Grant Narrative

1. Significance

The Northeastern University Libraries seek funding for the first major development phase of the Early Caribbean Digital Archive. This collaborative digital research collection is an interdisciplinary project that brings together scholars of early Caribbean textual and visual studies, network analysis, network visualization, and archival digitization. The project seeks to create a digital archive of culturally and historically significant materials from the 16th- to 18th-century Caribbean region, and to develop a scalable, repository-based publication infrastructure through which the collection can be managed, published, annotated, and studied.

This project will support and build on existing research on the Caribbean region in a way that is both timely and critically necessary. The history of the early Caribbean has been marginalized by its relegation to the colonial archives of nations in Europe as well as in the United States. However, the Caribbean plantation economy, fuelled largely by the exploitation of slave labor, was at the heart of the new Atlantic economy that served as the foundation of the modern world system in the seventeenth and eighteenth centuries. Historians and critics have thus increasingly turned to the early Caribbean as the site of the origin of modernity, particularly in relation to new systems of labor and racialization that have had enduring impact. At the same time, scholars in the digital humanities are turning their attention to the challenge of how to study regionality and — even more importantly — how to make categories of race, gender, class, ethnicity, and regional identity visible in the digital collections that underpin digitally mediated historical and literary research.

Physical archival collections are products of the history that creates them, and digital collections often mirror these configurations. As a result, although there is a wealth of digitized text available, it is not currently aggregated or organized in a way that makes research on the early history of the Caribbean archipelago possible; no pan-Caribbean digital or analogue archive of pre-twentieth-century materials now exists. Important existing collections include the digitized archives of dLOC (Digital Library of the Caribbean), the Early Americas Digital Archive http://mith.um.edu//eada/, at University of Maryland and Rice University, as well as a variety of library-based and commercial archives that include partial collections of early Caribbean materials (prominent among these are the John Carter Brown Library, the American Antiquarian Society, the British Library’s Burney Collection, the Adam Matthew Eighteenth-Century Journals Portal, and the digital collections of Readex/Archive Americana). None of these collections is devoted exclusively to early Caribbean texts, in part because of the imperial and colonial history of the Caribbean region that has caused print materials concerning the Caribbean to be fragmented and divided among archives in Europe, North America, and the Caribbean. It is quite difficult, for instance, to find full runs of eighteenth-century Caribbean newspapers (which were numerous) collected in a single archive: rather, many libraries hold scattered issues, but these issues are not brought together in any single location. In addition, limited financial resources make it difficult for national Caribbean archives (such as the National Library of Jamaica or the National Library of Haiti) to collect and digitize historical publications from their countries. Perhaps most importantly, the geographical dispersion of both materials and scholarly community poses obstacles to the kinds of dialogue and
collaboration that we understand as necessary to avoid simply recapitulating colonialist structures in new academic forms.

The inaccessibility of Caribbean archival materials, together with the buried nature of the history of enslaved peoples within an archive of texts written by colonizers, has thus rendered much of this history difficult to access. There is, then, an urgent need for a digital archive that brings together materials from the early Caribbean in a way that supports focused, collaborative research and can also ensure the long-term curation of the digitized materials and related scholarly products. The Early Caribbean Digital Archive aims to fulfill this role and to develop a new model for accessing and understanding historical print and visual archives. The Early Caribbean Digital Archive will use new digital tools to aggregate, reassemble, analyze, and bring into visibility this hidden history.

Our first goal in this effort is to uncover and make accessible a literary history of the Caribbean written or related by black, enslaved, creole, and colonized people, and also the surrounding economic and cultural history of the emerging commodity markets through which colonialism and systems of slavery were supported and elaborated. Both the contents of the archive, the digital representation of the documents, and the ECDA research interface will focus on foregrounding the buried and embedded slave narratives and patterns of commodity circulation that inhabit these texts, creating in effect a new scholarly canon of early diasporic African and creole voices. In this way, we are not only remediating original texts, but reframing the literary history of the early Caribbean as one where something new emerges—voices beyond the imperial history of the Caribbean.

Our second goal is to create an innovative interface for analyzing and accessing the texts. This interface will serve both as a commons that brings scholars together through discussion, annotation, and collaboration, and also as a laboratory that supports detailed research on texts and their history. We see three core functions of the interface. The first of these is to organize a scholarly community around the collective and on-going creation of the data in the archive, including activities of transcription and contribution. The second is to facilitate the formation and cohesion of a user community through tools that enable textual annotation and discussion. The last is to provide tools for analysis and visualization of the collection and subsets thereof. The interface will allow users of the database not simply to access texts by way of keyword searches, but to discover relations and mappings between texts as the materials are viewed. Such mappings will bring into visibility relations between locations of print activity across the Caribbean archipelago, as well as relations among individuals, such as printers, consumers, merchants, runaway slaves, missionaries, plantation owners, abolitionists, military figures, and colonial political figures. Such mappings can in turn help illuminate the cultural and political texture of the region: for instance, the relation between parliamentary legislation and the price of commodities as advertised in newspapers across a series of islands; the frequency of runaway slave advertisements in relation to changing laws concerning slave punishment or slave marriage; the geographic circulation of literary tropes; or concern over the effects of the Haitian Revolution in adjacent islands. In short, a wealth of interlinked textual and contextual information would enable a new understanding of a host of areas—literary, economic, political, and cultural—made visible in relations between people, places, and texts.

