While this experience had yielded important insights, during the HCRR grant period we tested the viability of Incite for the project in a more structured fashion, while generating additional seed data for review. The grant-funded graduate student deployed Incite in two additional undergraduate history courses taught by another professor at Virginia Tech. Sixty-five students participated in the testing program, with each student transcribing and tagging eight documents as well as reviewing transcriptions completed by their peers. A total of 520 documents were transcribed as well as peer reviewed.

Our preliminary review found that students produced high-quality transcripts and annotations, but Incite's learning-oriented "class-sourcing" approach, which is optimized for classroom usage,

was unlikely to scale up to process the complete dataset in a reasonable timeframe. We considered narrowing our focus to a subset of the surveys, but ultimately decided to keep the original goal of processing the complete dataset, which we felt would yield a more valuable resource for the scholars, educators, and the general public.

Transition to Zooniverse

After exploring a variety of crowdsourcing tools, platforms, and communities, we found that Zooniverse had the greatest potential to help us achieve our goals. Zooniverse describes itself as "the world's largest and most popular platform for people-powered research," and one of its strongest selling points is a community with over one million registered volunteers. Zooniverse was originally built for citizen-science projects in domains like astronomy and zoology, but having built up a critical mass of users and a dedicated staff, it has expanded in recent years to support projects in other research areas, including the social sciences and history. Among the latter are two significant World War I-related projects, Measuring the Anzacs and Operation War Diary. The platform leverages redundant user contributions and aggregation algorithms to ensure that the quality of crowd contributions meets or exceeds what trained specialists would produce.

Zooniverse's geographically distributed staff are primarily based at the University of Oxford, the Adler Planetarium, and the University of Minnesota, but they support research teams at other institutions to help them build and launch new crowdsourcing projects using the Zooniverse platform. We established a relationship with Zooniverse through Co-PI Luther's previous collaborations with Dr. Lucy Fortson, a professor at the University of Minnesota and a Zooniverse co-founder. Fortson and her colleagues at Zooniverse expressed great enthusiasm for The American Soldier, owing to the popularity of its subject matter, the characteristics of the dataset, and the type of crowdsourcing work we have in mind. We worked closely with Zooniverse co-investigator Dr. Laura Trouille to build a fully functional prototype of The American Soldier on Zooniverse. This development process required completing several key tasks. First, we imported all 65,000-plus scanned survey responses from the National Archives into the Zooniverse project builder, so that each could be viewed and analyzed by the crowd. Next, we experimented with a variety of workflows for guiding users through the process of transcribing and annotating a survey response.

Designing workflows proved to be more complicated than we first anticipated. While the scanned handwritten documents initially appeared to follow a predictable pattern, our more thorough review of the digitized rolls revealed many minor variations that could defeat a simple set of instructions. We created a coding scheme and systematically analyzed each roll for variants, identifying over 20 different formats. We then designed and tested a unique workflow for each

variant, including a customized tutorial, step-by-step instructions, and example content. Fortunately, Zooniverse's project builder allows customized workflows to be attached to individual data sets (for our purposes, each scanned microfilm roll). This customization will not only guarantee greater transcription accuracy but will also ensure that Zooniverse transcribers are not overwhelmed with complex workflow tasks and instructions.

Our evaluation and analysis of the entire collection of scanned microfilm rolls yielded other benefits. Although the handwritten responses are all anonymous, a large number of the scanned documents contain serial numbers or codes that were added by ARB staff. Additionally, many of the handwritten responses appear on survey pages that also have multiple-choice questions with responses. PI Gitre consulted another content specialist, the military sociologist Michael Hughes, also at Virginia Tech, to determine the nature of these codes and to assess their value. We concluded that it would be best to capture as much as possible from the scanned documents. This more thorough approach promises to improve post-transcription reconstruction, as it will allow us to make more approximate and, for some scanned rolls, perhaps very precise connections between free-comment responses and survey responses. With this goal in mind, we have designed Zooniverse workflows that not only ask volunteers to transcribe handwritten free-comment responses but also the adjacent multiple-choice questions and answers as well as serial numbers and ARB assigned codes where they exist.

We were able to conduct our assessment of Incite and of Zooniverse as an alternative platform because we made the decision to use NEH grant funding to hire a computer science graduate student rather than a history graduate student as initially planned. Specifically, we hired Luther's graduate student, Mr. Wang, who is the lead developer and designer of Incite and has expertise with the Omeka platform and crowdsourcing transcription projects. Hiring Mr. Wang allowed us to do far more under the Foundation grant than we had proposed. He administered our review of Incite and helped to develop a fully-functional project prototype on Zooniverse, creating workflows for each document variant and helping to create supporting content that might attract and retain Zooniverse users. He assisted PI Gitre with the preparation and posting of detailed "Field Guides" for the Zooniverse community to reference, including transcription guidelines and tutorials; guides for era-specific military rank, abbreviations, and acronyms; and an FAQ page. As well, Gitre and Wang created "Team" and "About" pages to round out the Zooniverse site. (See Appendix B for screenshots of the Zooniverse site.)

We first tested our Zooniverse project and workflows with students in PI Gitre's World War II Digital History course. Student testers were asked to provide feedback via a survey form that asked them to evaluate three main aspects of our project design, including task process

(workflow), tutorial, and field guide. In general, the survey affirmed that the Zooniverse project we created was a meaningful and engaging way for our student audience to interact with these rich primary sources. We provide some of their feedback in the Evaluation section of the White Paper. Project workflows were also tested and evaluated by project stakeholders, including the Advisory Board and a National Archives staff member. While we recommend the Zooniverse platform for digital humanities projects, we would counsel other humanities scholars to allocate sufficient time and resources for platform design and testing.

With our internal review complete, we submitted The American Soldier to Zooniverse so that it could undergo the platform's rigorous review process to become a Featured Project. Digital projects that earn Featured status receive the platform's highest level of promotion and the attention of the more than one million members of the community. We describe in the Evaluation section the feedback we received during what was a very successful project review. Based on discussions with Dr. Trouille and on statistics from related Zooniverse projects, like Operation War Diary and Mutual Muses, we estimated our entire corpus of 65,000-plus documents could be transcribed and annotated within a two-year window. We considered timing the launch of the Zooniverse site to coincide with the Implementation phase of the project, but we decided it would be better to initiate the Zooniverse project while we still had our NEH-funded computer science graduate student on the team. Using Zooniverse's projected schedule for review and approval, we planned the public launch to coincide with the anniversary of VE Day—May 8, 2018—commemorating the end of the war in Europe.

Assessment of ARB Survey Data and Planning for Data Conversion

From the outset of this project, it was clear that for all the wealth of information contained in the free-comment responses to make the most of these unique historical records they needed to be reunited, as best they could, with the original ARB surveys, as we have mentioned. While soldiers could write freely on any topic or issue, their responses were implicitly and, in some instances explicitly, guided by the surveys themselves—by the concerns that motivated survey design, by the sorts of multiple-choice questions posed, by the way individual questions were worded, etc.

Reuniting the survey commentaries with the rest of surveys would yield invaluable material not only about soldiers' opinions, behaviors, expectations, and experiences, but also their biographies. Soldiers were quite routinely asked about their education, time in service, training, so on and so forth. ARB staff also coded certain surveys during distribution to differentiate the unit and location of respondents, and, in a number of instances, to record respondents' race. While none of the free-comment responses, that we are aware of, note a soldier's name, a great deal can nevertheless be learned from the complete surveys by creating a composite picture of

individual respondents, and by situating these respondents in a larger social, institutional, and cultural context.

The Roper Center for Public Opinion Research

With this in mind, Gitre approached the Roper Center early in planning as a potential partner. In 1978, ARB's successor, the U.S. Army Research Institute for the Behavioral & Social Sciences (ARI), contracted with Roper, then associated with Williams College, to transfer the Research Branch's survey data from IBM punch cards to computer tape and to produce machine-readable frequencies and documentation. When the work was complete, ARI received a copy of the tapes and the codebooks, the Roper Center retained copies, and, at the ARI's request, the National Archives and Records Service, today's NARA, also received copies for archiving. At a later date, Roper transferred the data yet again, from computer tape to ASCII files. Several years ago, NARA did the same. While ASCII is preferable to tape, for long-term preservation, data in this format is still of limited use. ASCII formatting does not allow for analysis with current statistical analysis programs nor, crucially for our purposes, for designing and creating user-friendly tools.

