DATA PAPER

# Collection Objects from the Digital Occupy Archive

## Stephanie Becker[1], Heather McKee Hurwitz[2] and Anne Kumer[3]

[1] Freedman Center for Digital Scholarship, Kelvin Smith Library, Case Western Reserve University, US

[2] Department of Sociology, Case Western Reserve University, US

[3] Acquisitions & Metadata Services, Kelvin Smith Library, Case Western Reserve University, US

Corresponding author: Stephanie Becker (sxb891@case.edu)

The Occupy Archive's Collection Objects is a dataset of more than 400 digitized primary source materials from the global Occupy movement, openly available at DOI 10.17605/OSF.IO/6V9ZF. From 2019–2020, a team of research assistants, librarians, and Dr. Heather Hurwitz digitized over 1,200 pages of manuscripts and ephemera collected by Hurwitz at Occupy movement protests and meetings from 2011–2018. The team established a taxonomy of descriptive metadata and applied it to each object to facilitate searching, viewing, and downloading the data. Accompanying the dataset, the archive includes teaching tools and research guides to facilitate reuse.

## (1) Overview

**Repository location**

DOI 10.17605/OSF.IO/6V9ZF

**Context**

The Occupy Archive's Collection Objects originated as a physical archive of paper documents and Occupy movement ephemera collected by Hurwitz from 2011–2018. A team of research assistants digitized these primary source materials for broad reuse including the viewing and downloading of materials. Originally, Hurwitz collected media objects such as flyers, pamphlets, posters, postcards, art, and music for use in several research projects and her book, *Are We the 99%?: The Occupy Movement, Feminism, and Intersectionality* [5]. She conducted an ethnography of the main Occupy movement locations (New York City and the San Francisco Bay Area), engaged in in-depth interviews with participants, and explored many movement activities across the country (including the Occupy National Gathering in Philadelphia in 2012). The data collection was approved by the Institutional Review Boards[1] at the University of California Santa Barbara (Protocol Number: SOCL-TA-VA-050-6N) and Barnard College Columbia University (Project Title: Occupy Activism). In addition to examining the Occupy movement in general, to capture diverse experiences across the movement, Hurwitz's research was guided by the following research questions: How do Occupy movement participants build solidarity across gender, race, class, and sexual identities within the mass movement? Under what conditions do movement cultures exclude or alienate particular individuals and groups? How do gender, race, class, and sexuality processes influence contemporary social movements' dynamics and culture?

## (2) Method

As Hurwitz collected paper documents and movement ephemera (the collection objects), research assistants stored them chronologically in three-ring binders (see **Figure 1**). In 2019, Hurwitz formed a new project team to digitize the collection objects and upload digital surrogates to the Open Science Framework (OSF). OSF is an interdisciplinary platform for online access, storage, and preservation of data. As a team, we chose OSF because it does not require hosting fees or associating the project with a domain name. The platform is supported by Case Western Reserve University's (CWRU) Information Technology department, which provided ongoing user support for our team. We also chose OSF due to its technological capacities: several research team members could simultaneously upload scans, apply metadata, and add research and teaching documents into the archive's OSF site simultaneously from multiple computers. In order to manage the digitization workflow, project team members created folders in OSF corresponding to the organization of the physical binders. Each folder represents one of eight

**Figure 1:** Three-ring binders of objects that were later digitized.

physical binders that housed the collection, and each file is the scan or photographic reproduction of one collection object. There is also a folder called 'Contributions from the Occupy Community' that contains digital collection object donations from participants in the Occupy movement. This organization gives researchers a unique view of the collection's original order, which informed how the digitized dataset was formed.

### Steps

The project team met on a weekly basis to establish a strategic methodology for creating our dataset. We discussed and collaboratively made decisions on a workflow for systematically digitizing the collections, a platform for digital storage and access, and guiding overview documents for users. The collection digitization included developing a file naming system and scanning workflows. The storage and access platform included establishing an inventory spreadsheet, a folder structure in OSF, and a taxonomy for applying metadata. In addition to our dataset, we also included documents that describe the dataset creation process, research guides that list several objects for key subject themes, and a sample slide presentation and educational assignments to facilitate reuse among researchers and teachers using the archive.

### *File naming*

We conceptualized file names that would be unique and indicate the object's physical location in its respective binder. Each file name starts with OA for Occupy Archive, followed by the binder number, a short title to describe the object, and a number where the material is physically located in our digital filing system noted in our inventory spreadsheet. The format for the file names follow this

template: OAbinder#_title_object#. In OSF, the files are automatically organized into alphabetical order by file name. The file name's suffix indicates whether the file is a PDF or JPEG.