The ECDA has already made progress towards these goals. First, the project has formed robust ties to scholarly communities in early Caribbean studies. The project was initially conceived in
relation to the formation of the Early Caribbean Society, a group of scholars from around the world who work on early Caribbean texts. Regular engagement between the ECDA and scholars in this group will assure that the archive is being constructed in ways that scholars in this field will find valuable for both teaching and research. The ECDA has also established a formal partnership with the Digital Library of the Caribbean (dLOC) to assist in preserving digital versions of important source documents and ensuring open access to them. As part of this partnership, dLOC will serve as the point of contact with regional archives, gathering and aggregating digitized versions of source materials under an open-access license. These materials will then be made available to the ECDA for transcription, TEI encoding, and inclusion in our collection. This partnership enables us to focus our efforts on adding value to dLOC materials: whereas dLOC aims at acquisition and preservation, the ECDA is focused on the digital use and analysis of texts. As a result, our partnership with dLOC gives us additional reach by permitting us to acquire texts from archives around the world, and assists us in building visibility in the scholarly community. Further, a new journal associated with the Early Caribbean Society (also housed at Northeastern, and edited by one of the ECDA co-directors, Professor Nicole Aljoe) will foster links between scholarship in the field and materials housed in the ECDA. The project has developed an initial prototype site, using the open-source Omeka exhibit platform, through which we have begun experimenting with publication and transcription. Finally, the project has developed a draft specification for encoding ECDA texts using the TEI Guidelines, widely recognized as the key standard for the representation of humanities research materials in digital form.

We now seek funding for the next development phase of the ECDA, in which we take the project beyond what is possible through this prototype and establish a robust, sustainable platform for archiving and publication of digital materials as well as a scalable interface for exposing and interacting with them. Under this grant, we will accomplish three central tasks:

1. We will establish the essential infrastructure of the archive. The archive will be built using Fedora and Hydra, an open-source repository and interface framework that is widely accepted as an industry standard for institutional digital repositories. Materials will be housed in the Northeastern University Digital Repository Service under a memorandum of understanding that guarantees the long-term security and viability of the data.

2. We will digitize, transcribe, and encode an initial set of 100 texts to serve as the core of the archive, selected with the assistance of early Caribbeanist scholars to represent a wide range of genres and subject matter as well as key texts of current research interest. (A full listing is included in the Appendix.) This test corpus will be at the heart of the work we aim to accomplish during the grant period and will serve as a reference point for future data capture; its comparatively small size and wide generic range will permit us to undertake more detailed document analysis and finalize our encoding specification and schema in relation to the archive’s interface functions before undertaking a more extensive digitization effort. As part of this initial transcription effort we will also develop protocols for ongoing transcription and encoding through community contribution and crowd-sourcing.

3. We will develop an interface that not only supports traditional digital research activities such as browsing, searching, and reading, but also provides a collaborative research environment through which scholars can create personal research collections and exhibits, annotate texts, publish commentary, assist in transcription, and contribute texts to the collection. Because this
interface will be built using Hydra, which draws on a wide developer community and emphasizes reuse of code modules, the features we develop for ECDA will be repurposeable for other digital projects at Northeastern, and conversely can take advantage of features being developed for those projects, including visualization tools for network analysis and complex text and data mining.

By the conclusion of this development phase, we will have established a fully functioning repository and publication infrastructure for the ECDA, populated with a strong initial collection of texts and supported by an engaged research community, and we will also have established the accompanying methods, documentation, and work flow to carry the project forward after the completion of the grant.