Roper seemed an ideal partner, given the Center's reputation and its ability to convert the files according to best practices, in addition to its long history with this particular survey data. The project team at Virginia Tech along with staff at the Roper Center needed to address two critical issues for our joint venture to advance. First, on the technical side, we needed to evaluate ARB ASCII data files and assess the necessary resources for their conversion to a usable modern format that would allow us to more easily sort, compare, and analyze ARB's wartime data (via cross-tabs, for instance). In September, Gitre identified for Roper three data sets for test conversion and analysis. This work was completed in early November 2017, with Roper producing data files in four different formats (DAT, DTA, SAV, and XLSX). The documentation files for the ARB studies are very complete, compared to other data from this period, and while the metadata fields were limited by the original storage system, the metadata included in the Roper database was sufficient for conversion purposes.

Conversion of ASCII files requires the ability to read and understand the original documentation, including complex skip patterns, and to troubleshoot problems that may have arisen in the file conversion from column binary data formatting. Fortunately, the codebooks for the surveys include frequencies, which provided Roper with the necessary information for quality assurance checks. The Center converts ASCII files by first using Excel to lay out the variable locations and variable labels and then by developing a Python script to convert the spreadsheet to SPSS, CSV, and Stata formats. Afterward, the resulting sample files are checked against the original materials. Data of this age often reveal unforeseen problems during conversion, such as

incomplete documentation, which can slow the process. This automated method is nevertheless faster and more accurate than manual coding in SPSS or Stata. Based on this test data, Roper estimated that it would take approximately eight hours to convert each data file, with an additional hour for verifying conversion accuracy.

The second issue proved more complex. A non-profit organization, Roper operates on a membership model, with nearly three-hundred member institutions, which includes academic libraries, media organizations, government organizations, policy think tanks, and nonprofits. Virginia Tech Libraries is itself a member. Valuable as The American Soldier survey files are to the Center, they represent but a small sampling of the myriad 21,000 dataset files member institutions download each year through Roper's members-only website. Not only had Roper agreed to convert these ARB ASCII data files, as our NEH grant application outlined, but the Center had also agreed to build an online tool for our archive, similar to the one that is available to its own institutional members, called iPOLL. The tool they committed to designing would allow users of our archive to access and explore survey data in a highly customized fashion.

While Roper was keen to contribute, all parties knew that we would need to address the constraints of its membership model. Re-dissemination of data, for instance, is disallowed by its current terms and conditions, even in instances where Roper does not itself hold the original copyright. Furthermore, while Roper distributes ARB data files only to its members, NARA simultaneously holds and now freely distributes these same government records by way of its online catalog. These negotiations necessitated an investigation of the IP rights of ARB records. During the assessment, Gitre consulted with NARA staff, who conducted their own internal provenance and custodial-history review. This inquiry, completed in January 2018, affirmed that the files distributed via NARA's catalog are indeed federal government records and are therefore open access. This key issue resolved, we could proceed with planning.

Social and Decision Analytics Lab at Virginia Tech

Led by NEH's strong preference for free, public access for the projects it funds, and having ascertained the open-access status of NARA's records, the team based at Virginia Tech came to the conclusion that Roper's terms and conditions would constrain our ability to create a free, public-access website that optimized the humanities and social science sources that we intended to integrate into our digital archive. With the Advisory Board's approval and with the approval of our senior program officer at NEH, PI Gitre looked for an alternative partnership to undertake the conversion of ARB survey files. In late January he approached Dr. Sallie Keller, a professor of statistics at Virginia Tech who is also director of the university's Social and Decision Analytics Lab (SDAL).

Founded in 2013, SDAL has developed world-class statistical and data science capabilities and has formed long-term partnerships with a diverse range of non-profit, government, and corporate entities, from Arlington County, Virginia, to Procter & Gamble to the U.S. Department of Housing and Urban Development. Serendipitously, included among SDAL's current partners is none other than ARB's successor, the U.S. Army Research Institute for the Behavioral & Social Sciences. Today's ARI has contracts already with SDAL to undertake research that resembles some of what the Research Branch executed during the Second World War. Director Keller agreed to assist our team, starting with the conversion of ARB's ASCII data files. Sample files were shared with Dr. Aaron Schroeder, a senior data research scientist at SDAL, in January, 2018, to determine the best processes for converting the files and to estimate the necessary resources to convert the entire collection.

Dr. Schroeder conducted a successful first iteration of code for both reading metadata from the codebooks and parsing the ASCII data files using that metadata. The data files were found to be more complex than initially thought: the files are structured with a variable number of lines per respondent (e.g., if there were 3 "cards" of questions administered, then there are 3 lines of answer codes per respondent in that particular data file. If there were 2 cards, then 2, etc.); also, each line of card responses has a different set of column widths and, therefore, each line per respondent has to be parsed differently (e.g., respondent number is a 3-digit code and is part of the first line of response codes represented in positions 5-7, but this is not the same in line 2 if it exists, 3 if it exists, so on and so forth).

However, processing the files programmatically is still possible as the necessary information is contained in the provided codebooks (e.g., the columns widths required for each question sub-divided by question card). As the codebooks are ASCII text files, not ordered data files, it is necessary to write string parsing code to locate both the column information and the associated questions. Therefore, in order to parse the data file (the respondent's answers) correctly for each survey, the code must accurately determine, by first parsing the codebook, the number of cards administered and the column widths used for each card.

After successfully testing the initial code to parse both the codebooks and data files, Schroeder estimates that it will take approximately another 2 days to create additional code that will also pull the question text from each survey codebook and align the question text with associated columns in each survey. After this initial coding, it is estimated that each additional survey will take 2 days each to be successfully processed, namely, to allow for any necessary adjustments to the code for each particular survey and to verify that import is being handled correctly.

Historical Context for Digital Archive

Our team, from the very beginning, wanted The American Soldier to be more than an online repository for historical World War II sources. For users to get the most out of our site, to appreciate the diversity of interests and concerns the Research Branch studied while surveying half a million military personnel, the site would need to place these records into historical, social, political, economic, and institutional context. Our sense of this necessity was confirmed by reading the NEH HCRR white paper written by the team behind the Bryn Mawr-based *History of Women's Education Open Access Portal Project*. "The commonly expressed wish [was] for more robust contextualization of individual collection items," they highlighted (Pumroy et al., 2015). We felt that not only would individual items from our collection require contextualization, but so, too, would the U.S. Army Research Branch itself. As important as its contribution to the war effort was, no adequate history of ARB exists, and too little is known about its research beyond scholars with subject-related interests.

While we had intended to hire a graduate student in history to help prepare bibliographies, as mentioned above, we took advantage of the opportunity to hire a computer science graduate student who could help us create and launch a Zooniverse site for transcriptions. Gitre was able to use a second graduate assistant assigned to him by the Department of History at Virginia Tech to conduct some of the research we outlined in our original grant application, such as locating the Army bases where surveys were conducted, identifying keywords and research topics and concerns, so on and so forth. PI Gitre conducted research himself at the National Archives in College Park, MD, in May 2017, locating records not delineating in NARA's finding aid for ARB records (Reference Information Paper 78). While certain ARB records are well documented, such as the survey files and free-comment responses, large volumes of records elude easy discovery, owing to War Department's rapid expansion and reorganizations during the war, to the complexity of records management by the U.S. Army and Federal government, and to the vast tonnage of documents the war effort produced and was subsequently preserved. When ARB staff left the Branch, they took with them research materials and other documentation of historical significance.

Throughout the grant tenure, Gitre continued to search for, and found, relevant archival materials, consulting with other researchers and a former NARA staff member who worked with these records as far back as the 1980s. For all these reasons, our team decided that the work of contextualizing the historical materials to be disseminated through our digital archive would require additional research, time, and resources beyond what the planning grant provides. We felt comfortable shifting this work to the next phase of the project, which we outline in our Implementation work plan.

Additional Partnerships and Agreements

Wanting to create a robust digital archive that would appeal to, and be used by, a wide range of people—one that would attract a national, if not international, audience—we worked to cultivate external relationships with other institutions with the benefit of the NEH Foundation grant. NARA's assistance has been tremendously helpful, starting with the initial digitization of microfilmed ARB free comments. Very little was needed to persuade NARA's Office of Innovation of the value of these unique historical records. NARA has signed an LOI with PI Gitre to coordinate public release of project-relevant digitized records through the National Archives Catalog, to promote our Zooniverse crowdsourcing transcription project to NARA's citizen archivist community, to provide input on Zooniverse workflows, to assess the feasibility of accepting transcription data back into the National Archives Catalog. NARA has committed to participate in this project in other regards as required. In accordance with Federal policy, NARA has not received funds from this NEH HRCC Foundation grant.

