I. TABLE OF CONTENTS

I. Table of contents................................................................. 1
II. List of Participants............................................................. 2
III. Abstract............................................................................. 3
   A. Abstract
   B. Statement of innovation
   C. Statement of humanities significance
IV. Narrative........................................................................... 4
   A. Enhancing the humanities through innovation
   B. Environmental scan
   C. History and duration of the project
   D. Work plan
   E. Staff
   F. Final product and dissemination
V. Project budget................................................................. 10
VI. Biographies......................................................................... 11
VII. Data management plan..................................................... 12
VIII. Letters of support........................................................... 13
XI. Appendices.......................................................................... 17
   A. Sampling of woodcut illustrations
   B. Variants of woman with feathered fan
   C. Sample of variant illustration detail
   D. Illustration cataloging interface
   E. End-user “seed” illustration interface
   F. End-user “seed” illustration details
II. LIST OF PARTICIPANTS

Fumerton, Patricia. Professor and Director, The English Broadside Ballad Archive (EBBA), University of California, Santa Barbara.

Nebeker, Eric. Lecturer and Assistant Director, EBBA, University of California, Santa Barbara.

Palmer Browne, Megan. Doctoral Candidate, English and Medieval Studies; Woodcut Impressions Specialist, EBBA, University of California, Santa Barbara.

Stahmer, Carl. Research Scientist and Associate Director, EBBA, University of California, Santa Barbara.
III. ABSTRACT

A. Abstract
Focusing on the expansive English Broadside Ballad Archive (EBBA), http://ebba.english.ucsb.edu, containing over 2,000 distinct 17th-century woodcut illustrations, our proposed Ballad Illustration Archive (BIA) will allow unprecedented access to these hard-to-access images that are important cultural and artistic productions. Our project will make significant technological inroads through innovative integration of computer vision software and human cataloguing, delivering to the end-user a product which is technically cutting-edge and marked by careful scholarship. It will thus enable enhanced research in multiple humanities disciplines and also make these compelling images available to the interested non-specialist public. Ultimately, we see this project expanding to include a wider variety of early modern illustrations; we also expect it to expand the possibilities for future digital scholarship.

B. Statement of Innovation
BIA will create an innovative illustration search engine using state-of-the-art computer vision algorithms and metadata cataloging in order to provide users the ability to search images as thoroughly and easily as they can search texts. We will also use computer vision algorithms to create an innovative illustration-cataloging interface that will present catalogers with a relevance-ordered, computer-generated list of similar images, allowing them to compile thorough and accurate metadata.

C. Statement of Humanities Significance
BIA will offer unique access to broadside ballad illustrations, allowing them to be studied as artifacts in their own right. Art historians will be able to study the production techniques, motifs, and genealogy of these previously unavailable ballad illustrations; literary and cultural scholars will be able to trace relationships between texts and images; bibliographers will be able to use rich cataloging data to better understand the complexities of the printing industry.
IV. NARRATIVE

A. Enhancing the Humanities through Innovation

The Ballad Illustration Archive (BIA) requests critical start-up funding from the NEH in order to make the striking illustrations frequently printed on broadside ballads freely and fully accessible as artistic and historical artifacts. As a daughter of the English Broadside Ballad Archive (EBBA), http://ebba.english.ucsb.edu, at the University of California, Santa Barbara, BIA is well situated to undertake this task. Since 2003, and largely supported by three generous NEH Collections and Resources Grants, EBBA has been successfully making early modern broadside ballads fully accessible as texts, art, and songs. About 8,000 English broadside ballads of the 17th century survive, and EBBA manages the largest digital collection of these important objects, with 4,400 ballads archived (nearly 55% of extant 17th-century broadside ballads). Our project will extract the over 2,000 distinct illustrations from EBBA’s digital library and make them fully searchable using an innovative combination of computer vision programming and robust scholarly cataloguing. These innovations will transform both research and metadata cataloging practices for “cheap print” illustrations.

