Virtual Reunification of Dispersed Archival Photographs:
Developing a Collaborative Network and Framework
for Consolidating Fragmented Image Collections

University of Maryland, College Park

The University of Maryland's College of Information Studies together with its Digital Curation and Innovation Center (DCIC) and the Maryland Institute for Technology in the Humanities (MITH), seeks $50,000 for a Collaborative Planning Grant under the National Digital Platform funding priority to support a meeting of photo archivists, visual studies scholars, digital humanists, image processing experts, image digitization experts, and other stakeholders that will nurture a community of interested researchers and set an agenda for advancing and supporting virtual reunification tools for dispersed photographic archives collections.

Virtual Reunification of Dispersed Archival Photographs:
Developing a Collaborative Network and Framework
for Consolidating Fragmented Image Collections

University of Maryland, College Park

The University of Maryland's College of Information Studies together with its Digital Curation and Innovation Center (DCIC) and the Maryland Institute for Technology in the Humanities (MITH), seeks $50,000 for a Collaborative Planning Grant under the National Digital Platform funding priority to support a meeting of photo archivists, visual studies scholars, digital humanists, image processing experts, image digitization experts, and other stakeholders that will nurture a community of interested researchers and set an agenda for advancing and supporting virtual reunification tools for dispersed photographic archives collections.

1. Statement of Need

Scholars have noted the limitations of traditional descriptive standards and practices to represent photographic collections in archival care (Stewart 2010; Benson 2009; Schlak 2008; Mifflin 2007; Schwartz 2000; Bartlett 1996). Many factors present particular challenges that complicate working with photographs. Archival photographic materials often include multiple copies, versions, or formats of similar images. Photographs of the same provenance are often found in various locations or shared amongst several institutions. Format diversity, duplication, and dispersion pose deep challenges for library, archives, and museum (LAM) professionals and administrators attempting to represent photographic images scattered across many institutions (Punzalan 2014). Research users, likewise, have noted the limitations of traditional descriptive standards and practices to represent photographic collections in archival care (Stewart 2010; Benson 2009; Schlak 2008; Mifflin 2007; Schwartz 2000; Bartlett 1996). Moreover, geographically scattered collections present access challenges to scholars wishing to study images of common origin.

Researchers and archivists desire to reunify scattered photographic collections. This desire has been amplified by the seeming promise for easier unification and access in the digital realm. However, while various technical mechanisms, including advances in computer vision techniques and large-scale aggregation of digital collections, are available to enable such efforts, there has not been much coordinated action to overcome the challenge of dispersed image collections. In response to these problems, this planning project proposes to bring together notable stakeholders who can synthesize recent developments in image processing in order to ameliorate the consolidation of dispersed and duplicate photographic archives.

Virtual reunification is the strategy of producing a consolidated, digitized representation of scattered artifacts, literary and artistic works, and/or archival records of a single origin or common provenance (Punzalan 2014; Austenfeld 2010; Lynch 2009; Shenton 2009; Unsworth 2007; Henschke 2007; Deegan and Tanner 2002). The proposed planning grant will respond to some of the most difficult remaining challenges of developing aggregations into virtually reunified collections. These challenges include: What are the remaining technical, expertise, and methodological requirements necessary to provide consolidated access to dispersed photographic collections? How can we harness developments in digitization, linked open data, pattern

recognition, and computer vision to assist cultural heritage professionals to better understand and represent a collection that they share in common? What needs and capabilities would consolidated representation of such collections provide for researchers?

The cultural heritage field is at a significant point of technological transformation (Conway 2014). More and more items in the holdings of cultural heritage collections are being digitized. There is significant increase in developing tools to provide access to cultural items online. Users are becoming more acquainted with digital surrogates and are demonstrating sophisticated ways of interpreting digitized versions of visual artifacts (Conway 2010; Conway and Punzalan 2011). Digital humanists are creating more avenues for interacting and studying cultural heritage materials beyond the parameters of traditional scholarly communication and discourse (e.g., Moretti 2013; American Council of Learned Societies 2006). Library, archives, and museum professionals and administrators are more open to venture into collaborative digital projects as means to make their holdings widely accessible. Moreover, image processing and computer vision technologies in pattern detection and image consolidation have significantly advanced. Recent developments in linked open data have made it possible to connect metadata of collections held in various repositories (Voss 2012; Open Archives Initiative Protocol for Metadata Harvesting 2015). Despite the valuable aggregation of materials from multiple institutions in national digital platforms, and the existence of large aggregations offering the potential to "see" across collections, reunification potentials remain unrealized. On its own, aggregation will not naturally reconstitute scattered collections. Thus, widespread practices and developments, including mass digitization, availability of online representation tools, and large-scale aggregations offer promise for reunification, but more work needs to be done to coordinate efforts and to define priority areas among users in order to realize these emerging areas.

