Legacy records Procedural Documentation

Brice Lanham & Avery Blankenship (last edit: January 8, 2020)

This project began with a substantial amount of physical material to be organized and catalogued. Brice and Avery both began by going through various boxes of physical documents, separating them when necessary (such as in the case of dissimilar documents in singular folders, mislabeled folders, etc.). Each of these documents were then catalogued in a spreadsheet with metadata columns, including: document titles; document creators; intended recipients; dates; document type tags; and a brief descriptive abstract. These metadata columns were populated as the information was present and left blank in cases where data was not explicit on the document itself. The documents were also filtered for any personal information such as student names, addresses, etc.

After each of the documents had been sorted into three different categories ("to be included," "of possible interest," and "not connected to the writing program") and their metadata collected, Brice and Avery then made a second pass through the documents. This second pass was to guarantee that appropriate metadata had been collected as well as to ensure that any documents marked as "of possible interest" were definitively placed in either the "to be included" or "not connected to the writing program" categories. Once all of the documents were sorted, each of the documents marked for inclusion were assigned a distinct filename. The intention of assigning a filename was to link the documents—both physical and digital—to their respective metadata. Additionally, a sticky note with the filename was placed on the physical documents to ensure that eventual .pdf scans of each document would be appropriately named.

Once all of the physical documents were appropriately sorted and tagged with filenames, Brice and Avery began to scan each of the documents. The documents were all scanned into .pdf format, and in the case of oversized documents (such as those that exceed 100 pages) the scans were split into multiple .pdf files, as the scanner could only handle so many documents at once. In cases where documents were split into multiple .pdf files they were later re-combined digitally. Additionally, to ensure consistency, all scanned documents were correctly oriented (landscape or portrait, depending on the document) and all blank pages within the .pdf were removed. Finally, the digital scans were examined once more to ensure that all sensitive or personal information had been caught and filtered.

In order to test that the scanned files would be able to match up with the metadata entries in the spreadsheet, Brice and Avery selected a sample set of 100 documents that they felt were representative of the variety of documents that would eventually be uploaded onto the archive's website. With the help of Northeastern University's Digital Repository Services, the sample set of documents were queued and uploaded online, with their respective metadata attached.

Once the sample set of files had successfully been ingested into the DRS, the ingested files were moved into a new folder to indicate that they were already online and their metadata entries were moved to a separate spreadsheet. Initially, there were a total of 1852 documents and the eventual "yes" set of documents totaled roughly 659.

Avery & Brice's Digital Archive "Guidebook"

- Agenda: a list of to-do items, likely in a meeting or short period of time
- Assessment: placement test material/assessment of student
- Checklist: Literally just a checklist of things to do.
- Committee Description: A description of what a committee does/will do
- Course Description: A description of what a course entails/will entail
- Departmental Documents: a big binder full'o'shit
- Email: Electronic communication between one or more individuals, usually marked by a user's email address.
- Evaluation: the determination of a course's, instructor's, or the department's success
- Exam: a course exam/test
- Fax: a fax
- Floorplan: a top-down view of a building
- Flyer: an advertisement of information or an event on campus or in the department
- Form: a document that provides blanks to be filled (in some cases have been filled)
- Guidebook: a book-length description of how to produce/do something
- Guidelines: a description of how to produce/do something
- Handbook: a collection of rules or requirements
- Handout: a sheet with pertinent information for a specific event
- Image: you know
- Letter: a correspondence between one or more people that was not sent as a memo or email
- List: items organized into a column
- Job Listing: a listing for a job
- Job Description: a description for a job
- Memo: an informal correspondence that provides information about events; more often than not, says memo at the top.
- Minutes: notes/information about a department meeting
- Newsletter: an informal publication describing updates or news in/on the department/campus
- Notes: supplemental handwritten annotations, often regarding another document or event
- Outline: an example of the end product
- Overview: a summary review or examination of courses, requirements, and the writing program itself
- Policy: a department or campus wide rule
- PowerPoint: a collection of slides that are often used to present information
- Prompt: directions for an assignment
- Proposal: a document that proposes a plan, course, or intention
- Publication: a piece of writing that was published (by someone from the department… that we know of)
- Report: a write-up of findings following an examination of something
- Resolution: a determined plan of action
- Schedule: a plan applicable to one or more days, often over a greater range
- Statement of Role: a statement about the role of a department's project
- Statistics: Any numerical data about students/employees/the university/etc.
- Student Writing: writing completed by a student
- Syllabus: a course description and schedule for any given course
- Webpage: a printed version of a then-internet-accessible page