In its heyday of the first half of the 17th century, a broadside ballad was a single, large sheet of paper printed on one side with multiple eye-catching woodcut (or, occasionally, metal-plate) illustrations, a popular tune title, and an alluring poem. The illustrations themselves are strikingly graphic, frequently detailed, and often whimsical. Depicted subjects are remarkably varied, and include individuals and couples in flamboyant early modern fashions; elaborately-designed galleons in full sail; cityscapes; complicated and vivid continuous narrative scenes; and humorous or sinister demons, skeletons, and monsters (see Appendix 1). These images were often re-used on several ballads, sometimes from the same wood block and sometimes from a copy. We use the term “impression” rather than the commonly-used but much less specific term “woodcut” to remind users of what these images are—printed impressions made by inked blocks or plates rather than the blocks or plates themselves. Currently, no archive of early modern cheap-print illustration exists for the 17th century. This is primarily because the task of personally examining each extant ballad, chapbook, or tract would be extremely costly and time consuming. Indeed, it is largely due to the efforts of EBBA that intensive research into early modern ballad illustrations is currently possible at all.

EBBA’s quintessential example of a woodcut impression with many variants is a woman in a flower-decorated dress holding a feathered fan in her right hand (see Appendix 2). This image appears fifty-five times in the Pepys archive, thirty-five in the Roxburghe archive, and many more times in other ballad collections. At first glance, all such images seem the same. On further inspection, however, visible differences arise: for example, some impressions reverse positive and negative space, and sometimes corresponding pictorial elements are shaped differently (for example, see Appendix 3). These differences indicate that the impressions were printed from different blocks. At present, however, identifying the different blocks for all the various impressions is prohibitively time-consuming and often imprecise, as it is based on the determinations of individual cataloguers working over prolonged and intensive periods of time. Our proposed Ballad Illustration Archive, with its combination of machine- and human-created data, will significantly decrease the human time needed to make such distinctions while significantly increasing the amount of scholarly data available to researchers. Once BIA is fully developed with computer vision technology, it will be an image-centered archive which partners with EBBA’s already vibrant ballad-centered archive, opening new roads to scholarship in art history, cultural studies, and bibliography.

For art historians, access to these impressions will greatly advance understanding of early modern illustration practices. The on-the-fly computer-generated visual recognition will enable these scholars to focus on particular details of the impressions and to find other images that contain similar details. Cultural scholars will reveal new understandings of the relationships between images and texts, which will enhance our understanding of the period and make the images themselves much more comprehensible. Bibliographers will be able to use BIA to identify particular blocks or plates that were used to print
images that might otherwise seem identical. Such understanding will enable users to follow the history of these blocks, thus improving our understanding of how resources were transmitted between printers, refine conjectures about dating, and improve available metadata. Because the impressions have long been difficult to access and because there is so little art historical, cultural, and bibliographical scholarship around them, few scholars who would otherwise be interested in this important field have been willing to jump into the morass of sorting through them.

B. Environmental Scan

The fundamental problem facing an archive of visual material is how to make such material findable through simple, intuitive queries that return a meaningful result-set. To date, no satisfactory solution to this problem has been implemented in anything resembling a wide-scale, publicly accessible archive. Attempts to solve this problem have generally fallen into one of two general categories: 1) human cataloguing of archive items, either in a controlled cataloguing environment or through social tagging; or 2) attempts to apply computer vision algorithms to collections of images in order to automate the process of establishing “sameness” or “difference” between images.

Human cataloging

Because of the new capabilities provided by digital archiving, scholarly interest in early English woodcut illustrations has grown exponentially in the last several years. EBBA has been developing its woodcut impression catalog for over five years; other projects are beginning to show interest in developing similar tools. Discussion of three major projects follows—those in development by the British Printed Image to 1700 (Bpi1700) project, by the Folger Shakespeare Library LUNA Project, and by the Iconclass Project:

1. Bpi1700 (http://www.bpi1700.org.uk): This important and innovative project, undertaken by the United Kingdom’s Arts and Humanities Research Council, the Birkbeck University of London, the Victoria and Albert Museum, King’s College London, and the British Museum, is a digital archive of early modern book illustrations. It includes a wealth of important cataloguing information about illustrations from our period, but its scope does not extend to ballad illustrations. It also does not benefit from the latest advances in technology. For example, users can search by using cataloguing information, but they cannot perform image-centered searches. Furthermore, the database’s cataloguing structure, because it is set up for cataloguing fine art, is not always germane to the needs of our project. Categories such as “Artist,” and “Sitter,” for example, are almost never known for lowly ballad illustrations. Instead, we work from descriptions and keywords, moving from single images to larger categories of those images so that users can navigate the largely anonymously carved and multiply repurposed ballad illustrations.