We have reached out to other institutions invested in the history and memory of the Second World War, including the George C. Marshall Foundation, in nearby Lexington, Virginia; the National World War Two Museum in New Orleans, Louisiana; and the University of North Texas Libraries; and we continue to build on the existing relationship Virginia Tech has with Florida State University's Institute of World War II and the Human Experience, whose director, Dr. Kurt Piehler, is on our Advisory Board. External partners helped with our transcription drive, by hosting and promoting transcribe-a-thons, and will help to promote the digital archive when it is complete. During the Foundation grant, the Project Director also secured two additional historians to serve as advisory board members, Dr. Tom Thomas Guglielmo, Associate Professor of American Studies, George Washington University; and Dr. Beth Bailey, Foundation Distinguished Professor and Director, Center for Military, War, and Society Studies, University of Kansas.

External institutions are also contributing digital assets. To help us build a more robust archive, UNT Libraries, for instance, has granted us permission to use their extensive holdings of "Newsmaps," which were published by ARB's umbrella War Department organization, the Information and Education Division. The George C. Marshall Library is contributing its high-resolution scanned images of ARB's published reports, intended for wartime distribution across echelons, entitled *What the Soldier Thinks*. We also worked with Princeton University Press and the Social Science Research Council on rights access to a four-volume series published by ARB's research director Samuel Stouffer, known as *The American Soldier* (1949-1950). At the end of the grant period, we negotiated a Creative Commons Attribution NonCommercial ShareAlike 4.0 license for the volumes.

Audiences

This project began as a pedagogical innovation. In early 2016, PI Gitre collaborated with Co-PI Luther to deploy a digital history project in his World War II courses at Virginia Tech. Dr. Gitre wanted his students to engage in a new way with primary sources produced by soldiers during the war, while also seeding what is now The American Soldier Collaborative Digital Archive. Luther's graduate student Wang uploaded free-text responses from two scanned microfilmed rolls to an Omeka-based project website and helped to administer the project in the Project Director's classes. Luther, in turn, included Dr. Gitre's classes in proof-of-concept testing of the crowdsourcing Omeka plugin that Wang was developing for K-12 classroom application, Incite (http://incite.cs.vt.edu). Over the next year and a half, students in six Virginia Tech World War II courses transcribed, tagged, and analyzed approximately 2,200 handwritten documents. During the current year, as we also explained earlier in this White Paper, a seventh class contributed to the project hosted on the Zooniverse platform.

We believe that this student-centered approach has yielded exceptional advantages. We were able to scale the project from our initial 2016 "seed" up to an inter-institutional collaboration with relative ease; we simply incorporated additional students from scheduled courses, progressively expanding our audience. There were other benefits, however. Our goal is to create a project that will attract a broad audience, one that will come to The American Soldier with a wide spectrum of interests, historical knowledge, and technical skills, similar to the sort of student one might encounter in a class at a public university. Based on our testing experience, we are confident the crowdsourcing site we have already created will garner public attention and that the humanities sources highlighted by the project will interest not only students, but a diverse public that is invested in the memory of World War II, this time-period in our nation's history, and to the American experience of war writ large.

Still, we intend to maintain a commitment to this initial pedagogical orientation. Students beyond Virginia Tech will contribute to our crowdsourcing transcription efforts, as we note in our Implementation plan. Too, the Virginia Tech professor of history and social science education David Hicks will help to create curriculum using digital archive primary sources. Dr. Hicks has previously worked with co-PI Luther on another digital history project, *Mapping the Fourth of July in the Civil War Era*, funded by the National Archives/NHPRC. On this project, Dr. Hicks supervised the development of 10 assignment guidelines for the high school level and college level that are freely available online (http://incite.cs.vt.edu/m4j/incite/_help/teachers). These guidelines employ a historical source analysis scaffold called SCIM-C to support project assignments and assessments that focus on the analysis of historical sources as part of a historical

inquiry. The assignment guidelines we create will make the project as scalable and as appealing as possible for teachers to authentically and appropriately incorporate the crowdsourcing platform into their classes. A curriculum development plan has been included in our Implementation plan.

We aim to reach a broad, diverse audience with our project, but it is also true that a site designed for "everybody" is often not especially well-designed for anybody. To that end, we employed a user-centered design technique called *personas*, which are fictional users who represent concrete real-world needs and attitudes of intended audiences.[1] We developed four personas, each representing a primary user group that our project will serve: Amy (Student users), Bill (Enthusiast users), Carol (Scholar users), and Dave (Teacher users). We generated and refined these personas during our virtual stakeholder meetings with supplementary in-person meetings and email exchanges. We used the personas to anchor our requirements gathering and frequently referred to them in design discussions. The detailed personas are included in our Implementation plan, along with preliminary design and interaction requirements for the project's public website.

Evaluation

This project has been evaluated throughout development, beginning with the integration of this project in PI Gitre's World War II courses in 2016. Each semester, Gitre solicited feedback from students on all aspects of the project, both technical and intellectual. Approximately 40 to 70 students were enrolled in each of the sections that contributed. We altered our Omeka website after each semester, based on feedback from these students and also on our evaluation of project tasks, workflows, and objectives. When asked to evaluate these courses and the digital project, students highlighted the satisfaction they received by knowing that their efforts would have a public benefit. One student wrote, "I was a big fan of the transcription project because I was a part of history. I thought it was so cool that I was able to take an actual document from the WWII era, figure out what on earth they were writing about, and submit it for revising before it's made searchable online."

Another student spoke of how the free-comment responses helped them to understand a war that their grandfather had fought in but did not want to speak about. "My grandfather was a soldier in World War II, and growing up I never got to hear about his experience because he

[1] Lene Nielsen, *Personas – User Focused Design* (*Human–Computer Interaction Series)* (New York: Springer Science & Business Media, 2012).

refused to talk about what he saw, and he rarely told my dad about it either.… It was very eye opening to get to do the digital history project because it brought a whole new perspective to how the soldiers actually felt about being in the war. They are very much made out to be 'heroes who wanted to give their lives to serve' but that wasn't always the case and it was interesting to learn about that." Other students did write about the difficulty of transcribing or the value of their efforts. Overwhelmingly, however, the response has been positive, indicating to us that this project will attract a large community of transcribers on the Zooniverse platform.

As we outlined above, during this past year we took a more structured approach to testing, first by asking student-testers to complete an online survey after transcribing several documents. Their feedback affirmed that the process of transcribing on the Zooniverse platform was straightforward and enjoyable. One tester commented, "I love the entire transcription process! Sometimes handwriting is a tad hard to read, but it's still really fun to decipher." We also received positive feedback regarding supplemental materials we created to assist transcribers. Testers thought tutorials and the project "Field Guide" were clear, concise, and informative. Wrote one tester, "The field guide is extremely helpful because I wasn't sure about a few words and spent so long trying to figure it out, but then saw that the field guide told you what to do with those and it really helped." While most respondents liked the conciseness of tutorials, we did receive suggestions that helped us to refine our site, specifically regarding better language use and the inclusion of field-guide information into tutorials. Feedback that was specific to the Zooniverse platform was relayed back to Zooniverse's development team. Not surprisingly, some of our student-testers highlighted challenges that are often encountered while transcribing handwritten documents and when working with scanned microfilmed sources. (See Appendix C for student-tester feedback.)

After making some adjustments to our site, we submitted The American Soldier to Zooniverse for formal review. This process began in the second week of March and lasted approximately a week, with 27 testers assessing the site. These individuals were asked to review the difficulty of project tasks (i.e., transcribing the documents), documentation we provided to assist with these tasks, supplementary site information, the suitability of our project for Zooniverse, and the likelihood of continued participation. The responses show a high variability in perceived difficulty of transcribing these documents, similar to what we saw during project testing with Virginia Tech students, with some finding it very difficult to transcribe handwritten documents. Also like our student testers, Zooniverse reviewers did think our help text and additional project documentation to be adequate, informative, and interesting, although some wanted more contextual information—the very sort of information that our future website will provide. This review helped us to improve our Zooniverse site and, where we needed, to fixes a few "bugs."

Suggestions related to the Zooniverse platform were, again, passed on to Zooniverse's very responsive technical team.

