

The Grey Circuit: From Social Networking to Wealth Creation
13th International Conference on Grey Literature
Library of Congress, Washington DC
December 5-6, 2011

Title of Paper:

From Unprocessed Unknowns to Digital Diffusion: Uncovering Grey Literature at the Center for the History of Psychology

Topics Suited to Paper:

Subject-based Communities
Special Collections
Metadata

Authors Names, Affiliations, Contacts:

*corresponding author

Jodi Kearns*
Digital Projects Manager
Center for the History of Psychology
The University of Akron
Akron, OH 44325-4302
330.972.7952
jkearns@uakron.edu

Cathy Faye
Assistant Director
Center for the History of Psychology
The University of Akron
Akron, OH 44325-4302
330.972. 6096
cfaye@uakron.edu

Lynn Willis
Content Development Manager, PsycINFO
The American Psychological Association
750 First Street, NE
Washington, DC 20002
202-336-5689
lwillis@apa.org

Abstract

Collection Description: The Center for the History of Psychology (CHP) at The University of Akron is home to a special collection of grey literature that documents the history of psychology and related human sciences. The collection is comprised of 4,000 paper-and-pencil tests; 15,000 organizational newsletters, white papers, conference programs and proceedings, and membership directories; and 8,000 US military technical reports. These data, covering a broad variety of contemporary and historical topics, are an untapped but immensely useful resource for historical research, as well as contemporary meta-analytic studies and general literature reviews.

Problem/Goal: The central goal of CHP was to find a way to make these unknown materials available to scholars from a variety of fields. Digitization and mass distribution was the clearest solution. However, CHP lacked financial resources necessary for such a large project. CHP therefore partnered with the American Psychological Association (APA) to uncover, process, digitize, and provide access to its largely unknown grey literature documents.

Method/Procedure: A CHP librarian was hired to manage the project. Digitization was conducted onsite using rented equipment. Scanned images were uploaded to APA server for processing and assignment of metadata in order to provide access to the digital documents via PsycEXTRA, a searchable, digital repository widely used in the psychology community. CHP processes included archival arrangement, organization, and processing.

Costs: APA contracted with vendor Kirtas for onsite digitization. Costs were split between the grey literature project and a concurrent book scanning project. Two scanners were rented and two full-time technicians were hired to scan and then process files at a remote location. PDF files were then accessioned into PsycEXTRA and all files were investigated for copyright restrictions. First to be released in PsycEXTRA were files in the public domain and files for which agreements had been reached with copyright owners. APA indexed and released documents in PsycEXTRA and to CHP for onsite use. Documents remain held in a dark archive until permission is acquired from copyright owners for release.

Results/Case Study: The CHP/APA scanning project represents an attempt to build a partnership between subject-specific special collections and the disciplines with which they are most closely aligned. This partnership has proven to be a very successful one, resulting in the digitization and distribution of 20,000 documents from the CHP Grey Literature collection. The documents will be available onsite at the CHP, through the PsycEXTRA database, and on the CHP website.

The paper follows.

*From Unprocessed Unknowns to Digital Diffusion: Uncovering Grey Literature at the Center
for the History of Psychology*

The Center for the History of Psychology (CHP) at The University of Akron is home to a special collection of grey literature that documents the history of psychology and related human sciences, such as philosophy, biology, and anthropology. The collection comprises 6,000 paper-and-pencil tests; 15,000 organizational newsletters, white papers, conference programs and proceedings, and membership directories; and 8,000 US military technical reports. These data, covering a broad variety of contemporary and historical topics, are an untapped but immensely useful resource for historical research, as well as contemporary meta-analytic studies and general literature reviews. The collection grows daily through donations from psychologists and organizations, and nearly 2,000 paper-and-pencil tests, for example, have been added to the grey literature collection since the original abstract was submitted.