2. Impact

Many projects have explored bringing together dispersed collections using digital surrogates in order to improve context for collections and enrich interpretive possibilities. Digital projects such as the *Walt Whitman Archive*, *Rosetti Archive*, *Codex Sinaiticus*, and *William Blake Archive* demonstrate the value of digital projects to reunify works for researchers. However, this work has primarily focused on textual materials. Some projects have centered on digitized archival photographs, such as the U.S. Farm Security Administration and Office of War Information (FSA-OWI) (http://photogrammar.yale.edu), Japanese woodblock prints database and image (Ukiyo-e), and the work of early 20th century photographer, Edward S. Curtis (http://scalar.usc.edu/works/performingarchive/index). These digital projects demonstrate the feasibility of digitally reunifying photographs from various places. However, there have been no systematic efforts to analyze these projects or to identify future potential for other similar initiatives.

The proposed planning workshop will gather various experts that do not cross paths but are necessary contributors to accomplishing photographic virtual reunification. They will identify steps to overcome some of the most difficult challenges faced by cultural heritage institutions, including better ways of presenting digitized archival photographs, supporting virtual reunification as an inter-institutional endeavor, incentivizing greater cooperation among owning institutions, and developing the capacity for photographic digital reunification. Improving

technical potential and capacity to process large aggregations of digital material will help institutions to better see links and materials complementary to their holdings. Moreover, this work would allow LAM institutions to complete and improve contextual information about their collections, which may be of use to researchers and collections managers. By automating recognition of internal patterns, such as duplicate images across collections, as well as displaying this information to users in useful ways, national-scale aggregations can provide more valuable context to users.

3. Project Design

In order to accomplish these outcomes, the project will organize a two-and-a-half day planning workshop to be held in May 2016 at the University of Maryland, College Park campus. The PI, together with project partners, will implement the work in three stages: planning and preparation, the workshop itself, and coordination and facilitation of targets and goals identified at the workshop.

3.1. Selection/Identification of Participants

This workshop will seek to join constituencies that rarely cross paths: LAM professionals, photo archivists, developers with expertise in computer vision and pattern recognition, image digitization experts, digital humanists, scholars of visual archives, as well as representatives of large-scale digital aggregations. At this juncture, it is important to identify the desired expertise and focus areas for participants in this proposed planning workshop. Though not an exhaustive or definitive enumeration, below are the areas of expertise, selection criteria, and lists of potential invitees.

- **Cultural Heritage Workers**

LAM professionals and administrators with deep understanding of photographic archiving and processing (encompassing selection, acquisition, description, and creation of access systems and online representations) will compose a major group of invited participants. Invitees will include **Matthew Mason** (Chair, Visual Materials Section of the Society of American Archivists and Archivist, Beinecke Rare Book and Manuscript Library, Yale University), **Brett Carnell** (Head of Technical Services, Prints and Photographs Division, Library of Congress), and **Gina Rappaport** (Photo Archivist, Smithsonian's National Anthropological Archives). These experienced photo archivists will shed light on the issues and challenges of archival image processing, preservation, and access. *Representatives from institutions of research photographic materials, processes, techniques*, **Jamie Allen** (Associate Curator, Department of Photography at George Eastman House) and **Tracey Schuster** (Head, Permissions and Photo Archive Services Getty Research Institute) will bring current perspectives on standards on visual representation, access, and use.

- **Visual Archives Scholars and Photographic Digitization**

Scholars of archival photographic methods and representation, **Michelle Caswell** (Assistant Professor, UCLA) and **Joan Schwartz** (Associate Professor, Queen's University) will provide

their academic insights on the larger context of the social and historical impact of photographic archives access and use. *Scholars of photographic digitization*, **Paul Conway** (Associate Professor, University of Michigan School of Information), **Ricardo Punzalan** (Project PI, University of Maryland) and **Zack Lischer-Katz** (Ph.D. Candidate, Rutgers University School of Communication and Information), will provide insights on recent understandings on access to digitized images.