While reading another person's handwriting can be very challenging, the task of deciphering can be very rewarding, beyond the achievement of completing a task. Feedback from Zooniverse reviewers echoed what we have received from Virginia Tech students who have transcribed these unusual humanities sources. Namely, they helped members of the Zooniverse community to make personal connection with the war and to family members who are veterans. One reviewer wrote, "My dad was in WWII—the language & idioms are so like him. This is really great source material. So glad it's being preserved!" Another commented, "I loved this project—I went through a few pages and it was fascinating reading the opinions. I am a genealogist and being able to 'hear' the voices from the past is wonderful. I think sharing this with the public would give people a much deeper understanding of the military in WWII. This is information that is not available in books or anywhere else." Of those testers who completed project tasks, all agreed that The American Soldier was suitable to Zooniverse. (See Appendix D for feedback from the Zooniverse review.)

We proposed in our original NEH Foundation grant that we wanted to include an Implementation Plan in the White Paper and that we wanted to circulate the combined document during the grant period for public feedback. We proposed doing this with the expectation that we would adjust our project planning in response to public comment while still benefiting from the grant. We discovered that this additional step would require of us that we complete the White Paper sooner in the planning process, namely, while many details were still being sorted out. Our planning team agreed that we needed to capitalize on the momentum afforded by the Foundation grant by dedicating our efforts to Implementation planning and to the Zooniverse site launch. Though we did not circulate the combined document publicly, the project's Advisory Board offered excellent suggestions on how we could improve our plan in advance of submitting it to the NEH.

Project Continuation

We applied for a Foundation HCRR grant to plan for and to write an Implementation grant proposal, aiming for the next NEH HCRR program application deadline, 19 July 2018. By adding members to our project team, based on the growing complexity of the project; deepening institutional support at Virginia Tech; and forging additional collaborative relationships with external partners—all of which was made possible by the Foundation grant—we were able to achieve our objective.

Our original project team consisted of four members from Virginia Tech, two from the Roper Center, and three external advisory board members. The project's initial co-proposer, Dr. French, left Virginia Tech Libraries in the spring of 2017, though she stayed on with the project as a member of the Advisory Board. Keen to continue its support, Virginia Tech Libraries allocated other resources and staff to the project after her departure, including Corinne Guimont as Project Coordinator for VT Libraries, Christopher Miller as Digital Humanities Specialist, Nathaniel Porter as Data Consultant, and Michael Stamper as Data Visualization Consultant. For reasons outlined above, the Roper Center was replaced by the Social & Decisions Analytics Lab at Virginia Tech's Biocomplexity Institute. Daniel Chen has joined the project as Data Engineer, along with Dr. Gizem Korkmaz as Research Scientist, and Dr. Aaron Schroeder as Senior Research Scientist. We also brought in an additional internal collaborator with expertise in social psychology, Dr. Michael Hughes, and added two content-specialists to the Advisory Board.

Virginia Tech further committed itself by agreeing to store and host project data short- and long-term, within VT Libraries' data repository, VTechData, as well as in servers maintained by the university's Advanced Research Computing (ARC) unit. ARC provides the research community at Virginia Tech with a comprehensive ecosystem consisting of advanced computational systems, large-scale data storage, visualization facilities, software, and consulting services. Finally, VT Web Hosting has agreed to host the project's public website after it has been built. This infrastructure bodes very well for the success of our digital project moving forward.

The HCRR Foundation grant program allowed us to not only expand the team and to secure these necessary technological resources internally, but it also helped to attract interest to the project externally. In addition to granting us permission to add *The American Soldier* volumes to our project website in a digital format, the Social Science Research Council promoted the public launch of our transcription drive on Zooniverse and brought visibility to our project by recirculating publications in their own archives that touch on the work of ARB and by publishing a piece that PI Gitre wrote in their online journal, *Parameters*. Our local NPR station also helped to promote the project's public launch on Zooniverse by airing a short piece on the project, as did other media outlets, including *Stars and Stripes*. Appendix E lists a selection of media attention that went along with our VE Day launch. At present, our transcription portal on Zooniverse has attracted over 1,500 members, who have transcribed (in triplicate) nine of the forty-four digitized ARB microfilm rolls. We will be hosting a second daylong transcribe-a-thon to commemorate VJ Day at the end of August, as well as a third event to commemorate Veterans Day 2018. Team members will continue to promote both the transcription drive and the larger digital project at academic conferences, and this November PI Gitre has been invited to present

and promote the project at the Chancellor's Day Workshop for NYC Teachers, which will take place on the USS Intrepid. Of their own initiative, educators elsewhere have contacted the Project Director to contribute to the project by hosting their own local transcribe-a-thons. We anticipate that within eighteen months of the May 8th launch we will complete the transcription portion of the project.

At the end of the Foundation grant, we were still faced with two planning hurdles. Fortunately, both were resolved prior to the July 2018 submission. The first was the hosting of project data, and the hosting for the project website. Owing to the substantial size of project files, projected to top 750 GBs, we needed to find a cost-effective solution that facilitated long-term data preservation as well as real-time server access by the web application we intend to build for the project. This was no small matter. Fortunately, Virginia Tech already has the computational resources and infrastructure in place to meet both demands. Transcribe documents and their corresponding images will be batch zipped with metadata documentation and uploaded into VTechData for preservation and access purposes. VTechData ensures that published datasets receive persistent digital object identifiers (DOIs) while allowing researchers to assign Creative Commons licenses according to their public data-sharing interests. The web application we built for the project will pull data from ARC. ARC will host data on its DragonsTooth cluster. Designed to support general-batch, high-performance computing, the cluster is a 48-node, high-throughput, two-socket system. The project will be deployed using an on-premises cloud-system configuration.

The second outstanding issue was web development and design. After considering a variety of options, we approached Cast Iron Coding. The firm has extensive experience designing open-source web-based applications and websites for universities, non-profits, and other public-sector institutions, as well as for corporate clients. The firm has a record of success with digital humanities projects, such as the collaborative, Andrew W. Mellon-funded, open-source digital-publishing platform Manifold Scholarship, which Cast Iron Coding developed and designed in a three-way collaboration with the University of Minnesota Press and CUNY GC Digital Scholarship Lab. In our original NEH proposal, we had anticipated using another design firm and incorporating that firm in the Foundation planning process. While testing the procedures necessary to convert ARB's datasets and while designing transcription workflows on Zooniverse, it became quite apparent that we needed to have a better grasp of the underlying wartime records produced by ARB before engaging an outside firm to design a project website. We were also mindful of the cost of hiring a third-party firm for both phases of the project—that is, for both Foundation project planning and for Implementation project development. While anxious to secure a web developer, we can now appreciate that our digital project was well served by

our focus on the project's underlying historical records and the procedures and resources required to harvest this remarkable wartime collection. We were able to approach Cast Iron Coding at the end of the Foundation-grant period with a detailed plan that helped to secure the firm for project implementation. Similar to finding a sustainable solution for short- and long-term data storage, procuring a professional design firm for the project's website was no small matter.

Long-Term Impact

As Dr. Tom Ewing, the associate dean of Virginia Tech's College of Liberal Arts and Human Sciences, stated in his original letter of support for this project, The American Soldier Collaborative Digital Archive is consistent with not only with the College's scholarly, instructional, and service priorities and commitments, but also with the mission of Virginia Tech as a comprehensive, public, and land-grant university. "By designing a project that brings original source material to the attention of scholars at other institutions as well as the general public," he wrote, "this project serves the interests of both the National Endowment for the Humanities and Virginia Tech of putting knowledge in service of the public good. Finally, by building a digital resource in collaboration with university libraries and computer science, this project demonstrates how digital humanities can bridge the gap between conventional forms of analysis and new platforms for public documentation, engagement, and understanding."

The American Soldier is one of several collaborative, externally funded initiatives located at Virginia Tech that are aimed at bridging traditional humanities research and the data sciences, with the express goal of producing scholarship of contemporary relevance on the human dimensions of disease, medicine, mental health, security, and the social condition. Our team's partnership with Virginia Tech's Social & Decision Analytics Lab, for instance, will not only advance the project outlined in our White Paper and Implementation plan, but it promises to contribute to the Lab's ongoing collaboration with the U.S. Army Research Institute for the Behavioral & Social Sciences, which has produced more than $3.2 million in cooperative agreements. Additionally, our partnership with Virginia Tech Libraries aligns with their strategic goals to increase digital scholarship support through curating and making humanities research data openly accessible.