2. History, Scope, Duration

The work to be accomplished within the scope of this grant represents a portion of a larger project that has already made significant progress and will continue after the conclusion of this grant. As noted above, the ECDA first emerged out of the formation of the Early Caribbean Society. The ECDA has been supported with seed funding from Northeastern University in 2012-2013, and again in 2013-2014. We are now beginning a third year of work on the project with support from a Tier One grant from Northeastern University and fellowship funding for the project’s director, Elizabeth Maddock Dillon. Work accomplished thus far has involved research and organization, most particularly with respect to how to structure the data in the archive and how to transcribe and encode the texts in TEI/XML. In addition, we have identified and resolved issues of access, acquisition, and copyright. We have also created a public, teaching-oriented portion of the archive (using Omeka) that is being used for crowd-sourced transcription of texts and for teaching purposes on a test basis by a number of scholars at different institutions as of spring of 2014: http://omekasites.northeastern.edu/ECDA/. Our initial development work on the project can thus be divided into two main components:

1. Interface Prototyping in Omeka: On the Omeka site we have posted a small but growing collection of images and transcribed texts for users to browse and examine. We are also using Omeka’s Scripto plugin to support user transcription of early Caribbean texts (posted as PDF images on the site). The Omeka site thus enables us to learn in a preliminary way how users want to interact with the collection and to explore the process of crowd-sourced transcription. We expect crowd-sourcing to be undertaken primarily by other scholars and their students—particularly other scholars who are members of the Early Caribbean Society. The Omeka site also offers valuable pedagogical possibilities: students can annotate and “curate” sub-collections of texts on the site, can use Google maps to generate maps of texts, and can use the Omeka Neatline plugin (which supports more advanced mapping and timelines) to create exhibits of texts and images. By making materials available for use by scholars (for research and teaching) at this early stage, we are able to glean information about how the archive will be used and how it can best be structured, in its final form, to foster community and research.

2. Document Analysis and Schema Development in TEI: After initial transcription and display on the Omeka site, all texts will be encoded using the TEI (Textual Encoding Initiative), the current scholarly standard for literary and textual encoding. In the
summer months of 2014, a team of graduate students is working with Elizabeth Maddock Dillon, funded by an ACLS Digital Innovations Grant, to design the markup schema that will be used in the encoding. The encoding is substantial and detailed, in view of our aims to use it for textual analysis, and considerable work has gone into designing a schema that will be responsive to research questions of interest to scholars using the collection. In the fall of 2014, the graduate students will begin supervising undergraduate work-study students who will transcribe (where needed) and encode the core collection of texts that will form the first phase of the archive. Completion of the ACLS-funded work will position us to begin work on the next phase of the project—namely, developing the practical interface framework for using the TEI encoded texts, for which we seek support with this grant application.

During the project phase to be completed under the current grant proposal, as described above, we will establish the essential infrastructure of the archive in Fedora and Hydra; we will digitize, transcribe, and encode an initial set of texts and finalize the TEI schema and documentation that support the transcription effort; and we will develop an interface through which the ECDA community can study, teach with, and contribute to the ECDA.

The current grant proposal complements several other strands of effort for which we have sought separate funding. The ECDA forms part of a larger digital humanities project ecology at Northeastern University and has formed collaborative partnerships around several areas of common research and development interest. The most relevant of these emerges from the growing community of TEI projects supported by the Northeastern University Libraries, which now includes both the Women Writers Project (http://www.wwp.northeastern.edu) and TAPAS (the TEI Archiving, Publishing, and Access service, http://www.tapasproject.org), both recently relocated from Brown University. In partnership with these two projects, the ECDA has recently applied for NEH funding to develop a TEI-aware version of Bookworm (http://bookworm.culturomics.org), a data-mining and text visualization tool developed at the Harvard Cultural Observatory. This project would take advantage of the same repository storage and publication architecture that is being planned for the ECDA and would radically advance scholars’ ability to discover and examine textual and historical patterns within TEI-encoded digital collections. Future grant proposals are also being planned to develop additional interface tools that would build on the Fedora/Hydra architecture and would be usable by all of Northeastern’s TEI projects. By using this same architecture for its own publication, the ECDA would be able to take advantage of these developments as well.

Because of the very large quantities of relevant materials—already known, and still to be discovered—we do not anticipate that the ECDA will ever be “completed” in a conventional sense. The project is thus being framed with a very long-term mission and horizon of activity. However, with the essential infrastructure in place (and given Northeastern’s commitment to support that infrastructure as part of its support for the Digital Repository Service and the data contained within it), the project’s ongoing costs are comparatively minimal. Development effort to support additional major features will be funded through grants, probably in partnership with other projects at Northeastern. The project’s plans for sustainability are discussed in more detail in the Sustainability section below.
3. Methodology and Standards

Data Organization and Metadata

Metadata and data organization are critical issues for any large-scale collection, but for the ECDA they are of particular interest because of the nature of the collection. Unlike collections organized around genre, author, or language, the ECDA collection is regionally focused and deals with a domain in which many of the organizational categories that seem natural and significant to us (e.g., race, geographical boundaries, nationality, genre) are somewhat fluid, contested, or subject to historical or regional variation. For the ECDA to work as a research environment, we need to use these categories and make them visible to researchers, but in a way that takes account of these kinds of variability. Even more importantly, these terms need to be developed in a way that engages the community of researchers, archives, and other constituencies in the Caribbean region, and is sensitive to issues of cultural difference and colonial history.

