

Assistant Editor: Sarah Dorpinghaus, University of Kentucky. Contact Sarah at sarah.dorpinghaus@uky.edu if you would like to guest author a column or have a good idea to share.

Digitizing Modern Archival Collections: Addressing Complex Copyright and Privacy Issues

By Virginia Dressler and Cynthia Kristof, both at Kent State University Libraries

The article will describe a two-year grant project that was completed in 2018, particularly how copyright and privacy were addressed through the digitization of a modern archival collection. We will address incorporating a privacy review and will highlight how we dealt with copyright and permissions issues in archival collections.

The May 4th Collection is one of the largest archival collections held at Special Collections and Archives at Kent State University (KSU), with primary source materials totaling over 300 cubic feet represented in over 200 subcollections. The collection documents the events surrounding May 4, 1970, when the National Guard was called in as a result of a weekend of protests and unrest after the announcement of the US invasion of Cambodia.

The collection is open to the public and is used by researchers from all around the world. As the 50th anniversary of the event approaches, library staff at Kent State University Libraries have been amping up digitization efforts to provide open and free access to part of the May 4th Collection through its digital archive. In August 2016, University Libraries received a \$238,886 grant from the National Historical Publications and Records Commission (NHPRC) to digitize roughly 40 selected subcollections from the larger main May 4 archival collection.

Privacy Review

Addressing privacy in modern archival collections led to a lot of internal discussion on implementing a new review process early on in the grant project, particularly around unpublished materials. In the first digitized subcollection, our internal grant working group discovered the presence of Social Security numbers (SSN) in the form of student identification numbers within some materials. The working group was grateful to make this discovery before the materials were published in the digital repository, particularly as the printed format of the student identification number was not the same as the SSN. We discovered that the use of SSNs as student identification numbers had been a regular practice at KSU up until 2006.

A portion of the discussion around privacy review included personally identifiable information (PII) which is private information that could distinguish one person from

another, as well as other more nebulous forms of private information (for example, embedded information relating to religion, sexuality, etc.). Any information that can be used to distinguish one person from another and can be used for de-anonymizing anonymous data can be considered PII. This information includes proper names, aliases, identification numbers, addresses, date of birth, place of birth, race, religion, weight, activities, geographical indicators, employment information, medical information, education information, and financial information. The National Institute of Standards and Technology (NIST) "Guide to Protecting the Confidentiality of Personal Identifiable Information" is very helpful in identifying PII.

Concern over disclosing PII or other private information focuses on the harm or damage, including the potential for identity theft, embarrassment, or blackmail, that could arise by disseminating such information widely. Privacy breaches through dissemination of PII are hazardous to both individuals and organizations (who may be liable for the published information). The NIST guide also strongly recommends that institutions devise their own list of the kinds of information that should be protected. Additionally, certain laws such as FERPA and HIPPA provide some protections around certain types of information as they relate to educational or medical records, which lasts for the lifetime of the individual.

We implemented a new review process and workflow to identify potential PII and make redactions as needed before materials were published into the digital repository. We created a one-page guide for employees and students to refer to as they performed privacy reviews at the item level around common instances of PII (unique identification numbers, grade information, etc.).

Copyright

Often, digital projects bypass the copyright conundrum by focusing only works in the public domain or collections that otherwise do not raise copyright questions. For a modern digitization project, however, works potentially still under copyright could not be left out without omitting

(Continued on page 28)

(Continued from page 27)

essential portions of the collection useful to researchers and other information-seekers.

Kent State University clearly owns some materials in this collection, including official Kent State documents, meeting minutes, and other publications. These were digitized once privacy review was completed. For other materials, copyright status was difficult to determine. Many materials in this collection were unpublished and never registered for copyright. If we could determine and locate the author, creator, or copyright owner, we sought permissions. However, a great many materials were orphan works for which the copyright owner could be neither determined nor reached. For these materials, the project relied on fair-use evaluations.

The Association for Research Libraries' *Code of Best Practices in Fair Use for Academic and Research Libraries* (2012) proved helpful. The section titled "Creating digital collections of archival and special collections materials" (pp. 19–21) contains the following phrases, which rang true to our project:

- "published and unpublished"
- "copyright status is often unclear"
- "typically can be consulted only on-site"
- "unique assemblage or aggregation"
- "organized around a key topic, era, or theme"
- "Presenting these unique collections as a digital aggregate, especially with commentary, criticism, and other curation, can be highly transformative."

Roughly following the "Safe Digitization Workflow" flow-chart found on page 203 of Peter B. Hirtle's *Copyright and Cultural Institutions: Guidelines for Digitization for U.S. Libraries, Archives, and Museums* and utilizing an in-house adapted PDF version of Kenneth D. Crews and Dwayne K. Buttler's *Fair Use Checklist*, we developed a customized workflow and tracking system for copyright permissions and fair-use determination. The team often held lengthy discussions in biweekly copyright check-in meetings as to whether or not digitization might be considered a fair use, relying on the *ARL Code*, the *Fair Use Checklist*, and gut instinct as guidance.

At certain times, the team felt that although fair use could be invoked, maintaining good working relationships with local people and organizations was in the best interest of

everyone involved. Thus, our permissions contacts with copyright owners or creators were frequently less about copyright and more about communication. Just one example of this was the digitization of selected newspaper clippings from local newspapers. Many of these had been underlined and included with letters mailed to people associated with the May 4 events. Though this use seemed transformative, and we were digitizing a very small portion of any given newspaper, contacting them to ask for permission seemed like the considerate and prudent action.

Permission was most vigorously pursued when it came to digitizing highly creative works, such as art and poetry. An example of the latter was the digitization of photographs and paperwork associated with George Segal's *Abraham and Isaac* sculpture, originally commissioned as a May 4 memorial and later rejected by Kent State University administration. Segal's work is represented by Artist Rights Society (ARS). This was one of the easier permissions transactions, albeit more expensive than average.

We hope that as the 50th anniversary of the tragedy approaches, the digitized selections from the May 4th Collection prove to be a useful resource to researchers and the general public and that the combination of materials helps to provide some insight surrounding it. We also hope that results of the project help survivors, family, and friends to heal in some small measure.

Resources

1. Association of Research Libraries, *Code of Best Practices in Fair Use for Academic and Research Libraries*, 2012, <https://www.arl.org/code-of-best-practices-in-fair-use-for-academic-and-research-libraries>.
2. Kenneth D. Crews (formerly of Columbia University) and Dwayne K. Buttler (University of Louisville), *Fair Use Checklist*, <https://copyright.columbia.edu/basics/fair-use/fair-use-checklist.html>.
3. Peter B. Hirtle, Emily Hudson, and Andrew Kenyon, *Copyright and Cultural Institutions: Guidelines for Digitization for U.S. Libraries, Archives, and Museums* (Ithaca, NY: Cornell University Library, 2009), <https://hdl.handle.net/1813/14142>.
4. National Institute of Standards and Technology, *Guide to Protecting the Confidentiality of Personal Identifiable Information*, <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf>.