Associate Dean Ewing wrote that he was especially supportive of The American Soldier because it aligns with efforts at Virginia Tech to advance the field of Veterans' Studies. He wrote, "While we are familiar now with the long-term effects of military service as well as tools used by veterans, mental health providers, and educators to provide a framework for making sense of their experience, this project documents how these same issues were central to the ways that the

American military approached the mental health of soldiers during the Second World War." The American Soldier project will contribute to the development of Veterans' Studies as well as to several university-wide initiatives. As our Provost's Office explains, "Destination Areas combine Virginia Tech's existing strengths with novel transdisciplinary teams, tools, and processes that empower students and faculty to tackle the world's most pressing problems." The American Soldier contributes to three of the university's five destination areas: Adaptive Brain and Behavior, Data and Decisions, and Integrated Security.

Based on the public's reaction to the launch of our transcription drive on Zooniverse and on the positive press coverage the launch garnered, the team at Virginia Tech believes that The American Soldier is poised to make a marked contribution to our collective understanding of the American soldiers' experiences of World War II. We believe that the project will benefit not only professional historians and other scholars, but also the general public, given the utterly unique nature of the humanities collection produced by ARB. We look forward to seeing that transcription drive through to its completion and to the continuation of the project through its implementation.

Grant Products

- NARA-coordinated scanning of 65,631 microfilmed documents;
- Sample conversions of social survey data files to produce the necessary protocols to convert the complete data set, comprised of 183 data files, with nearly 300,000 respondents;
- A Zooniverse site to transcribe 65,631 handwritten "free comment" responses from WW2 soldiers (The American Soldier: A World War II Survey);
- Inventory of NARA archival records related to the Army Research Branch and The American Soldier;
- Inventory of survey topics, locations of surveys, units surveyed, etc.;
- Work plan for data re-unification;
- Interaction design and system requirements for digital archive;
- Plan for data management and sustainability;
- Plan for curriculum development;
- Cooperative agreements with National Archives and Social and Decision Analytics Lab;
- Letters of Support from internal and external stakeholders, including the Advisory Board;
- Public launch of Zooniverse site with associated transcribe-a-thon;
- White Paper and Implementation Plan;

- Edward Gitre, "The American Soldier: A World War II Survey," *Parameters*, 8 May 2018, http://parameters.ssrc.org/2018/05/the-american-soldier-a-world-war-ii-survey/

Appendix A: Project Accomplishments

- Project Director worked with NARA on scanning the complete collection of free-response documents, contained on 44 microfilm reels. PI Gitre engaged with NARA on quality control, leading to the rescanning of a subset of rolls. Digitized images will be uploaded to the National Archives Catalog. Copies were also provided to our team. Several copies were created after transfer to Virginia Tech for preservation purposes and for our Zooniverse project. NARA's decision to scan the documents and upload them into the National Archives, at their own expense, ensures wider distribution of these records and will increase the visibility for our project. Completed transcriptions will be ported back into the Catalog.

- Project Director hired a Graduate Research Assistant (GRA). We proposed hiring a History GRA but instead secured a Computer Science GRA. While the GRA did not create a project bibliography as proposed, having a computer scientist on the team allowed us to accomplish far more than originally proposed, including the creation of a crowdsourcing transcription project on the Zooniverse platform, with its million-strong community base. The GRA did perform other functions outlined in our proposal, such as managing team meetings and project deployment in Virginia Tech courses. Additional archival research was undertaken by PI Gitre in May 2017 at NARA, in College Park, MD, revealing the existence of archival records not delineated by NARA in the finding aid for The American Soldier (NARA Reference Information Paper 78).

- Technical Lead set up a customized Incite instance for American Soldier for deployment in VT courses. PI Gitre, co-PI Luther, and GRA Nai-Ching Wang coordinated testing of Incite, evaluated results, and made the decision to use the crowdsource Zooniverse platform for transcription of "free comment" documents. As a result of this decision, the Project Coordinator and Technical Lead did not need to determine server space requirements (storage, bandwidth, configuration) for Incite and backup processes for the digitized images, as originally proposed.

- Project Director and Technical Team created a Zooniverse project site, with accompanying FAQs, field guides, tutorials, additional project-related pages, and workflows for datasets. The Zooniverse site was submitted for review to become a Featured Project. The Technical Team made adjustments to the site based on the feedback submitted during this review positive. Of all testers who were able to complete project tasks, all evaluated our project as appropriate to Zooniverse. These accomplishments exceeded what was included in our original proposal.

- Project Director worked with Roper Center to conduct an analysis of survey data and to estimate the necessary resources for the conversion of all datasets from ASCII to Stata format, by conducting an initial test conversion of two survey files. This step was not detailed in our origin proposal. As outlined in our project's White Paper, midway through the year we decided to collaborate instead with Virginia Tech's Social and Decision Analytics Laboratory (SDAL). SDAL conducted an analysis of survey data similar to Roper and performed test conversions of select data files, to ascertain the resources necessary for complete dataset conversion. This was completed.

- We originally proposed creating a working plan with Roper Center technical staff for an enhanced prototype website, one that would integrate Incite/Omeka transcription data as well as contain a web-based data tool similar to what Roper already provides its members, called iPOLL. We will no longer work on the development of an iPOLL web-based tool with Roper, but use other tools for data integration, as outlined in our implementation plan.

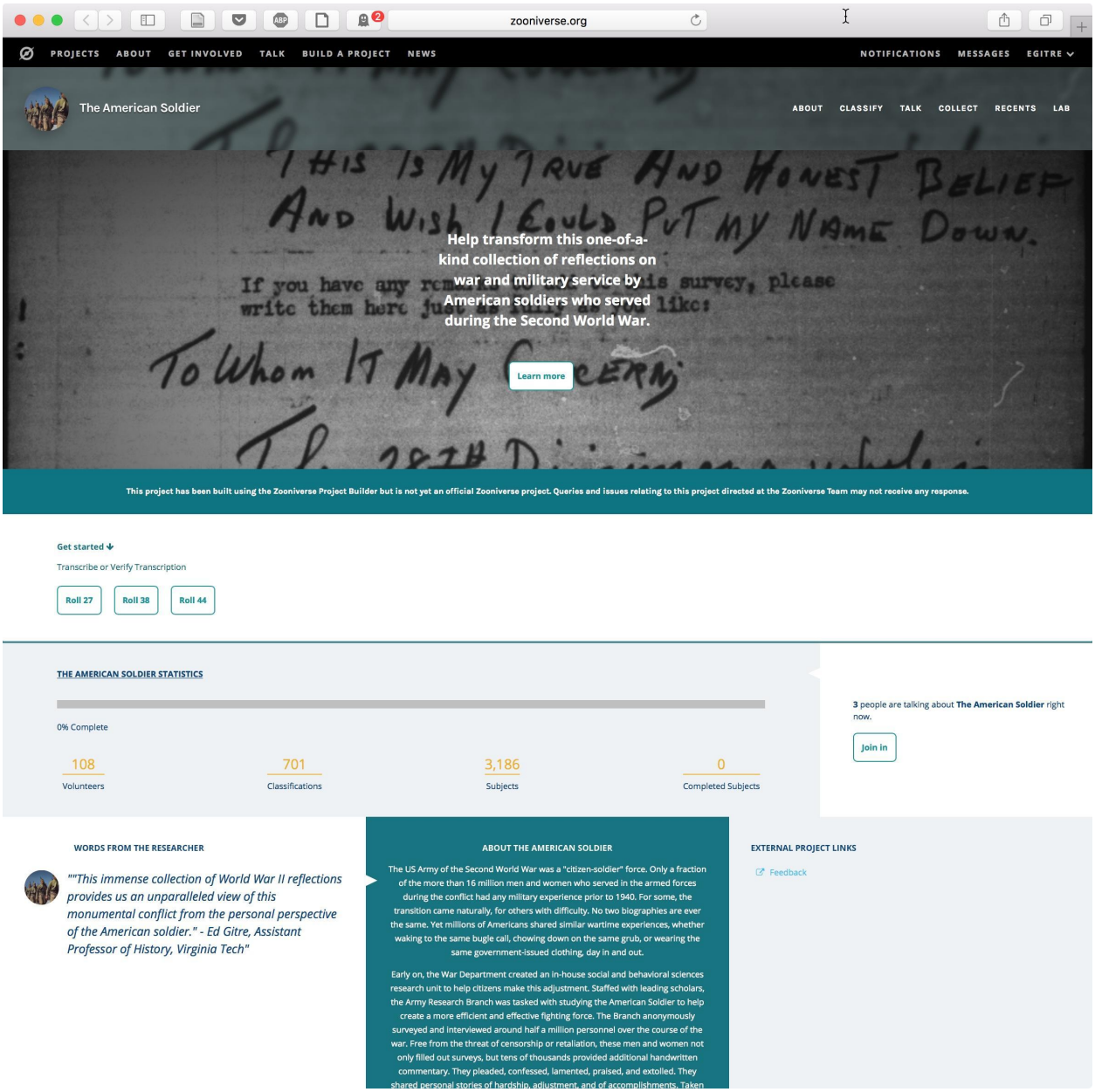
- Project Director and Coordinator had intended to hold virtual meetings with Interactive Mechanics to initiate the design of an Omeka-based website prototype with Incite integration and would do so while Roper created a parallel work plan for iPoll development. As we began to analyze our sources, we realized that we needed to have a much better grasp of the content of our project before initiating a prototype design with an outside web developer. With this in mind, we focused on the challenge of reintegrating ARB survey data with transcribed "free comment" responses and on finding a crowdsourcing platform to ensure that all 65,000-plus digitized images are transcribed for inclusion in the website that will be developed during project implementation. We did this while negotiating our collaboration with Roper (outlined elsewhere). We did not hold meetings with Interactive Mechanics while negotiating this relationship with Roper. Also, Virginia Tech Libraries, in its efforts to expand support for digital projects, committed itself to securing and approving web designers to assist with library-supported projects, including The American Soldier. That process is ongoing. Our technical team did, however, develop for the project website target audience "personas," interaction design requirements, and system requirements, which will guide project implementation. After the conclusion of the grant, we secured Cast Iron Coding as our web developer and designer.