In addition, for most readers this collection is significant not because of individual well-known items but rather because of the way it can represent the textual culture of a region. As a result, metadata serves to enable the discovery of groupings of items and associations between items that have been invisible to researchers. The metadata serves as a proxy for the individual item when researchers need to analyze the entire collection for patterns. Metadata descriptors for categories like racial identity, genre, geography, gender, and similar culturally defined categories thus carry a special weight even though they are also inherently difficult to systematize.

During the prototyping and document analysis process, the project has paid careful attention to the development of descriptors for these categories. These will be used in the formal document metadata: the TEI headers for transcribed TEI files, and the Fedora object records for other kinds of digital objects. They will also be used in the metadata about named entities, specifically persons and bibliographic items. During the development process, the project will engage its research community in a review of these terms as we finalize them for use in the ECDA interface and in final versions of the digitized documents.

Repository Systems

The version of the ECDA planned under this grant is designed to work closely with digital repository systems and to take advantage of their strengths as platforms for both data curation and data exposure. The Northeastern University Library’s Digital Repository Service (DRS) is built using Fedora (http://www.fedora-commons.org) and Hydra (http://projecthydra.org), which are open-source repository tools. Fedora provides the data storage and management component and Hydra provides a modular framework for developing specific interface components such as searching, metadata extraction, data ingestion workflow, data visualization, and so forth. Hydra is designed so as to permit its development community to share and build upon each other’s components, maximizing the possibility of component reuse. Individual Hydra components are conceptualized as distinct Hydra “gems”: modules of code that use a common specification and can be modified, extended, and reapplied in new contexts. These gems are in turn aggregated into Hydra “heads”, which are essentially assemblages of interface components that together constitute a distinct application aimed at supporting a specific kind of user experience. The Northeastern library has already developed Hydra head
for its Digital Repository Service, which includes a set of Hydra gems that handle the essential interface components of the DRS (such as browsing, searching, and data ingestion). The ECDA interface will constitute another Hydra head with quite a different look and feel, but several of the gems developed for the DRS will be useful for the ECDA as well. Similarly, we anticipate that the gems developed for the ECDA will be repurposable for other similar projects that use the DRS as their foundation. Crucial new gems to be developed under this grant include components to handle annotation, user accounts, creation of exhibits, and integration of social media into the ECDA commons.

The development process for the Hydra components will proceed in two stages. First, we will develop a prototype in Hydra that establishes the functionality of the ECDA site and enables us to do initial usability testing and interface design. During this stage we address basic architectural questions and issues of overall workflow, but we do not yet firm up any specific components as Hydra gems. Second, we will identify specific components that should be finalized as Hydra gems; these are components that lend themselves to modularization and reuse. This two-stage approach provides greater flexibility in the prototyping process.

Data Formats

Digital images (of texts, maps, other items) will be primarily sourced from the Digital Library of the Caribbean (dLOC), which will manage the acquisition and management of these materials. The ECDA will receive copies of these materials from dLOC for local storage in Northeastern’s Digital Repository Service (DRS). Where possible, these materials will be stored as high-resolution archival TIFF images in the DRS and converted to web-deliverable formats upon ingestion. In some cases local digitization practices at source libraries may dictate some other format, typically PDF or JPEG.

Transcriptions of primary source documents will be represented in TEI/XML. During the prototyping process, the project has developed a preliminary TEI schema based on an initial document analysis, and this schema will be tested and refined during the period July 2014 through March 2015 with support from Elizabeth Dillon’s ACLS Digital Innovations grant. This TEI customization represents an encoding that emphasizes detailed capture of the text of the source document, including structure and genre information, as well as identification of named entities (persons, places, organizations) and also of a set of specific content features that are of analytical importance to the project (references to commodities, flora and fauna, foodstuffs, enslaved persons).

TEI headers from the transcribed data will be indexed and searchable via the underlying Hydra infrastructure, and will be transformed into MODS records at ingest so that ECDA materials may be located from a generic repository search.

Community annotations will be expressed following the recommendations of the Open Annotation Core Data Model (http://www.w3.org/community/openannotation/), and stored in the DRS as first class objects. The project will use the Open Annotation plugin for Fedora developed at Brown University (http://brown-university-library.github.io/oac_web_service/). User profiles will not be stored in Fedora but will be stored in the Hydra head itself as a self-contained database.
Text Acquisition and Transcription

For the phase of development covered by this proposal, the ECDA will be working on an initial set of approximately 100 texts identified by scholars from the Early Caribbean Society as a particularly appropriate starting point for the collection. These texts cover a wide range of genres and formats including novels, poems, travelogues, scientific and agricultural texts, and newspaper articles and are closely implicated in current research on the early Caribbean; they will serve as an intellectual catalyst and a compelling start to the ECDA’s growing research environment. This list will be shared with dLOC, which will undertake to get digital images of the source texts from appropriate libraries for inclusion in dLOC, whence they can then be ingested into the Northeastern digital repository for use in ECDA. The initial set of texts are all in English; in subsequent phases of the project, we expect to include texts in other languages as well.