The Problems and Goal

The central goal of the CHP was to find a way to make a massive collection of unknown materials available to scholars and researchers from a variety of fields. Though digitization and distribution were an obvious solution, the CHP lacked personnel and financial resources necessary for such a massive project. A more traditional approach of finding grants to fund a project for a year was dismissed in favor of initiating and sustaining a relationship with a partner that had the capability to diffuse the content broadly by publishing rich “discovery tools” (Willet, 2009). The existing basic donor-repository relationship between the CHP and the American Psychological Association (APA) was developed into a working partnership where both

institutions found benefit: the CHP hired a certified librarian specializing in digital projects to manage the digitization initiatives, and APA received a mass store followed by an ongoing trickle of high quality images of documents to populate research sources for professionals and scholars of psychology and related human sciences. Such a symbiotic relationship permitted the CHP librarian to uncover, process, digitize, and provide access to its largely unknown grey literature documents that may have otherwise spent another several years in unprocessed collections of boxes.

A quick scan across literature discussing collection, description, and access to grey literature shows some common themes: that there is a lot of grey literature, it's valuable research material, and it takes a lot of time to adequately describe in an institution's system of retrieval (see Aina, 2000; Sulouff et al, 2005; Ranger, 2005). The grey literature collection at the CHP is plentiful and considered by patrons and donors tremendously valuable to the historical record, and, by the librarians and archivists, daunting by way of indexing, organization, and access. Further, the staff of the CHP recognizes that the scholarly appeal of the grey literature collection may be limited to a particular audience of researchers in the field of psychology, populous though this audience may be. The CHP itself is the largest collection of its kind in the world, and the only collection of its kind in the United States. The grey literature comprises only a small portion of the collections that also include manuscripts, artifacts, photographs, books, and film. With a global, though specialized, audience requesting access to dusty documents still in the boxes sent by the donor, the staff of the CHP recognized urgency in generating quick and complete availability and dissemination of this one-of-a-kind collection for a specialized, international patronage. This urgency was addressed in the partnership between the CHP and APA.

The Collections

The CHP's grey literature collection holds national and international military technical reports; organizational newsletters and circulars; conference proceedings, programs, and announcements; advertisements of new books and lectures; and other grey literature documents, but this collection was mostly unknown until the CHP/APA partnership. Presented with 85 boxes from the Saul Sells collection, from donor and CHP Board Member Dr. Ludy Benjamin, the Digital Projects Manager with a team of students unboxed and organized what turned out to be approximately 20,000 Air Force, Navy, and Army technical reports and studies ranging from the 1910s to the 1990s covering topics in survival, auditory and visual perception, medical experimentation, nutrition, statistical processes, and psychological testing. The Sells collection also includes newsletters, circulars, conference programs and proceedings, directories, and advertisements from assorted national and international organizations and psychological societies, special interest groups, and government offices. Documents in this grey literature collection were published from the 1880s until the present. Pulling also from processed and unprocessed collections from additional donors, the digitization team eventually queued 15,000 grey literature documents for digitization. Not every document that was uncovered qualified for digitization because the Digital Projects Manager was legally required (to the best of her knowledge and capability) to adhere to both Title 17 of the US Code and a selection policy written by APA. Documents were selected for digitization based on contracts made between APA and Content Owners regarding access through PsycNET and continued rights to the original work.

The Importance of the Collection to the Research Community

Though most contemporary psychologists rely on peer-reviewed, published literature, grey literature has tremendous potential as a useful resource for psychologists, including the more than 100,000 members of the American Psychological Association. Grey literature is being successfully used in a variety of disciplines to study a vast array of topics such as famine in Africa (von Braun, Teklu, & Webb, 1999), antibiotic use in low-income countries (Radyowijati & Haak, 2003), and the use of medical simulations in teaching (Issenberg, McGaghie, Petrusa, Gordon, & Scalese, 2005). Searches of the literature suggest that grey literature is rarely used among psychologists, though this perhaps reflects the simple fact that it is difficult to locate and obtain. Prior to the implementation of the PsycEXTRA database, there was no unified database for grey literature in the field of psychology. Scholars seeking such literature would have to search miscellaneous databases for conference proceedings, theses, and technical reports and then contact authors to obtain the text (Faragher, Cass, & Cooper, 2005).