- **Technical Experts in Digital Humanities, Linked Open Data, Platform Developers, Computer Vision, and Image Processing**

It is also important to include those involved in *successful online projects* and *technical development*. Invitees will include **Laura Wexler** (Photogrammar Project), **John Ressig** (Ukiyo-e), and **Jacqueline Wernimont** and **David Kim** (*Performing Archive: Curtis and the 'Vanishing Race'*). *Collections aggregators* such as representatives from DPLA, including representatives from its hubs, and **Frederik Truyen** of Europeana Photography, will share their knowledge on the potential of aggregated collections for reunification. **Laura Mandell**, coordinator of NINES (Networked Infrastructure for Nineteenth-Century Electronic Scholarship) will share her knowledge on aggregation of scholarly works. Representatives from *services and platforms* focused on visual collections such as **Ian McDermott**, collection development manager of Artstor and the project director of Mukurtu, **Kim Christen** have important insights on the various requirements and policy and ethical considerations in platform development. **Ed Summers** (Lead Developer, MITH) and **Trevor Munoz** (Associate Director, MITH and Assistant Dean for Digital Humanities Research, University of Maryland Libraries) will share their expertise in *linked open data and digital humanities*. **Jennifer Guiliano**, a core member of the Digging Into Image Data to Answer Authorship-Related Questions, will contribute her expertise in conducting multi-disciplinary image research at scale. Finally, expertise in computer vision, image processing and pattern recognition are important technical capabilities that will be represented at the workshop. Aside from the project's partner DCIC, the University of Maryland is also home to the Computer Vision Laboratory (CVL). DCIC faculty and staff with prior experience in image processing **Richard Marciano** (Professor, University of Maryland College of Information Studies) and **Greg Jansen** (Research Software Architect, DCIC) will be complemented by CVL experts, **David Jacobs** (Professor, University of Maryland Department of Computer Science) and his doctoral students.

3.2. Projected Performance Goals and Outcomes

This project would provide tangible and important contributions to both the scholarly research and cultural heritage communities by 1) discovering and communicating characteristics of successful virtual reunification for photographic collections; 2) identifying the necessary technical capability for photographic virtual reunification work; and 3) developing a strategy whereby these technologies could be integrated into workflows of cultural heritage organizations as part of their participation in national digital platforms.

The project will also contribute to professional practice. It will initiate steps that lead to the development of community-driven services for photographic virtual reunification, which can be used to identify and reunify other, similarly dispersed image collections. Such a platform in turn

can assist cultural heritage institutions in understanding and representing dispersed collections, including tracing provenance and original order for collections with duplication and metadata variation. This improved context and collection information for digital collections would also add an incentive for institutions to participate in efforts of linking cultural heritage collections online.

The specific tangible outcomes of the planning workshop shall be to:

- Build a community and set up working group structures for ongoing collaboration
- Communicate next steps and specific future challenges to focus community efforts
- Identify requirements or specifications – technical, policy, implementation, expertise
- Collect use cases

3.3 Workshop Structure and Agenda

Participants of the workshop will be asked to: 1) identify concerns, challenges, and barriers in providing a consolidated representation of a dispersed image collection in the context of national-scale digital platforms; 2) explore the potential of virtual reunification as a useful tool for both researchers and owning institutions; 3) identify additional stakeholders and partners; 4) devise a plan to develop both an active community for and technologies to support virtual reunification of image collections; 5) identify and assess appropriate technological platform for photographic virtual reunification; 6) articulate the role of national aggregation platforms such as DPLA and Europeana in improving photographic reunification.

The two-and-a-half day workshop shall be structured as follows:

Day 1 – Understanding Challenges and Opportunities

- Understand the expertise present at the workshop
- Enumerate the challenges of representing dispersed archival photographs
- Identify various critical technical and policy elements in virtual reunification
- Highlight key characteristics of virtual reunification projects
- Enumerate the needs of various stakeholders and how reunification can potentially provide strategies and solutions to current limitations and challenges in image processing

Day 2 – Developing Steps and Strategies

- Understand the potential of current collections aggregation efforts to support virtual reunification
- Identify relevant image processing, pattern recognition, and computer vision technologies in photographic collections representation
- Chart a photographic virtual reunification process that harnesses computer vision technologies, advances in collections aggregation, and helps cultural heritage workers represent dispersed collections

Day 3 – Planning Actionable Goals

- Identify future collaborative project(s)
- Plan next steps and meetings

3.4 Project Timeline

This project will be implemented in three stages: planning and preparation, the workshop itself, and coordination and facilitation of targets and goals.