2. Folger Shakespeare Library LUNA Project (http://luna.folger.edu): An online library of high-resolution images from the Folger’s collections, LUNA provides rich cataloguing information in an easy-to-navigate sidebar, and allows users to view multiple images at once, to zoom in and out, to download thumbnails, and to make permanent links to images. Though its images are high quality and its overall design stunning, this project has also not benefitted from the use of computer vision software, which would allow users to compare images. Nor does it offer keywording or short descriptions of each image, as BIA will. Also, of course, its work is focused on local holdings at its own library, while our project aims to catalogue and make freely available similar kinds of images across different libraries.

3. Iconclass (http://www.iconclass.nl/): This major cataloguing project, begun in the 1950s and now headquartered at the Netherlands Institute for Art History, is a subject-based image classification system hierarchically ordered on a complex tree system. It consists of 28,000 definitions and 14,000 keywords. It was built to be comprehensive across all forms and periods of visual art. But Iconclass impedes rather than helps ballad image searches due to the restricted media, size, and purpose of ballad illustrations. It is too large and unwieldy; it involves many hours of labor for cataloguers; and it results in a sometimes difficult-to-manage interface for end users. Finally, like the projects listed above, it does not include the computer vision software we propose, nor does it offer the same level of catalogued metadata, linking images with duplicates and variants.
**Computer vision algorithms**

While computer vision began to coalesce as a serious and unique field of study as early as the 1970s, the first truly successful attempts to model human visual recognition strategies with computers can be traced to the work conducted in the 1990s as part of the Defense Advanced Research Projects Agency (DARPA) and the United States Army Research Laboratory's FERET program. Research interest in the field has continued without interruption over the years, and there are several examples of successful attempts at achieving a workable system. The most notable early examples of these include the Webseek and Webseer search engines which attempted to provide Internet-wide image-based content retrieval, as well as a host of early commercial applications such as Informix Datablades, IBM DB2 Extenders, and Oracle Cartridges. Despite these early manifestations and a great deal of devoted research since that time, however, there is still nowhere on the web where a user can go, select a sample image, and find other images either within a single archive or across the net. The reason for this is surprisingly simple: While at a research level we have been able to perform image-based content retrieval for some time, at a computing level few users have had the computing power necessary to perform the system-intensive calculations required by the recognition algorithms. Thus, despite advances made in the algorithms in the past ten years, the lack of computing power has prevented their widespread application. Recent advances in computing hardware technology have now changed that equation, and effective application of image-based content retrieval is now within our grasp.

There are several open source Computer Vision engines. BIA will use Perceptual Image Diff ([http://sourceforge.net/projects/pdiff/](http://sourceforge.net/projects/pdiff/)), a Computer Vision engine developed as an open-source project under the direction of Hector Yee ([http://hectorgon.blogspot.com/index.html](http://hectorgon.blogspot.com/index.html)). It is not an end-user application, meaning you cannot install and run it on your desktop computer. Rather, it is a library of application classes that, when invoked, use “a computation model of the human visual system to determine if two images are visually different.” Implementing Visual Image Diff requires compiling and installing the library on the local server and the development of additional software that invokes the library and returns the results of its processes to the end user. In layperson’s terms, obtaining Perceptual Image Diff is like obtaining a pre-assembled and fully functioning engine for a car. But in order to make use of it, you have to build a car to put it in.