Scholars have suggested that grey literature is particularly useful for disciplines such as psychology that make use of meta-analysis, a statistical tool for combining the results of several research studies on a given topic (Conn, Valentine, Cooper, Harris, & Rantz, 2003). Meta-analyses and general literature reviews usually involve combing the published literature for studies on a given topic, and then reviewing and/or statistically combining the results from those published studies to arrive at a tentative conclusion regarding the state of research on that particular topic. Such studies are prevalent and useful in the field of psychology and often serve as the starting point for a new research study on a given topic. However, it has been suggested that because these review studies rely primarily on published findings, they suffer from a publication bias, where studies lacking positive results are less likely to be published. This

problem is thought to lead to an overestimation of the magnitude of a relationship or scientific finding (Thornton & Lee, 2000). Because grey literature represents the unpublished scientific record, it may help to ameliorate this bias and make such reviews more comprehensive and more representative of the field. The CHP's holdings, and specifically its large collection of technical reports, are a particularly useful and previously unknown source of data on everything from workplace satisfaction to the physiological and psychological effects of prolonged isolation.

The grey literature collection at the CHP is also a tremendous resource for historians of psychology and related sciences. The test collections document more than 100 years of changing practices in the measurement of aptitudes, personality, abilities, and preferences. The collection of military documents details the long and sometimes complicated relationship between the military and social scientists. Conference proceedings, reports, and agendas all document the ways in which particular aspects of these sciences grew and changed over time. Access to these documents would therefore be indispensable in helping historians trace the development of the field in a variety of ways.

Document Selection

APA staff visited the CHP to triage the collection for digitization. First priority was assigned to documents published by organizations with which PsycEXTRA already had an ongoing agreement. PsycEXTRA partners with nonprofit organizations, educational and research institutions, and for-profit companies to archive, index, and distribute their grey literature to the academic community. Some organizations included in this effort are Psychonomic Society, Society for the Psychological Study of Ethnic Minority Issues, US Air Force, Midwestern Psychological Association, Society for Industrial and Organizational Psychology, Southwestern

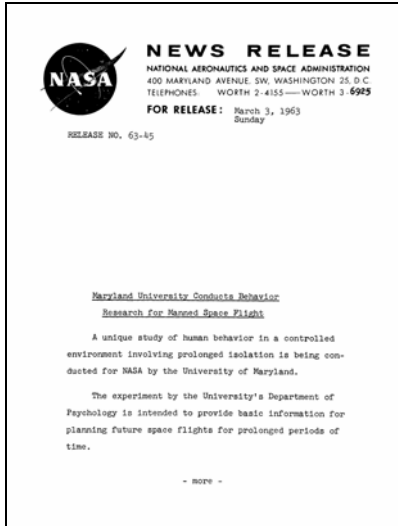
Psychological Association, Society for the Teaching of Psychology, Rocky Mountain Psychological Association, National Institute of Mental Health, Southeastern Psychological Association, Society for General Psychology, Federal Aviation Administration, US Navy, Western Psychological Association, Stockholm University, and National Institute on Drug Abuse.

Second priority went to public domain documents that filled a gap in PsycEXTRA holdings. Broad assumptions about current holdings were used to choose material rather than matching each document against the database, and as a result, some duplication occurred. Finally, documents published by organizations with strong ties to APA were chosen. The promise of archival documents scanned at no charge to these organizations was, in many instances, persuasive reason to partner with PsycEXTRA.

Below is shown a cross-section of organizations and societies whose collections were digitized with images from the CHP digital collections. [Aside: These highlights are going to appear to favor grey literature from the American Psychological Association, research departments of the United States military, and research laboratories of the United States government. The collection is largely composed of documents from these organizations, true, but the underrepresented organizations and societies may have granted permission under contract for us to include their grey literature documents in our digitization project, and subsequently in PsycEXTRA, but we have not obtained permission to necessarily include images in this document.]