Work Summaries

2019-2020

Brice Lanham & Avery Blankenship (last edits: May 12, 2020)

Summary of Work Done 2019-20:

Over the course of the academic year, the work of the project has been split into **Summer 2019**, **Fall 2019**, and **Spring 2020**

- **Summer 2019:**
  - Over the summer, Avery and Cameron finished sorting and recording metadata and filenames for the remaining documents.
  - The summer was also spent beginning the digitization process
  - Avery moved the boxes of material from the Writing Center to Neal's office. The documents were moved, once more, to the Barrs storage room.
- **Fall 2019:**
  - During the Fall, Brice and Avery finished scanning the materials (minus those materials that will eventually need to be split into their own pdf files)
  - Avery met with Sarah Sweeney to discuss developing a template for ingesting material into the DRS
  - Brice and Avery worked to develop documentation for the ingestion process and ingested a sample set of 100 documents into the DRS (which are all still privately available)
- **Spring 2020:**
  - The Spring semester was spent meeting in groups to discuss 4Cs presentations as well as details regarding the website
  - Alex began and finished recording metadata for the remaining box of syllabi (these still need to be scanned)
  - Avery met with Alex to discuss a workflow for post-COVID-19 work with the materials. Those instructions are here
  - Alex began the process of assigning aliases to person names in the metadata spreadsheet
  - Avery and Brice both began drafting some sample text for exhibits for the eventual website. Both of these documents can be located in the Drive where the scanned PDFs are
  - Avery and Brice both wrote narratives to describe the sorting and metadata recording process

- **Remaining Work**:
  - A second pass through files marked "no" to ensure that they should not be included in the archive
  - An additional pass through scanned documents to remove any documents that should be re-scanned to censor student information
  - Determination on which documents can legally be ingested into the archive
  - Development of the Wordpress site with relevant language surrounding use
  - Implementation of the CERES Exhibit Toolkit on the Wordpress site
  - Public permissions of the materials ingested in the DRS
  - Assigning aliases to remaining person names in the metadata spreadsheet
  - Splitting any PDFs which contain more than one document

Passwords and Usernames:

- WordPress: each team member should have their own username and password for the wordpress site which was created when the URL was assigned
- The DRS: each team member can access the collection in the DRS by using their MyNortheastern credentials (currently, the members who have access are Kyle, Neal, Avery, and Brice) Additional permissions can be granted by emailing Sarah Sweeney
- The Google Drive containing the scanned files:
  - [redacted]

DRS Resources and Links:

- https://dsg.neu.edu/services/drs/
- CERES Exhibit Toolkit: https://dsg.neu.edu/ceres/
- DRS user guide: https://repository.library.northeastern.edu/files/neu:cj82nc61v
- NEU Writing Program Digital Archive DRS page: https://repository.library.northeastern.edu/collections/neu:f1882336s
  - As of now, this collection requires that you login to the DRS (using your NEU credentials) and also have permission to view the collection. **(12/10/2019)**
- NEU Writing Program Digital Archive URL: **wparchive.northeastern.edu**

DRS Workflow and Ingestion Template:

- The template for recording metadata is located here
- The following folders within the shared Google Drive is as follows:
  - **Acquired folder**: This folder holds all material that the archive currently has possession of, but has not begun to ingest, yet. ***It is crucial that when new material is added to this folder, that the filenames are consistent across platforms, and that the filenames contain no non-alphabetical or numerical characters***. Upload the file to the "Acquired" folder
  - **Queued folder**: Once a file is ready to be tagged with metadata and ingested into the DRS, the file should be moved into the Queued folder, and its metadata should be logged in a copy of the template that is ***local*** to the Queued folder alongside the metadata of any other files in ***this same folder***.
  - **Ingested folder:** After a file has been ingested into the DRS, move that file into the Ingested folder. Files in the Ingested folder should be collected into their own sub-folders with the date of ingestion in the folder name. This way, we can keep track of dates so that files do not become lost.