BIA plans to build two different “cars,” each of which will be driven by the Perceptual Image Diff engine. First, we will integrate the engine into our cataloguing environment so that whenever a cataloguer is cataloguing an impression, the interface will use Perceptual Image Diff to automatically present the cataloguer with similar images as well as the cataloguing information for them (see Appendix 4). This will allow the system to “recommend” tags as well as compare the current cataloguing work with that already completed on similar impressions. The second BIA “car” to be powered by Perceptual Image Diff is our front-end website. We will develop a php and Javascript based plugin to EBBA’s existing website that will allow users to directly search all impressions for other occurrences of the same or a similar image. Using this new plugin, the user will select a “seed” image of interest and the computer will return a ranked and ordered list of similar images in the archive. This plugin will further provide the user with the ability to match the entire seed image or to select (using a mouse or trackpad) a sub-region of the image and search for all occurrences of the selected region within all other images in the archive (see Appendices 5 and 6). Returned results will always link the found image back to the ballad on which it appears and also use an index created by the Perceptual Image Diff engine and human cataloguing to provide cross-links to all related ballads and woodcut impressions.

**C. History and Duration of Project**

BIA is an outgrowth of EBBA, which has always understood the importance of and been committed to the cataloging of woodcut impressions. Indeed, cataloguing woodcut impressions has been an important part of EBBA’s development since its founding in 2003. In our nascent phase of development we experimented with a version of Iconclass as a means of describing the archive using a controlled
vocabulary (see Section B above for further discussion of Iconclass). But we quickly found that a
generalized taxonomy such as Iconclass did not meet the specific and specialized needs of a particularized
scholarly archive such as EBBA. Thus we set out to develop our own controlled vocabulary for use in
cataloguing the impressions based on our experience with the Pepys collection.

As EBBA expanded to include other ballad collections, it quickly became apparent that our second
attempt at a controlled vocabulary was still not sufficient to describe the wider range of subjects and
themes presented by the new additions. Happily, we were able to secure additional funding from a UCSB
Academic Senate grant in order to review our woodcut impression cataloguing practices and develop a
more effective system for describing and cataloguing the woodcuts. As part of this work effort, we
established a team of dedicated woodcut cataloguers, under the direction of Palmer Browne, who set out
to collaboratively review, categorize, and develop a logical and manageable system for better describing
all woodcut impressions from the period. This third-generation controlled vocabulary was fully
implemented and remains functional today via EBBA's advanced search interface.

The EBBA impressions team continues to hone the controlled illustration vocabulary, including re-
tagging impressions already catalogued if a new control word is added to the vocabulary. This work has
been facilitated by the development of an in-house impressions management database wherein woodcut
impressions that depict similar themes, figures, or events are associated with each other at the database
level (in what we internally call “woodcut impression groups”) so that cataloguers are able to identify and
retrieve related woodcuts. In addition to creating a computer vision engine, this grant would allow us to
improve our illustration-to-illustration linking system by making our cataloguing practices more granular
and their results available to end-users.

Duration of Project
The project requires an 18-month work effort that would combine EBBA's experience working with
broadside ballads in general and wood- and metal-cut impressions specifically with the application of the
latest technological advances in computer vision and automated image recognition.

D. Work Plan
The plan of work schedule listed below is based on EBBA’s already extensive experience to date with
cataloguing woodcut impressions and implementing various state-of-the-art technologies. The work effort
is divided into three parallel development tracks: 1) implementation of computer vision software and
integration into our current technology infrastructure; 2) preparation of woodcut impression images (see
below); and 3) human cataloguing of each woodcut impression image.

Implementation of computer vision software
It will take Stahmer working 49% time approximately 1 month to install and configure the Perceptual
Image Diff engine, 2 months to integrate the engine into our current cataloguing interface, and 3 months
to create a working version of an image searching interface plugin to our website application. As the
schedule below indicates, this work will be front-loaded at the beginning of the grant period as much of
the remaining work by Stahmer at 10% time involves further implementation of this software in the
cataloguing system and ongoing bug-checking and tuning of the system as it is implemented, as well as
direct oversight of the overall work-effort.

Preparation of woodcut impression images
Because our archive considers each impression a distinct artifact, we will create a separate file for every
single impression we catalogue. Thus, for example, the impression archive for the Pepys ballads will
contain 4,090 image files. In order to create these images, Palmer Browne and a graduate research
assistant will work with EBBA’s existing ballad images, using Adobe Photoshop to cut out and save each
individual impression. These files will be stored on EBBA’s existing servers.