The TEI transcription work will be managed by the project’s Research Assistants, who are also responsible for maintaining the ECDA schema and documentation (with oversight and assistance from Digital Scholarship Group staff). The transcription will be done in three stages. The initial text-level transcription, basic TEI encoding, and proofreading will be accomplished by work-study students, who are available to the ECDA at no cost (their labor is covered from University sources). These students will be trained and overseen by the Research Assistants. The RAs will also be responsible for the second, more detailed encoding pass through the text (which will also constitute a review of the first encoding pass), in which the content-level encoding will be added. This encoding will include identification of named entities (persons, places, organizations), development of the project’s personography and gazetteer, representation of embedded slave narratives, encoding of references to commodities, discourse features, and more detailed representation of specific genre features. At this stage, textual features that are unfamiliar or require further scrutiny will be flagged for the third pass, which will be done after discussion as needed with members of the project team, DSG staff, and members of the project’s research community. Our current time estimates for the first two encoding and proofreading passes are as follows:

- first pass transcription, average 10-15 pages per hour, with an average text length of approximately 100 pages, for an average total of 6-10 hours per text (1000 hours total)
- first pass proofreading and correction, average 10 pages per hour, for an average total of 10 hours per text (1000 hours total)
- second pass encoding and review, average 6 hours per text (600 hours total)

These numbers are based on information from the Women Writers Project, whose texts and encoding practices are of comparable complexity and whose encoding staff receive comparable training. Because the level of encoding for the ECDA is still being determined and because the transcriptional difficulties of specific texts are somewhat unpredictable, these numbers are necessarily approximate, but they are close enough to give us confidence that the work can be completed in the time and budget we are working with. During this phase of the project, we will also be exploring the possibility of having community members perform the initial transcription of texts, either in plain text through a crowd-sourcing interface (leaving the subsequent encoding and proofreading passes for ECDA staff) or possibly through outreach to classes (for instance, in the Boston public schools) and organizations in which students or community members might receive some guidance in using basic TEI.
The ECDA will also be experimenting with other methods of text acquisition, the most significant of which is a project to explore automated detection of relevant materials in large-scale digital collections such as HathiTrust, Google Books, or the Internet Archive. ECDA is currently partnering with David Smith, a faculty member in Computer Science at Northeastern, to seek funding from the Andrew W. Mellon Foundation to develop computational methods for detecting texts with content similar to those in an ECDA test collection; the texts detected could then be acquired by dLOC on behalf of ECDA. If this project is funded, the ECDA would additionally acquire a much larger set of texts which would be available for crowd-sourced transcription and encoding. The most significant of these texts would be staged for transcription by the project’s work-study students and RAs, time and resources permitting.

Methods of Community Engagement

The ECDA has developed a strong community engagement plan, which emphasizes outreach and support to scholars, educators, and students of the Caribbean. This plan aligns as well with the political, economic, cultural, and technical concerns of digital humanities organizations and initiatives such as GO::DH, THATCampCaribe, #guerrilladh, and transnational organizations such as the Partner University Fund, part of the French American Cultural Exchange. Our goals for the project are based on a deeper engagement with this region that over the life of the project will include assessment of and response to this community’s current research and archival needs. These include, most significantly:

- working with local archives (institutional, community-based, and individual) in digital preservation and open-sourced publication
- working with local scholars to acquire materials of note in their individual and collaborative research
- working with non-English speakers to promote multilingual digitization and publication of archival materials (this might involve support workshops with these scholars on transcription and TEI encoding practices)
- assessing and responding to technology needs (such as bandwidth and accessibility)

Perhaps most importantly, the ECDA will work closely with users in the region to think more critically about the building practices and function of the ECDA, and will actively solicit strategic involvement from users in the long-term development of the project. Specifically, we expect to convene a board of advisors that includes prominent scholars from the Caribbean as well as other geographic locations. In addition, we note that dLOC has robust ties with Caribbean archives and archivists, and we hope to build on these connections in order to link the ECDA with Caribbean scholars and potential users of the archive.