- Ultimately, the workflow is as follows:
  1. Locate a particular file in the Acquired folder which is nested in the larger Google Drive account
  2. Locate the template for ingesting ads into the DRS and make a copy of that template. The template will be located in the folder titled "Templates"
  3. Record metadata for the file following the instructions below
  4. Move the material to the "Queued" folder to indicate that the material is currently being ingested
  5. Once the file has been ingested, move the material into the folder titled "Ingested"

Recording Metadata:

We have worked with members of the DSG to develop a metadata schema for ingesting the WP documents into the DRS. The various categories are listed below with a description of their function.

- **Filename**
  - It is especially important to keep file names consistent across all templates, folders, etc.
  - Follow this basic model: DOCUMENTTYPE_TITLE_YEAR_SPREADSHEETROWNUMBER.pdf
    - Ex. Syllabus_English102_2009_0195.pdf
- **DRS Destination Collection**
  - Please enter a valid DRS collection URL
  - The files will be ingested into their own collection here:
    - https://repository.library.northeastern.edu/collections/neu:f1882336s
- **Date**
  - If the file has an available date, list it here. It's really important that these dates follow the same MM/DD/YYYY format. Each month and day value needs to have two digits (09 instead of just 9) and each year needs four digits (2010 instead of 10). Google Docs may try to override your formating, if that is the case, then change the data type of the spreadsheet into "text" rather than "date" and that should solve the issue
- **Title**
  - Write out the full title (not abbreviated) of the text as it appears on the title page (if applicable)
  - Note: some of the files will not have a filename. In these cases, you will need to write your own short, descriptive title since the DRS requires that all files have a title
- **Document Type**
  - This column is meant for recording the general form/type of the document. For example, the document may be a Memo, Email, Syllabus, etc. It is important that you double check spelling and keep your capitalization consistent as this will allow for the groupings of document types in the DRS.
- **Author**
  - If the file has an identifiable author, include it here. You should separate authors names with commas
- **Recipient**
  - If the file has a specified, intended recipient, then those names or group names should be listed here.
- **Content Tags**
  - This column is designed to capture a sense of what the general content of the file is. These tags should be short and descriptive
  - The first tag should always be the same as the Document Type tag. This way, we can tag a document as both "Syllabus" and "Guidelines," etc.
- **Tagged for Interest**
  - This should always say "Yes." If in the case that as you're reviewing a document, it seems unconnected to the Writing Program, then that file should be moved out of the Queued folder and should just remain in the Acquired folder.
- **Description**
  - This should be a brief, one sentence description of the document. This of this column as an abstract for the document. This description will be what users will have access to before viewing the larger document.

After Ingest Workflow

- Once all the files have been ingested there are a few steps to take to ensure that the process is documented and organized:
    - 1) Move the ingested files from "Queued" into "Ingested"
    - 2) Move the ingest spreadsheet from "Queued" into "Ingested" with the files. They should all be kept in their own folder within the "Ingested" folder. A sample folder path could be: > "Ingested" > "12/09/19 Ingest" folder. Make sure the spreadsheet is properly titled with the date of ingest, as well
- Additional post-ingest steps: double check how the item appears in the DRS, and if necessary, upload an appropriate thumbnail for the item

---

[*Kyle begins discussing the direction of the project with Professor Neal Lerner in 2019 and becomes Project Manager. Avery & Brice depart the project in Spring 2020. Kyle, Avery, Brice, & Neal draft an article for the* Journal of Writing Analytics.

Kyle creates a new shared Google Drive to organize tasks. Brice & Avery's original Google Drive is retained in its original form.]