- Project Director submitted an Implementation plan to the Advisory Board for comment and revision in advance of our final stakeholders meeting scheduled in April 2018. In that final meeting, stakeholders helped to identify the strengths and weaknesses of the Implementation proposal and made suggestions on how to make improvements. The Project Director simultaneously drafted the grant White Paper, working with other project

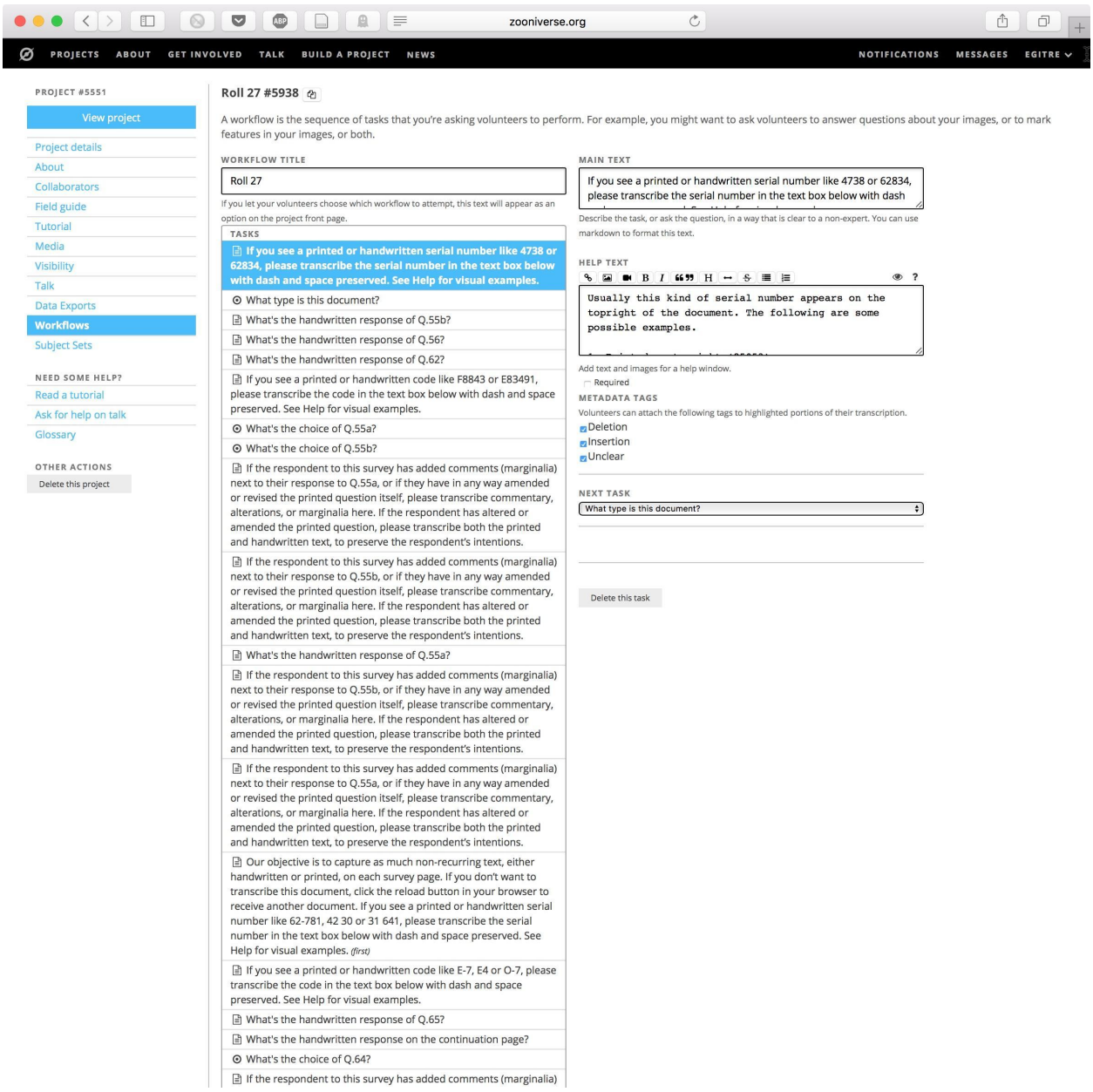
stakeholders, chiefly the Paper's co-author, Dr. Luther. The White Paper details work undertaken during our Foundations-level grant. It addresses the scope, scale, and outreach necessary for crowdsourcing the project; technological and interaction-design requirements for project hosting and the web interface; access to I&ED/ARB sources that have or need to be digitized and/or hosted; IP and copyright issues related to these sources; and other pertinent elements of design and project planning.

- Project Director obtain commitments from the Advisory Board and other project partners about material and work commitments necessary in the coming years to realize the final vision of a sustainable portal for World War II materials related to the surveys and Army Research Branch documents. With the White Paper in hand and a draft of the Implementation proposal completed, Project Director secured two additional members for the project's Advisory Board, Drs. Beth Bailey and Tom Guglielmo—the one a leading historian of war and society, the second a rising historian of war, society, and the civil rights movement.