4. Sustainability of Project Outcomes and Digital Content

The deliverables for this project fall into three categories, each of which requires a different kind of sustainability planning. Digitized page images and other image files (e.g. of maps) will be acquired and managed in the first instance by the Digital Library of the Caribbean, which takes responsibility for management of intellectual property issues surrounding these materials and for making them available on an open-access basis in perpetuity. These images, as well as all of the other scholarly content developed for the ECDA — including TEI transcriptions of documents, user-contributed annotations, and exhibits, and also the TEI ODD files and supporting documentation — will also be stored in the Northeastern University Digital
Repository Service. The Northeastern University Library has made a firm long-term commitment to the curation of scholarly data in the DRS, without cost to the project and without limitation in the case that the project staff changes or leaves Northeastern. This long-term curation also includes automated migration of standard formats (such as PDF, TIFF, and XML) into future versions and formats. Because all of the data formats used in this project are based on open standards and have no hardware or software dependencies, the only long-term curation concerns have to do with the evolution of those formats themselves, and the possibility that they may be replaced in common usage by other formats. Over the very long term, there may also be a need for more detailed curation and migration of data beyond what can be handled automatically (for instance, to future versions of the TEI Guidelines). The library’s Digital Scholarship Group is responsible for maintaining the viability of digital scholarly research at Northeastern and will assist projects with specialized data curation tasks.

The project’s user-related data, including ECDA user accounts, discussion forums, personal collection notes, and other data related to specific users will be stored as part of the ECDA project framework, in a database that will be maintained following the library’s standard data management practices including regular off-site backup and appropriate security for personal information.

The Hydra head being developed for the ECDA is planned explicitly as part of the library’s overall development and expansion of the Digital Repository Service as a system for scholarly digital publication and research. As part of the project’s design, this Hydra head and its component gems will serve as a model for other DRS-based projects at Northeastern. The library manages its software projects, including code management, issue tracking, documentation, and iterative releases in a public Github environment and offers its repository at https://github.com/NEU-Libraries. Code forked from external sources is also managed in this environment thereby maintaining a link to the original development trunk.

There is a further sustainability issue connected with the ongoing support for the ECDA as a project. The ECDA is currently in its early development phase which is being supported through a combination of internal funding from Northeastern and external funding in the form of grants. At the conclusion of this development phase, the project’s main ongoing activities will be community-driven: acquisition and transcription of further texts for the collection, annotation and commentary through the ECDA’s community forums, and development of exhibits and curricular materials by community members. Basic support for the project’s systems will be provided by the Digital Scholarship Group, which maintains the Digital Repository Service and associated publication systems. In addition, the library will continue to develop its Hydra framework for repository-based scholarly publications and as new features and interface tools become available they will also be integrated into the ECDA interface. Ongoing server costs are approximately $1000 per year; the library is currently developing a joint business model with Northeastern’s constituent colleges for the ongoing support of digital scholarly publications. The project would seek external funding for any major new development work.

5. Dissemination

The dissemination of the materials developed for this project will take several forms. The ECDA collection will be published as a formal digital project under the Northeastern University
The Library’s Office of Scholarly Communication and Digital Publishing, which will continue to serve as the publication rubric for the ECDA over the long term. The ECDA is committed to making all of its materials (including those contributed and developed by project contributors) available on an open-access basis for the long term. Moreover, an important part of the ECDA’s design planning for this project is provision for dissemination to users specifically in the Caribbean region. As noted in the Methods section above, the ECDA has developed a strong community engagement plan, which emphasizes outreach and support to scholars, educators, and students of the Caribbean. Based on this engagement, the most significant strands of our dissemination are:

- working with local educators to make materials accessible to their students
- developing curricular materials, including sample assignments for teachers to use that instruct students on how to transcribe and annotate texts, and how to create exhibits of texts related to their own interests
- working with local educators to introduce and mentor students in practices of digitization, encoding, composing, and analysis through guided classroom workshops and crowd-sourced transcription activities
- offering workshops and training through the NULab in a wide range of topics from platform use, to digitization practices, to advanced analytics for new and experienced users
- making presentations at conferences to other scholars about use of the archive for teaching and research purposes: conferences include the Early Caribbean Society, THATCampCaribe, American Literature Association, and the Caribbean Studies Association

Another important grant outcome to be disseminated is the Hydra head and its component “gems.” These are of greatest relevance to the Hydra developer community, especially since Hydra is designed with the goal of facilitating the exchange and reuse of Hydra components. The chief dissemination mechanism to this community is the Git open-source code repository, where Hydra components and documentation are maintained and through which updates and contributions of code are managed by the community. All of the Hydra code for this project will be maintained in Git.

In addition to our public Git repository, we will engage the Hydra community directly in the code development process by bringing the project to one Hydra coding event during the course of the grant (events like Hydra Camp or Open Repositories hack-fests are the regular mechanisms by which the Hydra community engages in face-to-face code development and sharing of progress) and by hosting a project sponsored Hydra event at Northeastern. While the typical Hydra Camp or hack-fest event focuses on general development within the Hydra framework, the Northeastern Hydra event will focus on Hydra applications that support the digital humanities.