### *Digitization*

Research assistants scanned the collection in Kelvin Smith Library's (KSL) Freedman Center for Digital Scholarship with Epson Perfection 4600 and Epson Expression 11000 XL flatbed scanners at a resolution of 300 pixels per inch in 24-bit color [2]. If the object only had one side or page, it was scanned and uploaded as a JPEG. If the object had multiple pages or sides, it was scanned and uploaded as a PDF. We did not run optical character recognition on textual objects because OSF does not have indexing capabilities. For three-dimensional objects, we coordinated with KSL's Digitization Technician, Naomi Langer to photograph the objects in KSL's Digitization Lab. For small items, such as buttons and pins, Langer used a Phase One iXG 100MP camera. T-shirts were scanned on a Scan Master 0, an oversize flatbed scanner in the lab's annex space.

### *Inventory*

As research assistants digitized the objects, they filled out an inventory spreadsheet to keep track of their work and to provide a complete inventory of the dataset for researchers. The inventory is a spreadsheet included in the Project Documentation section of the archive. The spreadsheet provides details about each object, including the file name, descriptive title, creator/author, original date, material type, number of pages, scanning status, tagging status, and notes (see **Figure 2**). To facilitate the search, identify relevant objects, and reuse the data, users may download the spreadsheet, perform text searches within

the spreadsheet, filter results, and easily target and examine particular objects.

### Taxonomy

In order to describe and tag the digitized collection objects, the project team worked together to develop a controlled vocabulary of descriptive terms, along with definitions and rules for applying them, known as a taxonomy (see **Figure 3**). The taxonomy was created with special consideration paid to the descriptive needs of the digitized objects, research uses of the collection [1], descriptive terminology favored by the collection's creators and researchers (the activist community), and library cataloging best practices [6]. Group discussion led to six categories of descriptive metadata: format, associated movement(s), subject description terms, location, gathering type, and date. Using the methodology outlined in the finalized taxonomy, the research assistants cataloged each object in OSF by applying terms from the taxonomy as descriptive "tags". Since OSF does not provide the functionality of searching within a specific dataset, the collection tag "Occupy Archive" was also applied to each digitized object. This helps researchers see that the object they are looking at is in fact from this collection, as opposed to a different collection hosted on OSF. The taxonomy created a valuable way for researchers to successfully search for specific content within the dataset using OSF's API. Through team conversations with OSF affiliated staff, we were informed that collection searching is a feature which will be rolled out in a future version upgrade of OSF. Since we tagged our data using the taxonomy, the collection objects will become easier for user searching when the new feature is released.

### Sampling strategy

The dataset includes the entirety of the physical collection that Hurwitz gathered at Occupy events using convenience sampling. There was no sampling process completed during the data creation process. The digitized objects that make up the dataset were created by several authors for a variety of purposes and audiences. If the object was in circulation at an Occupy event, it was collected, stored in three-ring binders, and digitized for the Collection Objects component of the Occupy Archive.

### Quality control

After scanning all the objects in a binder, uploading the files to OSF, updating the inventory spreadsheet, and cataloging each object using our project taxonomy, the research assistants then traded binders to check each other's work. They looked for physical objects skipped during scanning, cropping errors that cut off content present on the physical object, inconsistencies between the inventory spreadsheet and what appeared in OSF, and correct usage of the taxonomy on each object. As the students worked through each binder, they flagged

| File Name | Descriptive Title | Creator/Author | Original Date | Material Type |
|---|---|---|---|---|
| OA01_FemGA_001 | Fem GA: Calling for Gender Equality in the Occupy Movement and the World | Women Occupy | 07/01/2012 | Postcard |
| OA01_makeoutnotwar_002 | Make Out Not War: CodePink.org | www.codepink.org | N/A | Sticker |
| OA01_separateandequal_003 | Separate and Equal? | Mujeres Libres / inourheartsNYC | 2012 | Pamphlet |

**Figure 2:** Example of a small part of the inventory spreadsheet with information about each object.

| Term | Scope Note | Editorial Note |
|---|---|---|
| Article | A piece of writing or printed content that is published, self-published, or unpulished. Published works are most often found in a magazine, journal, newspaper, or book excerpt. Unpublished or self-published include scholarly papers, essays, and other types of writing. Includes whole newspapers / magazines. | For a piece of writing posted on a blog, use the term Blog Post. |
| Audio recording / Music | Compact discs, song sheets, chant sheets. | For any type of audio recording including music and broadcasts, and for paper-based song and chant sheets. |
| Banner / Flag | An oversize piece with hand or machine made graphics and / or words on in. These are usually made of fabric. | For oversized paper-based graphics, use Posters. |
| Blog post | Printout of a piece of writing on a blog. | For articles published in online journals, magazines, and other publications, or scholarly papers, self-published and unpublised writing, use the term Article. |