Stage	Duration	Activities
Planning and Preparation	October – November 2015	• Develop a final shortlist of invitees • Create workshop agenda (including dates and venue) • Design and publish project website
	December 2015	• Send out invitations to prospective participants
Workshop	May 2016	• Update project website to reflect final participant and agenda • Workshop proper
Coordination and Facilitation	May 2016 – June 2016	• Update project website to reflect future projects and expected outcomes

4. Project Resources: Personnel, Time, Budget

4.1 Personnel

Ricardo Punzalan, assistant professor of archives and digital curation at the University of Maryland College of Information Studies, will serve as project director and PI. He has recently published refereed articles on virtual reunification and dispersed photographic collections in the *Library Quarterly*, *American Archivist*, and *Archivaria*. His doctoral dissertation identified barriers and challenges of virtual reunification as a strategy for representing dispersed collections of ethnographic archival photographs, with a focus on materials attributed to Dean C. Worcester (1866–1924), a U.S. colonial administrator in the Philippines.

Two institutions at the University of Maryland with capacity and expertise in technical development and digital humanities will serve as project partners. The Maryland Institute for Technology in the Humanities (MITH) as well as the Maryland iSchool's Digital Curation and Innovation Center (DCIC) will provide technical expertise in digital projects and programming. In particular, DCIC and MITH will help in the identification and invitation of suitable workshop participants as well as assist in its planning, coordination, and implementation. The PI's

partnership with MITH and Maryland iSchool's DCIC guarantees that the proposed planning workshop will be successfully coordinated. But more importantly, it will ensure that the findings and recommendations identified during this event, in both technical and social areas, will be realized.

4.2 Budget

Funding requested for this project is estimated at $50,000.00 to cover the cost of organizing and hosting a workshop at the University of Maryland, College Park campus. The largest portion of anticipated costs is allocated for transportation, food and accommodation of 20 invited workshop participants coming from beyond the Washington, D.C. region.

5. Diversity Plan

The PI will ensure that members of underrepresented communities are invited to the workshop and that the topics covered in the workshop's agenda/program will reflect and address the needs of diverse populations. It is important to ensure that the staff working in this project reflects the background of diverse communities. Thus, the PI will give special attention to recruiting a student from an underrepresented population to work on the project. The University of Maryland College of Information Studies has a diverse student population and is a pioneer in offering a specialization purposely designed to cater to the information needs of diverse populations. Specifically, we will seek to include Native American participants, both researchers and community members, who will provide critical input on access concerns to ethnographic collections held by many of the participating institutions. The project will complement and contribute to an already established infrastructure of diversity and inclusion within the College.

6. Communications Plan [Not required for Planning Grants]

7. Sustainability [Not required for Planning Grants, National Forum Grants, or Research Grants]

8. Summary

The goal of this planning workshop is to bring together experts from various communities capable of identifying concrete ways to productively work together to chart future steps to achieve virtual reunification of digitized photographic collections. This project will provide critical leadership for improving preservation of and access to archival visual materials in the digital realm, and in so doing it will advance access efforts to enhance national digital collections by addressing ongoing challenges to the representation and use of digital photographic materials. The project will not only bring in crucial technical perspectives and knowledge of extant capabilities in digital image processing, which will improve institutional understandings and capacity to create better tools for reunifying dispersed collections from large-scale aggregations, but it will also increase value for researchers by incorporating needs of users to improve access based on community needs.

Virtual Reunification of Dispersed Archival Photographs:
Developing a Collaborative Network and Framework
for Consolidating Fragmented Image Collections
University of Maryland, College Park

Project Timeline

This project will be implemented in three stages: planning and preparation, the workshop itself, and coordination and facilitation of targets and goals.