Some Content Owners with Images from Mass Digitization Project

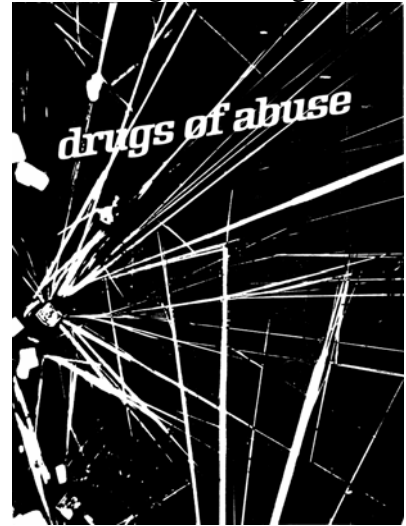
NASA



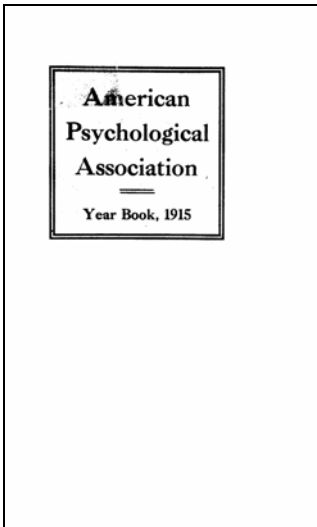
Southwestern Psychological Association



Bureau of Narcotics and Dangerous Drugs



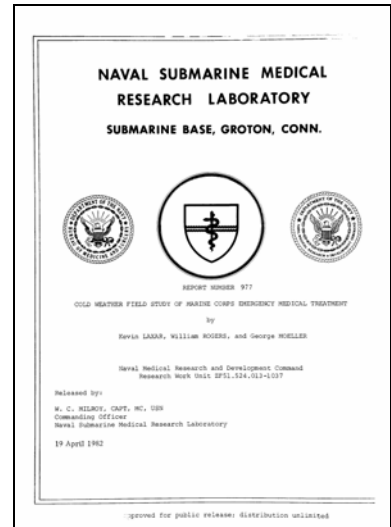
American Psychological Association



Rocky Mountain Psychological Association



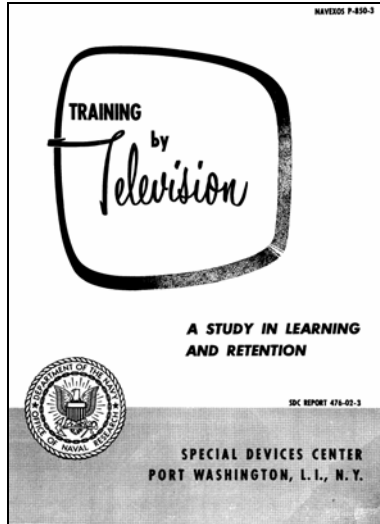
Naval Submarine Medical Research Laboratory