**Figure 3:** Portion of the taxonomy for object format.

objects for a copyright review. While fair use requirements were met in the creation and sharing of the dataset, there were specific objects that posed privacy and/or copyright concerns. Students flagged the objects in question and Hurwitz reviewed them with KSL's Scholarly Communications and Copyright Librarian, Mark Clemente. For objects that included personal identifying information, a research assistant contacted the individuals for use permission or blurred out personal information if the individual could not be reached. There were a number of music CDs that included tracks from a variety of artists, which raised copyright concerns. For these objects, we logged the information in the inventory spreadsheet, but did not include a representative file in the dataset. Outside of individual object evaluation, we also wrote a copyright statement[2] to include with our dataset that underlines the educational fair use, the attribution of the content/author of objects in the inventory spreadsheet, and a take-down procedure should any user dispute the use of an object.

After completing our quality control process, we recruited library staff on KSL's Freedman Center for Digital Scholarship Team to test OSF's API using our tagging system (see **Figure 4**). We tested the search functionality with people who were not directly involved in creating the dataset in order to gain an understanding of how a user might interpret the documentation.

Once the Freedman Center staff validated the tagging and search methodology, we changed the permissions on our OSF site from restricted to openly available.

## (3) Dataset Description

**Object name**
Occupy Archive's Collection Objects.

**Format names and versions**
PDF and JPG.

**Creation dates**
2019-09-01–2020-04-01.

**Dataset creators**
Heather McKee Hurwitz, PhD (Conceptualization, methodology, funding acquisition, project administration, supervision, writing – original draft, writing – review and editing).
Stephanie Becker (Methodology, project administration, supervision, resources, writing – original draft, writing – review and editing).
Anne Kumer (Data curation, methodology, supervision, validation, writing – original draft, writing – review and editing).
Jason Choi (Investigation, visualization).
Kyle Jones (Investigation, visualization).
Zoe Nguyen (Investigation, visualization).
Riley Simko (Investigation, visualization).
Zoe Wang (Investigation, visualization).
Naomi Langer (Investigation, methodology).
Mark Clemente (Investigation, methodology).

**Language**
English with occasional Spanish.

**License**
Creative Commons Attribution 4.0 International Public License.

**Repository name**
Open Science Framework.



**Figure 4:** Sample of a search within the Occupy Archive in OSF using the metadata in the taxonomy to search the 'Occupy Archive' collection for objects cataloged with a specific format and year. In this example, the search reveals a list of objects tagged with the format 'Flyer' and year '2012'.

## (4) Reuse Potential

This dataset fills a need in history, media studies, social movements, and sociology to share online rapidly evolving activism. The collection objects capture a range of typically ephemeral social movement activity and records the media and cultural products of the Occupy movement. The data is publicly available online to other researchers, students, activists, and the general public to learn from the movement.

### Researcher reuse

The digital Occupy Archive not only includes the Collection Objects dataset of digitized primary source materials, but also provides a supplementary overview and documentation to inform researchers who will reuse the data. The workflow for digitization explains the methodology and data structure. The metadata taxonomy reveals the rationale of the search terms linked to objects and provides instructions for users to search the data. These materials will assist researchers to locate sets of objects in order to conduct original research and make meaning by examining the data. Researchers can also download the data and the associated inventory for digital scholarship projects such as visualizing the metadata or mapping objects to their locations using a geographic information system.

### Educator and student reuse

Students and educators may use the data to learn about social change, democracy, to organize digital exhibits, and create original teaching activities and assignments. Educators will find interest in the lecture slides and the lesson plan for a project assignment. Each of these teaching materials were test-run in Hurwitz's Spring 2020 classes at CWRU.

The slide presentation and lecture were designed for use in Introduction to Sociology classes, which often neglect learning about social movements and social change because textbooks lack updated information about contemporary movements. Yet, these movements are important to contemporary students. This lecture provides a way for students to "go into the field" and examine the Occupy movement through the media created by the movement. The team of research assistants designed the 50-minute Introduction to Sociology slide presentation and lecture.[3] The lecture provides an orientation to understanding contemporary social movements, explains basic social movement concepts, illustrates those concepts by featuring several digitized collection objects, and includes a series of multiple choice questions to stimulate student engagement. Students can raise their hands en masse to answer questions presented throughout the lecture, or answer using iClickers.[4]

The extended class activity/assignment to create a digital exhibit is designed for upper level sociology, politics, history, or American studies classes but may be tailored to students at other levels or in other types of classes. Digital exhibits are curated collections of archival objects annotated with contextualizing information. Students select primary documents, analyze the data, create webpages using Scalar,[5] and link their pages together to create a digital exhibit.