Stage	Duration	Activities
Planning and Preparation	October – November 2015	• Develop a final shortlist of invitees • Create workshop agenda (including dates and venue) • Design and publish project website
	December 2015	• Send out invitations to prospective participants
Workshop	May 2016	• Update project website to reflect final participant and agenda • Workshop proper
Coordination and Facilitation	May 2016 – June 2016	• Update project website to reflect future projects and expected outcomes

DIGITAL STEWARDSHIP SUPPLEMENTARY INFORMATION FORM

Introduction:
IMLS is committed to expanding public access to IMLS-funded research, data and other digital products: the assets you create with IMLS funding require careful stewardship to protect and enhance their value. They should be freely and readily available for use and re-use by libraries, archives, museums and the public. Applying these principles to the development of digital products is not straightforward; because technology is dynamic and because we do not want to inhibit innovation, IMLS does not want to prescribe set standards and best practices that would certainly become quickly outdated. Instead, IMLS defines the outcomes your projects should achieve in a series of questions; your answers are used by IMLS staff and by expert peer reviewers to evaluate your proposal; and they will play a critical role in determining whether your grant will be funded. Together, your answers will comprise the basis for a work plan for your project, as they will address all the major components of the development process.

Instructions:
If you propose to create any type of digital product as part of your proposal, you must complete this form. IMLS defines digital products very broadly. If you are developing anything through the use of information technology – e.g., digital collections, web resources, metadata, software, data– you should assume that you need to complete this form.

Please indicate which of the following digital products you will create or collect during your project.
Check all that apply:

	Every proposal creating a digital product should complete	Part I
	If your project will create or collect	**Then you should complete**
☑	Digital content	Part II
☐	New software tools or applications	Part III
☐	A digital research dataset	Part IV

PART I.

A. Copyright and Intellectual Property Rights

We expect applicants to make federally funded work products widely available and usable through strategies such as publishing in open-access journals, depositing works in institutional or discipline-based repositories, and using non-restrictive licenses such as a Creative Commons license.

A.1 What will be the copyright or intellectual property status of the content you intend to create? Will you assign a Creative Commons license to the content? If so, which license will it be? http://us.creativecommons.org/

All materials will be assigned a Creative Commons License CC BY-NC 3.0, which allows for sharing and reuse of the materials under conditions of attribution and non-commercial use, unless otherwise noted. We make a commitment to publishing any articles under this open license, which is currently the standard for many professional journals, such as the American Archivist.

A.2 What ownership rights will your organization assert over the new digital content, and what conditions will you impose on access and use? Explain any terms of access and conditions of use, why they are justifiable, and how you will notify potential users of the digital resources.

The University of Maryland (UMD) will not assert ownership rights of the digital products, specifically the project website, research outputs, and intellectual content. Any published materials will be copyright by the authors but will be made openly and freely available online (through the UMD institutional repository, DRUM http://drum.lib.umd.edu/), but UMD will not assert ownership rights to research products or intellectual content. The project website will be hosted by the UMD, which will require permission from UMD to host and serve content, but the UMD does not assert ownership over the research products.

A.3 Will you create any content or products which may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities? If so, please describe the issues and how you plan to address them.

No. Data containing personal information of workshop participants would not be made public without first obtaining informed consent and permissions from participants; if consent is given, such data would be anonymized and made available under the same CC BY-NC 3.0 license unless restricted according to participants' requests.

Part II: Projects Creating Digital Content

A. Creating New Digital Content

A.1 Describe the digital content you will create and the quantities of each type and format you will use.

The project website will be created using WordPress. Reports, white papers, or research articles pertaining to the workshop shall be uploaded to the site and may be downloadable in .pdf format. Any photographs uploaded to the site shall be in .jpg.

A.2 List the equipment and software that you will use to create the content or the name of the service provider who will perform the work.

N/A

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to create, along with the relevant information on the appropriate quality standards (e.g., resolution, sampling rate, pixel dimensions).

N/A

B. Digital Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

N/A

B.2 Describe your plan for preserving and maintaining digital assets during and after the grant period (e.g., storage systems, shared repositories, technical documentation, migration planning, commitment of organizational funding for these purposes). Please note: Storage and publication after the end of the grant period may be an allowable cost.

The project website shall be maintained during the active life of the project. The site will not contain unique information. Project reports, white papers, and research article related to the project will be made openly and freely available online (through the UMD institutional repository, DRUM http://drum.lib.umd.edu/).