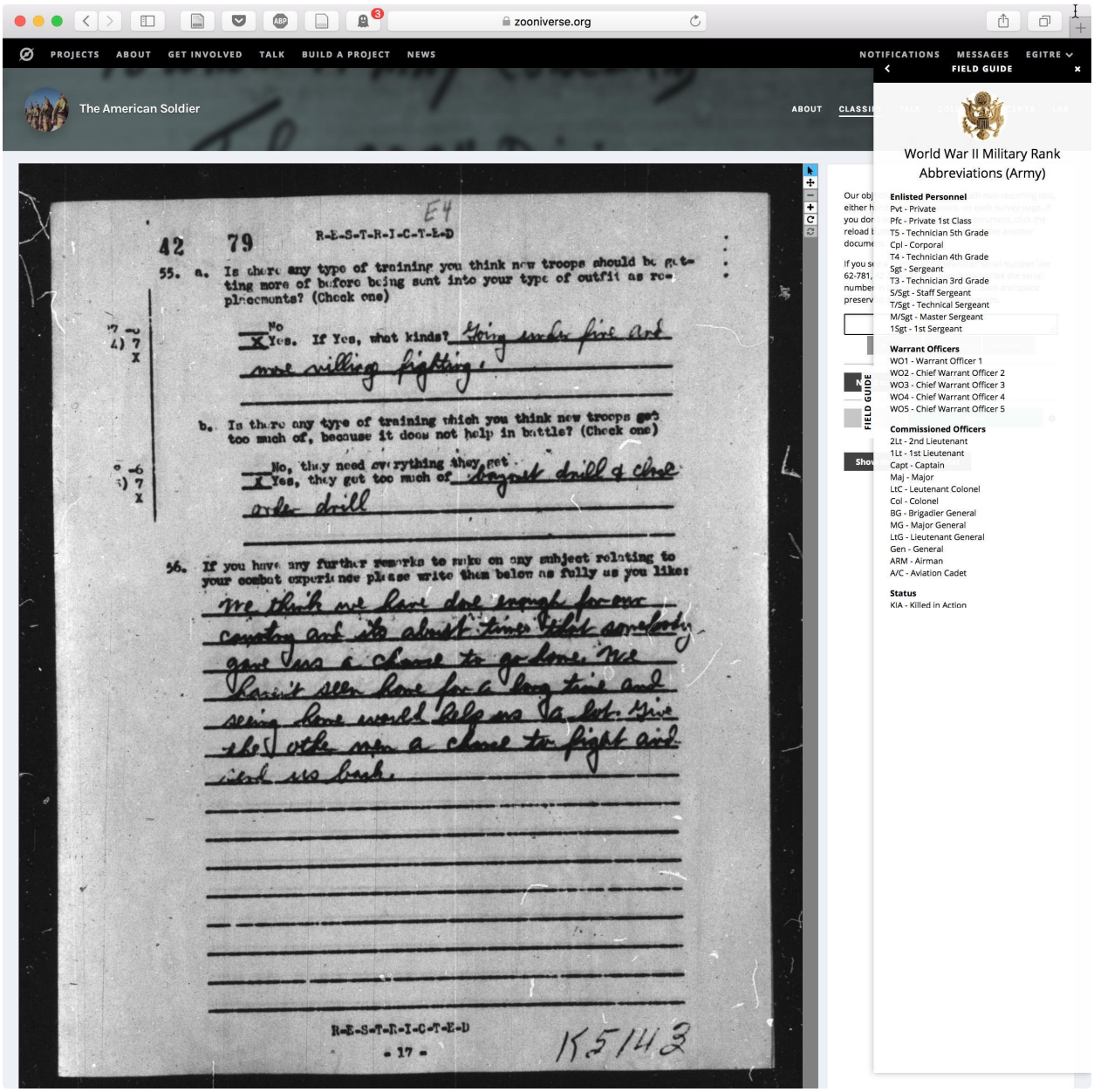
Appendix B: Zooniverse Site



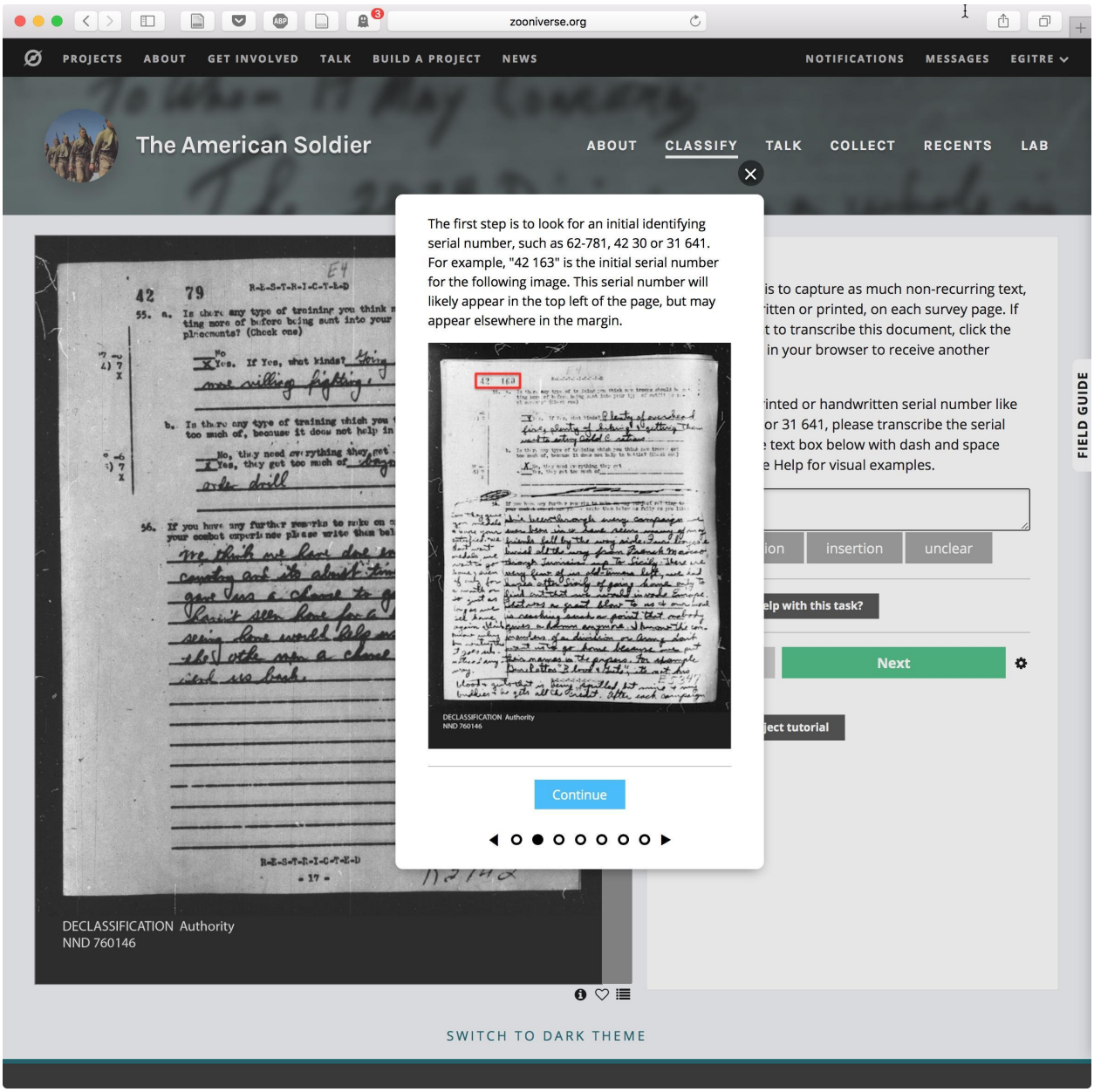
front page of portal



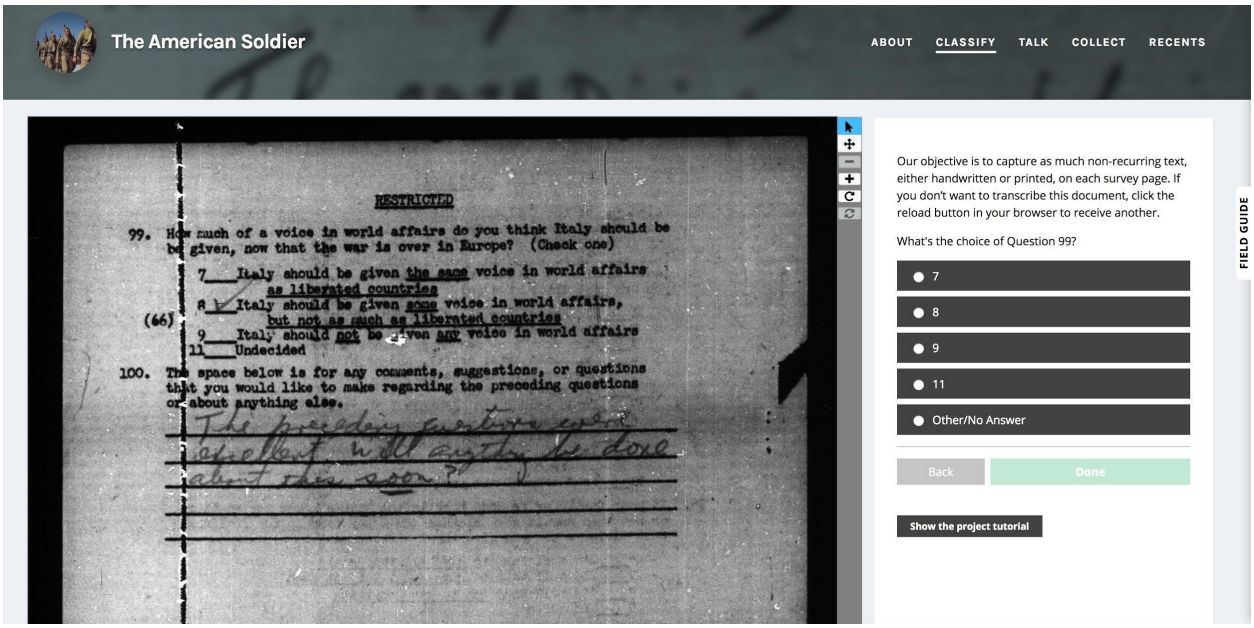
sample transcription workflow



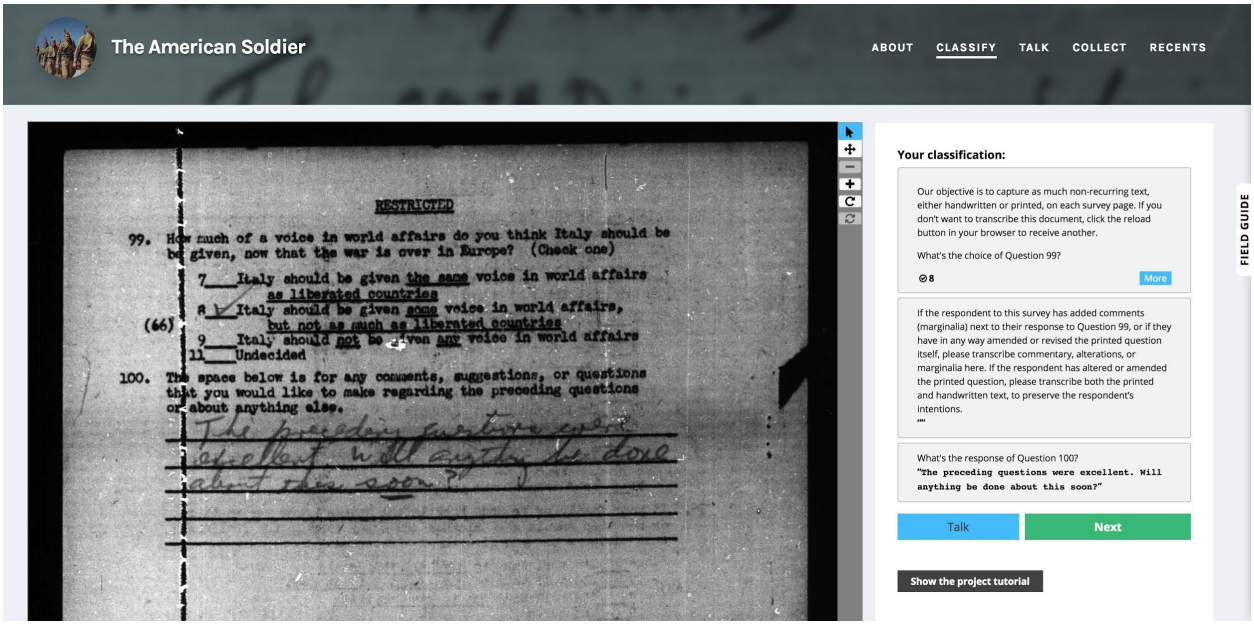
sample field guide


zooniverse.org
PROJECTS
ABOUT
GET INVOLVED
TALK
BUILD A PROJECT
NEWS
NOTIFICATIONS
MESSAGES
EGITRE
The American Soldier
ABOUT
CLASSIFY
TALK
COLLECT
RECENTS
LAB
The first step is to look for an initial identifying serial number, such as 62-781, 42 30 or 31 641. For example, "42 163" is the initial serial number for the following image. This serial number will likely appear in the top left of the page, but may appear elsewhere in the margin.
DECLASSIFICATION Authority
NND 760146
Continue
FIELD GUIDE
insertion
unclear
Next
SWITCH TO DARK THEME

tutorial



workflow task



transcription verification

Appendix C: Feedback from Student-Testers & Sample Feedback

Feedback on Process

Their feedback showed that the task process was straightforward, enjoyable, and easy to follow. Here is a sample of the responses we received:

> *"The process is very enjoyable, it feels rewarding and nothing gives any trouble."*
> *"The process is neat. The instructions are pretty straight forward and I feel that anyone can pick this up and work on it."*
> *"I love the entire transcription process! Sometimes handwriting is a tad hard to read, but it's still really fun to decipher."*

Feedback on Tutorial

The feedback also showed that the design of the tutorial was clear, concise, and helpful. When asked about the tutorial, typical responses include:

> "*The tutorial was very clear and easy to follow.*"
> "*Very informative and leaves nothing to question.*"
> "*The tutorial and description is very clear and concise. It is simple and straightforward which makes these tasks simpler.*"

Feedback on Field Guide

The field guide was considered concise and very helpful, as some responses highlight:

> "*The field guide is extremely helpful because I wasn't sure about a few words and spent so long trying to figure it out, but then saw that the field guide told you what to do with those and it really helped.*"
> "*Field guide was very helpful in knowing how exactly how to transcribe the documents*"
> "*The field guide was helpful if you needed help. The description was clear, people who try to transcribe this should have no problem figuring out what to do.*"

A sample of responses is as follows:

"The tutorial is limited in information which is fine since it directs you to the field guide which was much more helpful."
"the field guide is much better and should be mentioned in the tutorial."
"I did no know there was a field guide so making it more prominent could help with identifying that."

Appendix D: Summary of Feedback from Zooniverse Review

Task Difficulty: The responses show a high variability in perceived difficulty of the tasks, ranging from very easy to very hard. This may show the readability of handwritings of different soldiers is very different and/or there is a big individual difference in reading handwritings.