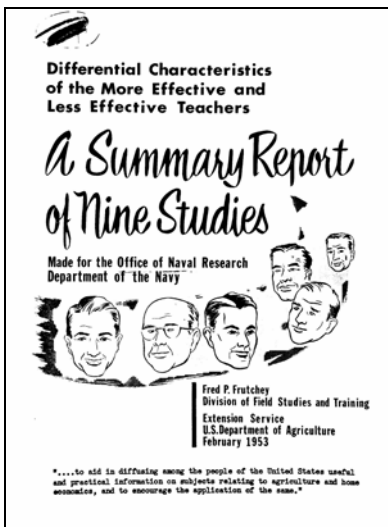
Special Devices Center

Works Progress Administration

National Institute of Mental Health

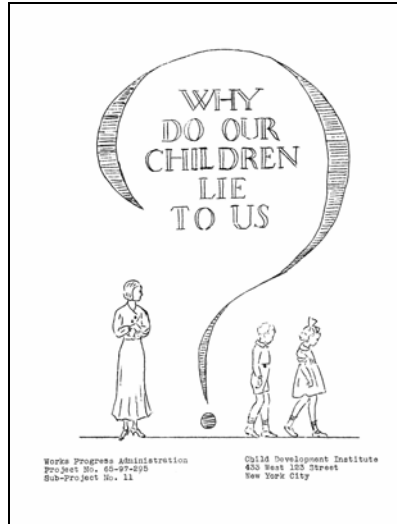


US Department of Agriculture



The Mass Digitization Project

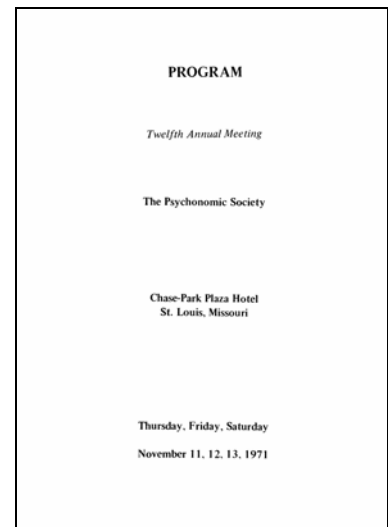
University of Akron and American Psychological Association legal teams prepared a contract to outline the relationship and responsibilities of both partners. For example, it was important to the CHP and to the University that, even while APA holds the right to distribute digitized documents to database subscribers, students, staff, and faculty had a right to unrestricted access to CHP documents in both physical and digital forms (Kaufman, 2007) and it



US Department of Labor
Children's Bureau



Psychonomic Society



was essential to both partners that careful regard was given to the rights of active copyright owners (Radice, 2009).

Digitization was conducted onsite at the Center for the History of Psychology (then, the Archives of the History of American Psychology) using two Kirtas book scanners that allowed rapid digitization each with a robotic arm for turning pages and two Canon EOS SLR cameras mounted above the document for grabbing images. APA rented Kirtas scanners and hired the technicians to run them. In phase one of the project, Kirtas technicians captured and manipulated images of document pages and delivered them to APA server. During phase two of the project, the Kirtas machines were replaced with a large-surface flatbed scanner for continued digital reproductions, and images continued to be uploaded by the CHP digitization team to APA server.

During the first phase, each document received a unique barcode by which it could be identified in the Kirtas system. At a later phase, the PsycEXTRA team would rename these files within its own system, while maintaining the original barcode as part of the metadata so the original documents could be easily located when necessary.

Three concerns arose in regards to working with hired technicians and rented machinery. Because the grey literature digitization project ran concurrent to a rare book digitization project, the same scanners captured the materials. First, much of the grey literature in the CHP collections is not in book form, so book scanners may not have always offered the most efficient digitization method. Second, common archival practices prohibit the use of automatic feeds for making copies of historical documents, and because of this, it was usually impossible to rely on the automatic features and robotic arms to turn fragile pages of document artifacts. Third, the technicians were experts in using their equipment for rapid mass digitization, though not

necessarily experts in careful archival practices, document handling, document organization, and appraisal of document condition and sustainability through the process (Teper, 2011).

Physical documents were organized four times as a result of the many necessary phases from discovery to repository. Those phases are: 1- initial organization of documents by institution, military branch, or government office in order to categorize unknown materials; 2- secondary alphabetical organization of publishing or sponsoring agencies within government or otherwise broad categories, like the Air Training Command, Air University, Human Resources Directorate, and School of Aviation Medicine (few of many laboratories within the Air Force publications); 3- tertiary organization of documents within categorized groups, usually by report numbers, years, volumes, or issues; (In this order, documents were assigned a barcode and queued for digitization); 4- final reorganization of scanned documents and placement in vertical files for on-site viewing.

The Costs of Mass Digitization

Costs were split between the grey literature project and a concurrent book scanning project. This drove down overall project costs due to economies of scale and the guarantee that machines would never be idle while material was being prepared for scanning.

APA contracted with Kirtas for onsite digitization. We rented high-end planetary scanners based on the needs of the book scanning project. Full-time technicians were hired and trained by Kirtas to operate the machines. High resolution image files were sent via FTP to a Kirtas facility where they were made into pdfs, optimized, and OCR'ed. Kirtas sent finished files to PsycEXTRA in Washington, DC, for use in the database.

Scanning began in March of the first year with two scanners and two technicians. In October of that year, we stepped down to one Kirtas scanner as the bulk of the work on the concurrent book project was complete. In June of the following year, we concluded our relationship with Kirtas and moved to a flatbed scanner procured by APA and operated by CHP staff.