We estimate 25% time for Palmer Browne and 100 hours for the RA over 6 months to complete this task.
**Human cataloguing of woodcut impression images**

The human cataloging of ballad impression images will expand on the impression cataloging already completed by EBBA (placing impression images in the database, associating them with related images, and keywording the images). Palmer Browne, Nebeker, and a graduate student cataloguer will divide these large groups of related images into sets of variants. In the case of the lady with the feathered fan, for example, one group of impressions shows a horizontal crack across the skirt of the woman’s dress (see Appendix 2). A cataloguer noting this would use the software to find all similarly-shaped cracks and, studying the returns, would group all the images with this variant into yet a sub-set of the larger lady-with-fan group. Then the cataloger could establish a chronology based on the patterns of wear. With the aid of computer vision technology, the entire process will take the cataloger approximately one hour of work for each image, a process that might otherwise take sixty hours.

We expect it will take Palmer Browne, Nebeker, and a graduate researcher approximately twelve months to complete this human cataloguing, working respectively at 25% time, 10% time, and 200 hours.

**Schedule of Work**

**July 2012 – December 2012:**

- Fumerton works 15% time in a supervisory capacity.
- Stahmer works 49% time installing, testing, configuring and modifying as necessary for this particular use image recognition/computer vision software packages (PerceptualDiff) and integrating them into back-end cataloguing interface. Stahmer begins work on integration of image recognition software into front-end website.
- Palmer Browne works 25% time designing initial user interface for new front-end image search functionality.
- Nebeker works 10% time researching specialized cataloguing information.

**January 2013 – June 2013**

- Fumerton continues to work 15% time in a supervisory capacity.
- Stahmer works 10% time continuing development of front-end image search functionality. Upon completion Stahmer works on direct project oversight and fixing any bugs in implementation of new software.
- Palmer Browne works 25% time preparing woodcut impression images for cataloguing and ingestion by image recognition software, extracting stand-alone woodcut images from larger ballad images and saving each as an independent file. Palmer Browne begins work in performing Quality Assurance (QA) testing of visual search interface.
- A graduate Research Assistant works 100 hours aiding Palmer Browne.
- Nebeker continues to work 10% time researching specialized cataloguing information.

**July 2013 – December 2013**

- Fumerton continues to work 15% time in a supervisory capacity.
- Stahmer continues working 10% time on direct project oversight and fixing any bugs in implementation of new software and fine-tuning image search algorithms.
- Palmer Browne continues working 25% time cataloguing woodcut impression images and performing Quality Assurance (QA) testing of new visual search interface.
- Nebeker works 20% time researching specialized cataloguing information.
• A graduate Research Assistant works 200 hours aiding Palmer Browne and Nebeker in cataloguing.

Subsequent to the NEH Grant
Woodcut impressions for all broadside ballad collections subsequently archived by EBBA will be catalogued and archived using the new system so that all will be visually searchable across the EBBA database. If EBBA funding continues, these will include the early Crawford collection of 1,466 ballads at the National Library of Scotland, Edinburgh; the 16th-century broadside ballads in the library of the Society of Antiquaries, London; holdings at the Chetham Library, Manchester; and John Bagford’s three volume collection as well as Narcissus Luttrell's three volume collection, together with the Huth ballads at the British Library. We are currently collaborating with the Bodleian Library to make our combined archives mutually searchable—both ballads and individual illustrations. A long-term goal of our project is to catalogue and facilitate widespread access to all early modern English woodcut impressions—an enormous and exciting task which has never before even been considered, but will be increasingly feasible as our project gains ground.

E. Staff
Patricia Fumerton: Project Co-Director. Fumerton will act in a supervisory role and have editorial control.
Carl Stahmer: Project Co-Director. Stahmer will implement Perceptual Image Diff computer vision and build the interfaces for end users and catalogers. He will also act in a supervisory role.
Megan Palmer Browne: Project Manager. Palmer Browne will assist Stahmer in designing the user interface, extract stand-alone woodcut impressions, catalog woodcut impressions, perform quality assurance, and supervise research assistants.
Eric Nebeker: Research Specialist. Nebeker will research specialized cataloging information such as ballad printers and publishers and ballad dates.