C. Metadata

C.1 Describe how you will produce metadata (e.g., technical, descriptive, administrative, preservation). Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

N/A

C.2 Explain your strategy for preserving and maintaining metadata created and/or collected during your project and after the grant period.

N/A

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content created during your project (e.g., an Advanced Programming Interface, contributions to the DPLA or other support to allow batch queries and retrieval of metadata).

N/A

D. Access and Use

D.1 Describe how you will make the digital content available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

Project reports, white papers, and research article related to the project will be made openly and freely available online (through the UMD institutional repository, DRUM http://drum.lib.umd.edu/).

D.2 Provide URL(s) for any examples of previous digital collections or content your organization has created.

aeri2015.umd.edu; vos.umd.edu

Part III. Projects Creating New Software Tools or Applications

A. General Information

A.1 Describe the software tool or electronic system you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) the system or tool will serve.

A.2 List other existing digital tools that wholly or partially perform the same functions, and explain how the tool or system you will create is different.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your new digital content.

B.2 Describe how the intended software or system will extend or interoperate with other existing software applications or systems.

B.3 Describe any underlying additional software or system dependencies necessary to run the new software or system you will create.

B.4 Describe the processes you will use for development documentation and for maintaining and updating technical documentation for users of the software or system.

B.5 Provide URL(s) for examples of any previous software tools or systems your organization has created.

C. Access and Use

C.1 We expect applicants seeking federal funds for software or system development to develop and release these products as open source software. What ownership rights will your organization assert over the new software or system, and what conditions will you impose on the access and use of this product? Explain any terms of access and conditions of use, why these terms or conditions are justifiable, and how you will notify potential users of the software or system.

C.2 Describe how you will make the software or system available to the public and/or its intended users.

Part IV. Projects Creating Research Data

1. Summarize the intended purpose of the research, the type of data to be collected or generated, the method for collection or generation, the approximate dates or frequency when the data will be generated or collected, and the intended use of the data collected.

2. Does the proposed research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity already been approved? If not, what is your plan for securing approval?

3. Will you collect any personally identifiable information (PII) about individuals or proprietary information about organizations? If so, detail the specific steps you will take to protect such information while you prepare the research data files for public release (e.g. data anonymization, suppression of personally identifiable information, synthetic data).

4. If you will collect additional documentation such as consent agreements along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

5. What will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

6. What documentation will you capture or create along with the dataset(s)? What standards or schema will you use? Where will the documentation be stored, and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

7. What is the plan for archiving, managing, and disseminating data after the completion of research activity?

8. Identify where you will be publicly depositing dataset(s):

Name of repository: _____

URL: _____

9. When and how frequently will you review this data management plan? How will the implementation be monitored?

Virtual Reunification of Dispersed Archival Photographs:
Developing a Collaborative Network and Framework
for Consolidating Fragmented Image Collections

University of Maryland, College Park

Bibliography

American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences, *Our Cultural Commonwealth: The Report of the ACLS Commission on Cyberinfrastructure for the Humanities and Social Sciences* (American Council on Learned Societies, 2006), available online http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf.

Austenfeld, Anne Marie, "Virtual Reunification as the Future of 'Codices Dispersi': Practices and Standards Developed by e-codices—Virtual Manuscript Library of Switzerland," *IFLA* 36(2) (2010): p. 145-154.

Bartlett, Nancy, "Diplomatics for Photographic Images: Academic Exoticism?," *American Archivist* 59 (Fall 1996): p. 486-494.

Benson, Allen C., "The Archival Photograph and Its Meaning: Formalisms for Modeling Images," *Journal of Archival Organization* 7 (2009): p. 148-187.

Conway, Paul, "Digital Transformations and the Archival Nature of Surrogates," Archival Science 14 (2014). DOI: 10.1007/s10502-014-9219-z

Conway, Paul and Ricardo L. Punzalan, "Fields of Vision: Toward a New Theory of Visual Literacy for Digitized Archival Photographs." Archivaria 71 (Spring 2011): 63-97.

Deegan, Marilyn and Simon Tanner, *Digital Futures: Strategies for the Information Age* (London: Library Association Publishing, 2002).

Henschke, Ekkehard, "Digitizing the Hand Painted Bible: The Codex Sinaiticus, Its History and Modern Presentation," *Libri* 57 (2007): p. 45-51.