For the TEI community, an important deliverable from this project is the ECDA’s TEI data and TEI schema customization (represented as a TEI ODD file, which is a meta-schema that encapsulates both the schema specification and the accompanying element documentation), as well as the additional documentation concerning transcription and encoding methods. These materials will be disseminated via the ECDA’s web site, and also via the ECDA’s presence in TAPAS (the TEI Archiving, Publishing, and Access Service) where the ECDA will maintain a
test collection. At the time of writing, TAPAS is scheduled to launch in fall 2014, and the ECDA has registered as an early adopter project.

A final important mode of dissemination for this project will be the eventual inclusion of ECDA metadata records in the Digital Public Library of America. The Northeastern University Libraries participate in the Digital Commonwealth project (https://www.digitalcommonwealth.org/), and all publicly available materials from this project will have their metadata harvested by and included in the Digital Commonwealth portal. Once in the Digital Commonwealth portal, the metadata records will be harvested, in turn, by the Digital Public Library of America (http://dp.la). No intervention by project of Library staff will be required during this portal-to-portal migration.

### 6. Work Plan

#### May 2015 - October 2015
- Hire Hydra developer (Project director)
- Develop prototype architecture (Hydra architect)
- Initial meetings to develop functional specifications (project team)
- Begin building prototype: commons and annotation, repository deposit and access (Hydra developer and architect)
- Recruit user testing community (RA)
- Begin TEI encoding (RA and work-study students)

#### November 2015 - April 2016
- Continue development of commons and repository functions (Hydra developer)
- Begin user testing and community outreach (RA, project team); refinement of commons and repository functions (Hydra developer)
- Hire web designer
- Continue TEI encoding (RA)

#### May 2016 - October 2016
- Complete prototype (Hydra developer)
- Begin work on interface in Bootstrap (web interface developer)
- Attend Hydra camp for initial project exposure to Hydra community (Hydra developer)
- Identify components of the prototype that should be extracted and converted into gems for reuse by other humanities projects. (Hydra developer and architect, with input from Hydra community)
- Begin developing Hydra gems from prototype components (Hydra developer and architect)
- Continue user testing and community outreach (RA)
- Continue TEI encoding (RA and work-study students)

#### November 2016 - April 2017
- Attend Hydra community event, with focus on refinement of existing gems, getting feedback (Hydra developer)
• Organize and host a “Hydra for Humanities” event (hackfest/user group) meeting at Northeastern (Hydra developer and architect)
• Implement gems version of project (Hydra developer and architect)
• Continue user testing (RA)
• Continue interface development in Bootstrap (interface developer)
• Continue TEI encoding (RA and work-study students)
• Begin user documentation (RA)
• Begin documentation of gems (Hydra developer)

May 2017 - October 2017
• Host humanities-focused Hydra event at Northeastern, with focus on final polishing of gems and dissemination of results (Hydra developer and architect)
• Complete documentation of gems (Hydra developer and architect)
• User testing, bug fixing (Hydra developer and architect)

November 2017 - April 2018
• Final assessment (project team)
• Prepare white paper and final report (project team)
• Dissemination (project team)

7. Staff

David Cliff is Senior Digital Library Developer in Northeastern University Snell Library, and will serve as co-principal investigator for this project. His work focuses on the developing and expanding the campus-wide Digital Repository Service, customizing a local installation of VIVO and the development of custom data loaders for VIVO, coordinating the work of other programmers and developers, as well as the supervision of cooperative education student employees. He will be responsible for the overall technical architecture of the project, working closely with Patrick Yott, and for overseeing the work of the rest of the technical team. He will also be responsible for developing the prototype Hydra components into finished “gems.”

Elizabeth Maddock Dillon is Professor of English at Northeastern University where she is Founding Co-Director of the NULab for Maps, Texts, and Networks and teaches in the field of eighteenth-century transatlantic literary studies, with a focus on early America and the Caribbean. She is also the co-director of the Futures of American Studies Institute at Dartmouth College. She is the author of New World Drama: The Performative Commons in the Atlantic World, 1659-1859 (Duke University Press, 2014) and The Gender of Freedom: Fictions of Liberalism and the Literary Public Sphere (Stanford University Press, 2004) which won the Heyman Prize for Outstanding Publication in the Humanities at Yale University. Together with Michael Drexler, she is co-editor of the forthcoming volume, The Haitian Revolution and the Early U.S.: Histories, Geographies, and Textualities (University of Pennsylvania Press). She has held research fellowships at the American Antiquarian Society (NEH) and the John Carter Brown Library at Brown University. In fall of 2014 she will be a Distinguished Visiting Fellow at the Advanced Research Collaborative at the CUNY Graduate Center; in spring 2015 she will be a visiting fellow at the Charles Warren Center at Harvard University. She has also been named an ACLS Digital Innovations Fellow for 2014-2015 in support of work on the Early Caribbean Digital Archive. She will serve as co-principal investigator for this project and will
provide general oversight over the scholarly and intellectual integrity of the project and serve as a liaison to the ECDA community. She will also contribute to the development of the specifications for the ECDA interface and work flow, and oversee the work of the Research Assistant.