2020-2021

New Team Orientation

Kyle Oddis (created: September 2, 2020)

All (first steps)

- Read *Journal of Writing Analytics* Research Note Draft

Camila

- UX for Wordpress, site architecture, research with Kyle other open access projects (~25-30 in spreadsheet minimum)
    - List of pros and cons, design features
        - Ex. https://www.digitaltransgenderarchive.net/about/overview
            - https://kjrawson.net/trans-gender-timeline/
        - This might be helpful for a general search tool: https://archive.org/search.php?query=course+AND+subject%3A%22Northeastern+University+%28Boston%2C+Mass.%29%22&sort=-date&page=2
        - A great place to start is to look up some of those projects cited/mentioned in the *JWA* draft, and then google around to see what other kinds of open access archives or projects exist within/outside of academia (what can we pull from "industry" in terms of usability?) → create a spreadsheet in your folder? Up to you how to organize this
            - What kinds of tools & visualizations do you see? What are the narratives of the projects like?
- List of institutional archives with links (in person) -- visible archives (WAC Clearinghouse)
    - HBCUs / community colleges?
    - Categorize by two-/four-year research institutions, liberal arts colleges, etc.

Shannon

- Read "Report Draft" in the "Shannon" folder
- Qualitative coding training (with Kyle); Review syllabus scoring protocol
- Create first research project exhibit for Syllabi using Content Spreadsheet (Research Copy)
    - Qualitative coding
    - Incorporating SPSS (quantitative)

Sofia

- OCR/ICR, transcription

Joe

- Metadata, policy language

Kyle Oddis (created: September 24, 2020)

AY2020-2021 Workflow

All (GO Team!)

- Regular documentation upkeep and progress checks
- Publication opportunities (?! later)

Neal (P.I.)

- Qual coding practice for first exhibit & develop coding scheme (with Shannon & Kyle)
- OGC Follow-up
- Meetings with RAs/DRS/other people as necessary
- Potential external servers/hosting?
- Grant funding proposals
- Hiring/job postings/interviews as needed

Kyle (PhD Project Manager)

- Set goals for end of fall 2020 and end of spring 2021 for archive overall & for each RA
- Manage task lists and materials/workflows (bi-weekly/monthly)
- Create & revise documentation for qualitative coding procedure (with Shannon & Neal) along w/ all other documentation needed for general project management and archive administration
- Organize materials & documentation into G-Drive
- Create datasets from digitized archive material (batches to be coded and OCR/ICRd)
- Get access for RAs to WP as needed
- Meetings with RAs/Neal/DRS/other people as necessary
- Archive branding/maintenance of WP site
- Interviews as needed
- Proposal/draft feedback as needed

Shannon (MA Research Assistant)

- Qualitative coding practice for first exhibit (with Neal & Kyle)
- Develop coding scheme (with Neal & Kyle)
- Code & count (end of FA20)
- Data analysis & write up (end of SP21)

Camila (MA Research Assistant)

- Research UX and accessibility on existing digital projects & physical archives, create catalog/inventory (end of FA20)
    - Focusing on accessibility, navigation, narrative, data visualization options
    - What are more specific limitations of our platform? Create documentation
- Work with Kyle & DRS to get site architecture in place (end of SP20)
    - Work with Shannon & Kyle to create first exhibit

Sofia (UDG Research Assistant)

- Transcription, OCR/ICR
- Work with Joe on metadata revising/tagging
- Work with DRS on OCR'd files

Joe (MA Research Assistant)

- Metadata schema assessment & revision as needed
- Work on permissions & policy language

Kyle Oddis (last edit: November 13, 2020)

Updated RA Task List

*Your individual RA folders should contain all the materials required to complete your respective tasks.

Please plan to meet again with Kyle during the first week of December. Add your schedule availability to the "RA Availability" doc in the shared Google Drive folder.

Camila – Thanks for your work so far!