Lynch, Clifford A. "Special Collections at the Cusp of the Digital Age: A Credo," *Research Library Issues: A Bi-Monthly Report from ARL, CNI, and SPARC* 267 (December 2009). (Available: http://www.arl.org/resources/pubs/rli/archive/rli267.shtml).

Mifflin, Jeffrey, "Visual Archives in Perspectives: Enlarging on Historical Medical Photographs," *American Archivist* 70 (Spring/Summer 2007): p. 32-69.

Moretti, Franco, *Distant Reading* (Verso, 2013)

The Open Archives Initiative Protocol for Metadata Harvesting 2015 (Available: http://www.openarchives.org/OAI/openarchivesprotocol.html)

Punzalan, Ricardo L., "Archival Diasporas: A Framework for Understanding the Complexities and Challenges of Dispersed Photographic Collections," *American Archivist* 77(2) (Fall/Winter 2014): 326-349.

Punzalan, Ricardo L., "Understanding Virtual Reunification," *Library Quarterly* 84(3) (July 2014): 294-323

Schlak, Tim, "Framing Photographs, Denying Archives: The Difficulty on Focusing on Archival Photographs," *Archival Science* 8 (2008); p. 85-101.

Schwartz, Joan M., "Coming to Terms with Photographs: Descriptive Standards, Linguistic 'Othering' and the Margins of Archivy," *Archivaria* 54 (2000): p. 142-171.

Shenton, Helen. "Virtual Reunification, Virtual Preservation and Enhanced Conservation," *Alexandria* 21(2) (2009): p. 33-45.

Stewart, Brian, "Getting the Picture: An Exploratory Study of Current Indexing Practices in Providing Subject Access to Historic Photographs," *The Canadian Journal of Information and Library Science* 34(3) (2010): p. 297-327.

Unsworth, John, "Virtual Reunification of Scattered Archives," *Symposium on Manuscript Matters, British Library* (London, October 20, 2006). Retrieved from people.lis.illinois.edu/~unsworth/bl.06.ppt.

Voss, Jon, "Radically Open Cultural Heritage Data on the Web," *Museums and the Web*, San Diego, August 2012. (Available: http://www.museumsandtheweb.com/mw2012/papers/radically_open_cultural_heritage_data_on_the_w.html)

Original Preliminary Proposal

Virtual Reunification of Dispersed Archival Photographs:
Adding Value to Large-Scale Digital Aggregations
University of Maryland, College Park

Summary
The University of Maryland's College of Information Studies, in partnership with the Maryland Institute for Technology in the Humanities (MITH), seeks $50,000 for a Collaborative Planning Grant under the National Digital Platform funding priority to support a meeting of archivists, technologists, humanists, and other stakeholders that will develop specific plans for incorporating virtual reunification of dispersed photograph and image collections as an additional social and technological service atop the basic aggregations provided by national-scale digital platforms.

Statement of Need
Archival photographs often appear in multiple copies, versions, or formats. Photographs of the same provenance are often found in various locations or shared amongst several institutions. Format diversity, duplication, and dispersion pose deep challenges for heritage professionals and administrators attempting to represent photographic images scattered across many institutions. Moreover, scattered collections unduly burden researchers wishing to access images of common origin. Aggregation of materials from multiple institutions in new national digital platforms is valuable but aggregation cannot meet all of these challenges. The Digital Public Library of America (DPLA), for instance, now provides a single point of access to millions of cultural heritage items. However, DPLA lacks capability to identify similar items or bring together duplicates that can hinder users' capacity for insightful discovery.

Virtual reunification is the strategy of producing a consolidated, digitized representation of scattered artifacts, literary and artistic works, and/or archival records of a single origin or common provenance. The proposed planning grant will respond to some of the most difficult remaining challenges of developing aggregations into virtually reunified collections. These challenges include: What does it take to provide consolidated access to dispersed photographic collections? How can we harness developments in digitization, linked open data, pattern recognition, and computer vision to assist cultural heritage professionals to better understand and represent a collection that they share in common? What would consolidated representation for such collections provide for researchers?