Help Text: The responses agree the help text was indeed helpful and adequate to help complete the tasks although some people were not able to open the tasks.

Additional Information/Capability: When asked about what additional information or capability may be helpful or interesting, the responses generally think the current amount of information is enough. Some suggest more context about documents and some more user interface for the image such as brightening and larger display of the image.

Suitability for Zooniverse: Except for those who were not able to open the project (about 3 participants), the rest participants (about 25) unanimously agree the project is suitable on Zooniverse.

Participation: Except for those who were not able to open the project, about half of the responses were positive about future participation when the project becomes official. Two of them even mentioned about bringing their friends. Only five of the participants specifically say no for future participation.

Other comments: The responses also mention the project is interesting and even relate to their family or themselves as WWII veterans. The responses also recognize the importance and potential values once the project is completed. However, they also mention the tasks (handwritings) are not easy and bring up some bugs and suggestions to improve the process.

Appendix E: Press Related to the Public Launch of Zooniverse Site

"Victory in Europe Day to be commemorated with national effort to make available insights of American soldiers during World War II," *VT News*, 1 May 2018, https://vtnews.vt.edu/articles/2018/05/clahs-american-soldier-project.html.

> Press Release Recirculated
> CBS Fox 59, WVNSTV, 3 May 2018, https://www.wvnstv.com/news/education/victory-in-europe-day-to-be-commemorated-with-national-effort-to-make-available-insights-of-american/1157524658; *Military News*, 9 May 2018, https://www.militarynews.com/become-a-part-of-history-with-crowdsourced-project-the-american/article_df41427e-52c3-11e8-8cc4-5778c321caf9.html; *War History Online*, 6 May 2018, https://www.warhistoryonline.com/instant-articles/victory-europe-day-commemorated.html

Robbie Harris, "What They Were Thinking: WWII Soldiers' Insights to Become Public Data Base," *Radio IQ WVTF*, 8 May 2018, http://wvtf.org/post/what-they-were-thinking-wwii-soldiers-insights-become-public-data-base

Rodrigo Ugarte, "Understanding the American Soldier: The SSRC and Social Science in World War II," *Items*, 8 May 2018, https://items.ssrc.org/understanding-the-american-soldier-the-ssrc-and-social-science-in-world-war-ii/?_cldee=YmV5ZXJjaGVuLjFAZ21haWwuY29t&recipientid=contact-be41f918f646e311a360001cc477ec84-ca1700220de84c3faef9b476e7ba182d&esid=dd203560-ff51-e811-80cb-005056ab0bd9

National Archives Catalog email, https://mailchi.mp/nara/0rjknzxchj-763325?e=ce163cc854

Will Morris, "WWII in a new light: Anonymous soldier surveys tell stories of morale, race relations," *Stars and Stripes*, 7 May 2018, https://www.stripes.com/news/wwii-in-a-new-light-anonymous-soldier-surveys-tell-stories-of-morale-race-relations-1.525772