Costs fell under three general categories: scanning machine rental, per-page scanning charges, and personnel. Rental costs were fixed per machine and steady throughout the project. The per-page scanning charge was more expensive for the grey literature than for books. The grey literature contained fewer images per file on average, increasing set-up and processing times. Because the automated page turner often could not be used, images were generated more slowly.

APA contracted the hiring and management of onsite staff to CHP. In addition to the project manager and scanning operators, students at the University of Akron were hired to help with retrieval and restacking material. Additional personnel were assigned to the PsycEXTRA database to accession and index the new material. Technical staff at University of Akron and APA were also called upon to deal with problems that crept up at the intersection of the various file handling systems.

The Resulting Benefit for APA: PsycEXTRA

PsycEXTRA staff match PDF files to content publishing organizations and investigate files for copyright restrictions. Those documents for which we have an agreement with the content owner, or which are determined to be in the public domain, are accessioned into

PsycEXTRA production system. Other documents are retained in a dark archive while staff reach out to the organizations for permission to reproduce the work in PsycEXTRA.

PsycEXTRA staff then create bibliographic records. These records, like all PsycEXTRA records, include abstracts, are indexed with controlled vocabulary, and classified using the same subject system as that used for PsycINFO. Each record for a CHP-sourced document includes the sentence: “This historic document is included through collaboration with The University of Akron, The Archives of the History of American Psychology, University Libraries.”

Currently almost 15,000 documents from CHP have been released into PsycEXTRA. Created in 2004, PsycEXTRA has about 230,000 records covering historical and current grey literature on subjects relevant to the behavioral science community. Inclusion in PsycEXTRA makes these documents available to individual and institutional subscribers throughout the world.

The Resulting Collections for the CHP

The CHP/APA scanning project represents an attempt to build a partnership between subject-specific special collections and the disciplines with which they are most closely aligned. This partnership has proven to be a very successful one, resulting in the digitization and distribution of 20,000 documents from the CHP Grey Literature collection. In addition to being available through PsycEXTRA, the documents are available to researchers onsite at the CHP and digitally to students, faculty, administration, and staff of the University of Akron, and up to 10% at any time as part of the CHP web-presence. Further, physical arrangement of documents on site at the CHP is now organized in a system of vertical files, currently using 60 sets of filing cabinets with 10 more on deck to accommodate the daily growth of the physical collection.

Access to both physical and digital versions of documents remains a priority for the CHP, because, to the archivist and to the historian, “the physical manifestations in which [writings] originally were expressed” (Radice, 2009) are often as important to the historical record as the contents of those documents.

The rapid mass digitization of the grey literature forced the digitization team to worry later about descriptive and bibliographic metadata in order to maintain a constant availability of documents in the scan queue while the equipment was still under the rental agreement. As a result, the CHP staff and researchers are now faced with a massive backlog of documents to index in its onsite system and collections—a recognizable dilemma in rapid and mass digitization projects (Gueguen, 2009), even without addressing legacy data for digital objects. Not all documents digitized are yet available in PsycEXTRA nor did all documents uncovered in the process qualify for digitization. CHP staff and students have created to date more than 50,000 records in local databases, thereby opening access to this impressive collection of grey literature. During the metadata creation phase, the PsycEXTRA team added the following byline to the abstract of each document, thereby enabling easy identification of the CHP materials in PsycEXTRA for the CHP community: “This historic document is included through collaboration with The University of Akron, The Archives of the History of American Psychology, University Libraries.”

Sustaining Relationships

At least three distinct relationships have been sustained as a result of the partnership: 1- the Center for the History of Psychology and the American Psychological Association; 2- the

donor and research community who supply and access previously unprocessed documents in CHP collections; 3- the broader research community and APA.