F. Final Product and Dissemination
The goal of this project is to provide easy and open access to the images, citation information, and subject matter of ballad illustrations. Our completed archive will consist of computer vision software which can show related images to users in real time supported by a robust catalogue which will provide a multitude of information about each individual image, about related images, and (through already-existing EBBA architecture) information about the ballads on which these images were printed. The resulting wealth of data will allow scholars and other users an unprecedented—indeed, a previously impossible—look into the rich world of early modern print illustration as art and artifact.

In the second six months of the grant, the Early Modern Center, in conjunction with EBBA, will seek UCSB funding to hold a colloquium to publicize the new archive. The colloquium will feature distinguished scholars of early modern illustration and of technology in the humanities, and it will be announced on all major early modern listservs. In addition, we will promote the Ballad Illustration Archive from UCSB’s English Department website as well as from the Early Modern Center website. The NEH will be recognized as a major sponsor of the archive on both these sites. We also plan to foreground the Archive in canvassing for more government funding for the NEH by attending the Humanities Advocacy Day in Washington, as we did in the spring of 2007. When the archive is complete, we will announce our accomplishments via the NEH Grant through media outlets across the world, including notification of major search engines. Finally, as usual, EBBA team members will continue to deliver presentations and publish articles on this new development of the project.
VI. BIOGRAPHIES

Co-Director of Project: Patricia Fumerton
Fumerton is Professor of Renaissance Literature, founder and director of the English Broadside Ballad Archive and founder of the UCSB English Department's Early Modern Center, and much-published expert on 17th century broadside ballads.

Co-Director of Project and Computerized Development Specialist: Carl Stahmer
Stahmer holds a Ph.D. in English from UCSB in digital humanities and is Associate Director of the English Broadside Ballad Archive. He was also Research Scientist at the Maryland Institute for Technology and the Humanities (MITH) and currently serves on the Steering Committee for the Networked Infrastructure for Nineteenth-Century Electronic Scholarship (NINES).

Project Manager: Megan Palmer Browne
Palmer Browne is a doctoral candidate in English and Medieval Studies at UCSB and Woodcut Impressions Specialist for the English Broadside Ballad Archive.

Research Specialist: Eric Nebeker
Nebeker holds a Ph.D. in Early Modern Studies from UCSB, specializing in ballad and print history and has published several articles on ballads. He is Assistant Director of the English Broadside Ballad Archive and a Lecturer in the UCSB English Department.
ID # 1508: Fashionable Man  
ID # 1447: Fashionable Woman  
ID # 0767: Courting Couple  
ID # 0002: Galleon in Full Sail  
ID # 0573: Cityscape
ID # 0534: Devil with Flail Admonishes Group

ID # 0093: Three Skeletons Rise from Graves

ID # 1625: Monster

ID #1496: Narrative Scene
Appendix B: Variants of Woman with Feathered Fan

Block “A”
Characteristics: White skirt; short, wide fan; large, rounded left eye; very full skirt; both feet showing
State Changes: Loses border; loses left eye; acquires crack in edge of fan

Block “B”
Characteristics: Black skirt; break in top corner of fan; sole of right shoe white; not holding glove
State Changes: No significant state changes in current archive

Block “C”
Characteristics: White skirt; long, narrow fan; straight, small eyes; narrow skirt; one foot showing
State Changes: No significant state changes in current archive

Block “D”
Characteristics: White skirt; rounded fan; dark hair; rounded features; both feet showing
State Changes: Loses fan

Block “E”
Characteristics: Black skirt; pointed feathers in fan; dark hair; rounded features; both feet showing
State Changes: No significant state changes in current archive

Block “F”
Characteristics: Black skirt; rounded feathers in fan; large forehead; full mouth; both feet showing
State Changes: Break in skirt (throughout current archive)
Search for Impressions in Database:

Words in Description:
feathered fan

Current Image to be Catalogued:

Source Ballad Information:
20663. Pepys 2.39
The Wanton Wife of Bath.
Printed for J. Clarke, W. Thackeray and T. Passinger
1685

Impressions in Group:

Top Match: # 0005
Known Variants

Description: woman; holding feathered fan; arm at waist; flowers on dress
Keywords: nobility / aristocracy; woman / women

Possible Relations

# 9779  # 1521  # 1595  # 1600
Keywords: nobility / aristocracy; woman / women; outdoor scene