The assignment was designed for five 50-minute sessions at KSL. Digital Scholarship Librarian, Amanda Koziura, led the sessions and guided the students in using Scalar. They browsed and selected objects from the dataset, worked in small groups to link those objects with others of their classmates, and wrote object annotations with insights they gleaned from supplemental scholarly readings about the Occupy movement [3–5]. The activity outlines a hands-on learning experience about digital and contemporary activism. Students leave the class with the ability to critically evaluate social movement media and framing, as well as develop their own digital exhibits using Scalar. The Occupy Archive is a repository of contemporary social movement objects ripe for additional creative remote-learning assignments where students can analyze and contextualize the data.

### Public reuse

Activists and the general public may find additional uses for the dataset. Activist groups may analyze the ways Occupy movement participants framed messages and distributed educational materials. Artists and graphic designers may find inspiration in the visual materials. Historians will find the variety of data, especially primary source materials about the Occupy movement, useful to illustrate political issues, economic issues, and democratic participation in the 2010's.

The Occupy Archive's Collection Objects is an open source dataset that affords many options for reuse. We welcome reports from researchers, educators, and the public on ways they have reused the data.

### Notes

[1] The Institutional Review Board (IRB) at each institution is a committee of university and community members who review social and behavioral studies that include human subjects. IRB approval indicates that a researcher followed established protocols for the ethical involvement of and data collected from human participants.

[2] The Occupy Archive is made available for purposes of education, research, and scholarship. Materials in the Occupy Archive are either in the public domain, are made available by permission, or are permitted by an exception such as fair use. When possible, attribution has been given to rights holders; however, due to the nature of archival collections, we are not always able to identify this information. Please examine the Occupy Archive Inventory for attribution information. The Occupy Archive site is licensed under the Creative Commons Attribution Non-Commercial (CC-BY-NC) 4.0 License; however, individual materials in the Archive may be protected by copyright and/or related rights legislation. If you wish to use any work in the Occupy Archive beyond what is permitted by copyright and/or related rights legislation, you may need to obtain permission from the rights holder(s).

If you have questions or concerns about the use of specific works in this archive, please contact Heather Hurwitz: heather.hurwitz@gmail.com, hxm56@case.edu. To help us identify the work in question, please indicate: 1) your relationship to the copyright holder, including if you are the copyright holder, and 2) the works with which you are concerned, including a URL to the work in the Archive. We will respond promptly to such requests and shall remove the work(s) in question should it be determined that use in the Archive is not permitted.

[3] In addition to designing the lecture, the undergraduate research assistants gave the lecture to two different classes of 70 students each.

[4] iClickers are small remote-control devices. They stimulate student participation in large lecture halls. College and university instructors often require students to purchase and use iClickers. Instructors need iClicker software and instructor devices to use them in the classroom. See more information here: https://www.iclicker.com.

[5] Scalar is a free open source online tool that allows users to create web pages that combine images, files, other websites, and the user's own writing. For more information, see: https://scalar.me/anvc/scalar/.

## Additional Files
The additional files for this article can be found as follows:

- **Inventory Spreadsheet.** Full inventory spreadsheet of Occupy Archive's Collection Objects. DOI: https://doi.org/10.5334/johd.20.s1
- **Taxonomy.** Taxonomy for Occupy Archive's Collection Objects. DOI: https://doi.org/10.5334/johd.20.s2

## Competing Interests
The authors have no competing interests to declare.

## Author Contributions
- S.B. – Methodology, project administration, supervision, resources, writing – original draft, writing – review and editing.
- H.H. – Conceptualization, methodology, funding acquisition, project administration, supervision, writing – original draft, writing – review and editing.
- A.K. – Data curation, methodology, supervision, validation, writing – original draft, writing – review and editing.

## References
1. **Booth B, Blair M.** *Thesaurus of sociological indexing terms* (3rd ed). Sociological Abstracts. 1992.
2. **Freedman Center for Digital Scholarship.** *Scanning.* Kelvin Smith Library Research Guides. 2019-09-04. https://researchguides.case.edu/scanning/home.
3. **Hurwitz HM.** Gender and race in the Occupy movement: Relational leadership and discriminatory resistance. *Mobilization: An International Quarterly.* 2019a; 24(2): 157–176. DOI: https://doi.org/10.17813/1086-671X-24-2-157
4. **Hurwitz HM.** #FemGA #SayHerName #NotHereForBoys: Feminist spillover in US social movements 2011–2016. In: Reger J (Ed.), *Nevertheless they persisted: Feminisms and continued resistance in the U.S. women's movement.* Routledge; 2019b. DOI: https://doi.org/10.4324/9780203728628-7
5. **Hurwitz HM.** (forthcoming). *Are We the 99%?: The occupy movement, feminism, and intersectionality.* Temple University Press.
6. **Library of Congress (US).** *Library of congress subject headings.* Library of Congress. 2020-08-17. https://id.loc.gov/authorities/subjects.html.