The CHP and APA continue to hold a professional relationship to digitize grey literature collections and to make the documents accessible via PsycEXTRA and other APA-published databases. Currently, the CHP Digitization Center is focused on selection and digitization of documents in the CHP Test Center: paper-and-pencil tests designed to measure intelligence, aptitude, mental and emotional capacity, occupational preferences, academic achievement, and perception. Years of publication of this historical test collection ranges from the 1890s to the present, though not all documents qualify for digitization under APA projects. Careful selection based on public domain and copyright registration renewal results in a stream of documents scanned for PsycTESTS , a new addition to the PsycINFO suite of databases. CHP Digital Projects do not, at this point, digitize materials under Section 108 of Title 17 of the United States Code, that is, to preserve content simply for the sake of generating an archival facsimile (U.S. Copyright Office, 2009) as part of the digital projects partnership with APA. PsycEXTRA is an example of APA's continuing commitment to provide a full range of information products to the psychology community. Archiving information in PsycEXTRA guarantees that these documents will have an electronic home for today's and tomorrow's researchers. This partnership has drawn attention to CHP collections and growing access to archival documents, which inevitably has increased visibility in the research community and strengthened relationships with donors and researchers across the globe. Separate from the CHP researchers can access these documents that have been integrated into the databases of the American Psychological Association to enrich access to historical documents in their collections.

Concluding Thoughts

The grey literature collection at the Center for the History of Psychology was the catalyst for a functional and continuing relationship with the American Psychological Association that produced a research database to diffuse otherwise unknown historical documents. The Center employs digitization experts and supporting students in order to continue document discovery and diffusion on the digital landscape that have increased visibility of collections. The benefits of the collaboration continue to place the right documents into the hands of historians, psychologists, and scholars around the globe.

References

- Aina, L. O. (2000). Grey literature and library and information studies: a global perspective, *International Journal on Grey Literature*, 1(4), 179-182.
- Conn, V., Valentine, J., Cooper, H., & Rantz, M. (2003). Grey literature in meta-analyses. *Nursing Research*, 52, 256-261.
- Faragher, E. B., Cass, M., Cooper, C. L. (2005). The relationship between job satisfaction and health: a meta-analysis, *Occup Environ Med*, 62, 105-112.
- Gueguen, G. and Hanlon, A. M. (2009). A collaborative workflow for the digitization of Unique Materials, *Journal of Academic Librarianship*, 35(5), 468-474, [doi:10.1016/j.acalib.2009.06.001](https://doi.org/10.1016/j.acalib.2009.06.001).
- Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee, G. D., and Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review, *Med Teach*, 27(1), 10-28.

- Kaufman, P. B. and Ubois, J. (2007) Good terms: Improving commercial-noncommercial partnerships for mass digitization, *D-Lib Magazine*, 13(11/12).
- Kirtas Technologies Automatic Book Digitization, <http://www.kirtas.com/bookscanners.php>
- Radice, A. (2009). Digital Debates, *Conference Proceedings from Web Wise, Institute of Museum and Library Services*: Washington, DC.
- Radyowijati, A. & Haak, H. (2003) Improving antibiotic use in low-income countries: an overview of evidence on determinants, *Social Science & Medicine*, 57(4), August 2003, 733-744, 10.1016/S0277-9536(02)00422-7.
- Ranger, S. L. (2005). Grey literature in special libraries: Access and use, *Publishing Research Quarterly*, Spring, 53-63.
- Sulouff, P., Bell, S., Briden, J., Frontz, S. and Marshall, A. (2005). Learning about grey literature by interviewing subject librarians: A study at the University of Rochester, *C&RL News*, July/August, 510-515.
- Teper, J. H. and Shaw, E. F. (2011). Planning for preservation during mass digitization projects, *portal: Libraries and the Academy*, 11(2), 717-739.
- Thornton, A., and Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, 53(2), 207-216. doi: 10.1016/S0895-4356(99)00161-4
- U.S. Copyright Office (2009). *Circular 21: Reproduction of copyrighted works by educators and librarians*, U.S. Copyright Office, <http://www.copyright.gov/circs/circ21.pdf>.
- von Braun, J., Teklu, T., & Webb, P. (1999). *Famine in Africa: Causes, responses, and prevention*. Baltimore, MD: Johns Hopkins University Press.
- Willett, P. (2009) Mass digitization and its impact on interlending and document supply, *Interlending & Document Supply*, 37(3), 143 – 148.