1. Begin research on physical archives (Neal can send some leads and provide an outline of what information to focus on here relative to what you focused on in your inventory of digital archives)
2. Create a spreadsheet (perhaps in the same sheet doc just in a new tab?) that's specific to physical archives
3. Create documentation in your RA folder for your researching/thought process to include your notes (how you decided on certain labels/characteristics, etc.)
4. Compile a list of links and brief (1-2 sentence descriptions) of each archive based on the info you collected
5. Depending on when the physical archive research is completed (so don't worry about finishing this by 11/30 if you can't get to it yet), start thinking about constructing some sort of potential comparative analysis of information you found (digital v. physical?) for us internally
   Deadline: 11/30

Shannon – Thanks for your work so far!

1. Create documentation in your RA folder for your qualitative coding process to include your coding process notes for each batch of files
2. Work through qualitatively coding Batches 1, 2, and 3 (the idea here is that you'll code a handful of syllabi each week) – Check in with Kyle about pacing: if you're able to do more than one batch in one week, you can move through at your own pace!)
3. Create a spreadsheet of code counts that we can translate easily into SPSS to do some quantitative analysis beyond basic frequencies (e.g., hopefully we can do some regression analysis later)
   Deadline: 11/30

Sofia – Welcome to the team!

1. Read article proof: "Possibilities for a Public-Facing Digital Writing Program Archive in the Age of Analytics" – Under "JWA Article Manuscript" in shared Google Drive folder
2. Practice OCRing sample texts in your folder, then OCR Batch 1 of archive documents, determine a method for double-checking accuracy (you won't need to deal with handwriting in the earlier batches)
3. Save OCRd files in a separate folder
4. Construct a spreadsheet that allows you to keep track of which files have and have not been OCRd using the naming conventions outlined in the article/in the Master Content Spreadsheet (check in with Kyle separately if you want some help/confirmation when setting this up)
   Deadline: 11/30

Joe – Welcome to the team!

1. Read article proof: "Possibilities for a Public-Facing Digital Writing Program Archive in the Age of Analytics" – Under "JWA Article Manuscript" in shared Google Drive folder
2. Review "permissions" and policy notices on the public archives listed in the spreadsheet Camila constructed, note the language used, talk with Kyle and Neal about composing a policy statement for our archive
3. Read through metadata PDF document for the DTA
4. Review our archive's current metadata schema
   Deadline: 11/30

[*Through several team members graduating/cycling through their degree programs, the project undergoes several revisions; spreadsheets are re-configured several times; documentation is reorganized; team acquires Blackboard LMS files after transition to Canvas LMS (these files would have otherwise been lost).*

Joe & Kyle work to organize these files and attempt to follow organizational procedures inherited from Brice & Avery. Kyle encounters generic difficulties with file naming and categorizations—implementation pause continues and team focuses on research development & training.

Joe works to organize Canvas files; RAs work on redacting acquired files and turning PDFs into plain text (.txt) files to plan for potential research uses. Joe & Shannon work on qualitative coding with Kyle & Neal.

Kyle delves into Inquiring & Imagining new systems and orients comprehensive exams to facilitate inquiry that will aid in project development. Following comprehensive exams (February 2021), the project undergoes transformation and new project management systems are put in place to apply conceptual foundations.]

2020-2021

WP Archive Workflow Checklist

Kyle Oddis (created: March 21, 2022)

Remainder of March

Week of 3/21

- Kyle send updated (new) spreadsheet
- Kyle email Sarah Green (see Urmi highlight in fair use document)
- Kyle finalize WP site architecture
- Kyle confirm WP access for Urmi
- Kyle & Neal follow up on opt-in/opt-out form w/ Laurie Nardone & Rachel Molko
- Kyle follow up with Neal on assigned tasks
- Urmi finish draft of fair use statement
- Urmi update reference list/bib for syllabus project to include in WP exhibit page
  - Include resources Kyle send + readings used for metadata
  - Put in G-Drive folder
- Urmi check attribution format (for citing the archive using multiple style guides; e.g., APA, MLA, etc.) for both the archive project and specific records