Project Plan
The project will organize a two-day workshop to be held in May 2016 at the University of Maryland, College Park campus. This workshop will seek to join constituencies that rarely cross paths: representatives of cultural heritage institutions sharing a common dispersed collection, developers with expertise in computer vision and pattern recognition, digital humanists, scholars of visual archives, as well as representatives of large-scale digital aggregations. Participants will be asked to: 1) identify concerns, challenges, and barriers in providing a consolidated representation of a dispersed image collection in the context of national-scale digital platforms; 2) explore the potential of virtual reunification within such aggregations as a useful tool for both researchers and owning institutions; 3) identify additional stakeholders and partners; and 4) devise a plan to develop both an active community for and technologies to support virtual reunification of image collections as part of aggregation services delivered by national platforms for cultural heritage materials.

To produce tangible and useful outcomes, the planning meeting will ground itself in a case study of ethnographic photographs as starting point for addressing the large and complex topic of reunification of image archives. The ethnographic photographs of Dean C. Worcester, an early-twentieth-century American colonial administrator in the Philippines, are currently dispersed among at least 11 institutions in the U.S. and Europe. The ethnographic content involves Indigenous communities, which necessitates careful attention to cultural sensitivity issues. The photographs themselves document some of the earliest American anthropological explorations in the Philippines produced from 1890 to 1913. Although many of the images have been digitized, the digital collection is currently not linked, yet various sectors have expressed enthusiasm to do so: researchers,

Indigenous source communities in the Philippines, as well as the owning institutions themselves. In other words, this collection represents both opportunities and limitations faced by other collections with the same context and nature. The strategy of this project is to tackle one complex case to best mirror the realities and limitations of other, similar dispersed collections.

Relevance to National Digital Platform and Impact
Many projects have explored the possibilities of improving interpretation and context by bringing together dispersed collections using digital surrogates. Projects such as the Walt Whitman Archive, the Rosetti Archive, the *Codex Sinaiticus*, and the William Blake Archive demonstrate the value of digital projects to reunify works for researchers. However, this capacity has primarily been focused on textual materials and has never been explored on digitized archival photographs.

This project will identify steps to overcome some of the most difficult challenges faced by cultural heritage institutions, including better ways of presenting digitized archival photographs, supporting virtual reunification as an inter-institutional endeavor, incentivizing greater cooperation among owning institutions, and developing the capacity for photographic digital reunification. Creating greater capacity to combine dispersed collections will help institutions to better see links and materials complementary to their holdings. Moreover, this work would allow the institutions to complete and improve contextual information about their collections, which may be of use to researchers and collections managers. By automating recognition of duplicate images across collections, and other internal patterns, as well as displaying this information to users, national-scale aggregations can provide more valuable context to users.

Projected Performance Goals and Outcomes
This project would provide tangible and impactful contributions to both the scholarly research and cultural heritage communities by understanding the characteristics of successful virtual reunification for photographic collections with ethnographic content and in developing a strategy and identifying the necessary technical capability for photographic virtual reunification.
Its contributions in practice will be to initiate steps that lead to the development of community-driven services for photographic virtual reunification, which can be used to identify and reunify other, similarly dispersed image collections. Such a platform in turn can assist cultural heritage institutions in understanding and representing dispersed collections, including tracing provenance and original order for collections with duplication and metadata variation. This improved context and collection information for digital collections would also add an incentive for institutions to participate in efforts of linking cultural heritage collections online.

Project Director and Partners
Ricardo Punzalan, assistant professor of archives and digital curation at the University of Maryland College of Information Studies, will serve as project director and PI. His doctoral dissertation identified barriers and challenges of virtual reunification as a strategy for representing dispersed collections of ethnographic archival photographs, with a focus on materials attributed to Worcester. He has recently published articles on virtual reunification and dispersed photographic collections in *Library Quarterly* and *American Archivist*.

Two units at the University of Maryland, with capacity and expertise for technical development and digital humanities, will serve as project partners. The Maryland Institute for Technology in the Humanities (MITH) as well as the Maryland iSchool's Digital Curation Lab will provide technical expertise in digital projects and programming. The PI, in partnership with MITH and the iSchool's Digital Curation Lab, will plan, organize, and host the proposed planning workshop.

Estimated Budget
Funding requested for this project is estimated at $50,000.00 to cover the cost of organizing and hosting a workshop at the University of Maryland, College Park campus. The largest portion of anticipated costs are for participant travel to the workshop.