Julia Flanders is the Director of the Digital Scholarship Group in the Northeastern University Library and a Professor of the Practice in the English Department; she also directs the Women Writers Project. She has a PhD in English Literature and has served in positions of leadership in the digital humanities and TEI communities for over a decade: on the TEI board of directors (Chair, 2004-2005), on the Steering Committee of the Alliance of Digital Humanities Organizations and CenterNet since their founding, and as president of the Association for Computers and the Humanities (2008-2011). She has extensive experience managing digital humanities projects and working with sustainable funding models for digital humanities services. In addition to extensive expertise in TEI encoding, she also has deep knowledge of digital project design and management and digital humanities research methods. She will serve as co-principal investigator and will provide guidance and oversight on the design of TEI schemas and encoding specifications.

Amanda Rust is Assistant Department Head, Arts & Humanities in Northeastern University’s Snell Library. Her work focuses on library services and instruction for the arts and humanities and, more recently, archives outreach and the digital humanities. She holds offices in a national forum for literature librarians, and has presented or written on topics like information literacy and instruction, Drupal and website design, theatre archives, library special collections and Wikipedia, and the digital humanities. She will contribute to the planning and design of the ECDA interface, user testing, and community engagement processes.

Sarah Sweeney is the Digital Repository Manager for the Northeastern University Library and is responsible for managing and maintaining Northeastern’s Digital Repository Service (DRS). The DRS is the primary storage and discovery tool for Northeastern University's digital assets, which includes documents, photographs, and audio and video files from various Northeastern offices, research groups, and departments. Sarah is actively engaged with the Northeastern community and works with many campus groups to ensure their digital resources are secure and accessible in the DRS. She is skilled in various digital metadata tools, practices, and standards (such as metadata schemas, XSLT and XPath, ontologies, and content modeling), and traditional library cataloging standards and procedures. Sarah has a BA in English from Hartwick College and an MLS from Simmons College, and she is currently engaged in coursework to develop her understanding of XML and RDF-based Semantic Web technologies. She will contribute expertise in metadata standards and repository systems, and serve as a liaison for the project with the Digital Repository Service and its development team.

Patrick Yott, who will serve as principal investigator for this project, is the Associate Dean of Libraries for Digital Strategies and Services at Northeastern. He has been working in the digital library arena since the mid 1990’s and has held strategic positions at the University of Virginia (1995-2001) and Brown University (2001-2010). He has taught numerous workshops on XML/XSLT, PHP/MySQL and other digital technologies for the Association of Research Libraries, the American Library Association, the New England Library Network and other regional consortia. Patrick is the lead architect of Northeastern’s Digital Repository Service and
related systems and led the initial development of the Brown Digital Repository Service prior to accepting a position at Northeastern in 2010. He will provide general oversight over the project.

**Hydra developer, to be hired:** This position is responsible for developing the Hydra head.

**Web interface developer, to be hired:** This position is responsible for the web interface design and development.

**Research Assistant, to be hired:** The Research Assistant is a humanities graduate student at Northeastern University with expertise in digital humanities, text encoding, and a relevant humanities field. This position will oversee the work of transcription and encoding, and will perform the advanced encoding passes. The RA will also lead the user engagement and user testing efforts for this project.

**Work study students, to be hired:** These positions are undergraduate students at Northeastern University.

### Budget Justification

This grant budget covers salary and benefits for two staff positions: a Hydra programmer at 50% FTE for two years, and a web interface development position at 75% time for one year. It also includes funding for a graduate research assistant for two academic years. Additional staff time will be cost-shared by Northeastern University: Amanda Rust (Research and Instruction librarian) at 5% FTE; Sarah Sweeney (Digital Repository Manager) at 5%; David Cliff (Senior Digital Library Developer) at 10% FTE; Julia Flanders (Director of the Digital Scholarship Group) at 5% FTE; Patrick Yott (Associate University Librarian for Digital Strategies and Services and co-PI on the grant) at 3% FTE; and Elizabeth Maddock Dillon (Professor of English) at 5% FTE. Benefits are calculated at 26.4%. Indirect costs are calculated at 54.5%. In addition to salaries and benefits, the grant budget covers travel costs for attendance at two conferences to present the project’s work, and also travel costs to three Hydra camps. The budget also covers event costs for Northeastern University to host a Hydra event.