Week of 3/28

- Kyle & Urmi input drafted text content to WordPress (during Monday meeting– 1hr)
- Urmi ccreate exhibit page for syllabus project
- Kyle put together data vis for exhibits
  - Upload media files to WP
- Kyle update WP pages with existing documentation from G-Drive
  - Add descriptions to pages
- Kyle make a logo (don't like the five I already made)

April

Week of 4/4

- Urmi WP syllabi from 2016-2021 permissions
  - Opt in/opt out for adding to WP archive site
  - Kyle revise ingest spreadsheet (again!)
- Urmi check for UX on pages in WP preview (during Monday meeting)
- Review data vis/design options with Urmi (during Monday meeting)
  - Alt text/accessibility considerations
- Urmi read & copyedit/revise WP page text
- Kyle & Urmi check in with Sarah on ingest (this depends on whatever issues we might encounter after Kyle sends spreadsheet and files over)
  - Follow up on adding potential files following opt-in/opt-out responses for AY20-present

Week of 4/11

- Kyle build WP exhibits with ingested files from previous AY RA work
- Urmi work on Decolonizing syllabus project exhibit page (by this point Urmi will "own" this page/presentation of it based on project goals and status)
- Kyle continue documentation & site text
- Kyle continue working on UX design of site

Procedural Note on Redaction in Syllabi

Urmi Parekh (last edit: December 16, 2021)

I started with redacting professors' personal information from the syllabi. For Batch 1, I removed the whole information, e.g.: "Phone: xxx-xxx-xxxx". But from Batch 2, I decided only the number needs to be removed and not the subtitle, because it can be useful to know what personal information professors would decide to add to their syllabi. Like, some professors only had their office phone and email, whereas others had their home phone and fax information as well. So, I redid the redacting on Batch 1 for consistency.

- When the word 'Professor' is used as a prefix, e.g. Professor Frankenstein, it is redacted. But if it is used as a subtitle, e.g. Professor: Mr. Frankenstein, it is not redacted.
- Some syllabi have names of professors whose syllabus (or part of it) was used. Those names have been redacted as well.
- Some professors have mentioned books they've written in the 'about me' section. That information has not been redacted, but I've made a note of it in the s/s with records of redacted materials.
- Since the articles are publicly available, they don't need any redacting.
- I added <redacted> in place of names redacted from word/text documents.
- Some google docs and PowerPoints have comments in them. Since there is no way to redact names from comments, I have highlighted the text added as a comment and mentioned that it's an 'added comment'. Certain comments which are only about adding a paragraph have been accepted and no note is made, since the structural change is not very significant and does not impact the written information in the document.
- The excel sheet has 2 tabs. 'File Content' tab is for information redacted within the files. 'File Name' tab is for information redacted from the file names.

[Nearly all RAs worked on redaction procedures at some point. Approaches to these procedures and the organizational systems RAs used varied. Kyle had each RA decide on the organizational system and method that worked best for their workstyle and which best facilitated usability and task deliverables—all RAs were encouraged to document and reflect on their procedures and processes pertaining to task lists and project workflows.

Kyle met periodically with each RA (weekly, biweekly, or bimonthly depending upon scope and troubleshooting needs with specific software or tools). There was a great deal of freedom in procedural approaches to completing tasks; however, each RA was versed in project vocabulary, goals, and objectives.

RAs were also encouraged to pursue research interests of their own as they were inspired by/interested in various records. When RAs wanted to explore or research a topic or wanted to develop research method skills, Kyle worked with them to practice and develop methods to aid their research programs. This produced several exhibits that currently exist on NUWPArc. Kyle has revised & compiled these exhibits by formatting content to fit with the user interface redesign; exhibit text is drawn directly from RA documentation and notes (whatever documentation is not included here from RAs is included in the exhibits).]

2022-2023

[RA Procedural documentation for this year is still being developed.

Kyle's documentation involving updated ingest procedures and processes has been added to NUWPArc in a separate file to reflect "finalized" procedures executed in Spring 2023.

NUWPArc launched to the public on April 14, 2023.]