**Handbook for Collecting Newspaper Endorsements**

This handbook outlines the process for collecting newspaper endorsements and ensuring the data is accurate and complete. It includes steps for research assistants (RAs) and the quality assurance (QA) team.

**Contributors**: Kevin DeLuca[[1]](#footnote-1), Jim Snyder, Tyler Simko, David Beavers, Grace Aitken, Andrew Beingessner, Will Forbes, Sabrina Goldfischer, Khuan-Yu Hall, Owen Hannon, Patricia Hughes, Sasha Jones, Zoe Kava, Natasha Khazzam, Matthew Kind, Marina Legoretta, Julia Lin, Emily Mao, Nat Markey, William Porayouw, Alexandra Schoettler, Evan Sun, and Karen Yang.

**Data Description**

[tbd]

**PART 1: Data Collection Process**

**Step 0: Assignment**

* You will be assigned a newspaper (papername, ID number, and years) to collect endorsements from, by someone on the QA team.

**Step 1: Locate and Save Endorsements**

* Use newspapers.com to locate the assigned newspaper and the year for which you are collecting endorsements.
* Start with the editorial page on Election Day (the first Tuesday of November each year, unless Nov. 1st is a Tuesday; then it’s the second Tuesday in November).
* If endorsements are not found on Election Day, check the editorial page(s) for each day leading up to the election. Search back to 10 days before Election Day (the Sunday two weeks before). If no endorsements are found by then, move on to another year.
  + Note: Some papers do not make endorsements every year or may have stopped endorsing after a certain year. Make a note in the metadata about which years you checked, even if there are no endorsements in those years. For details, see metadata notes below in the final Folder Preparation section.
* Once you find the endorsement page for a specific year:
  + Save a clipping of it on the newspapers.com account.
    - Typically, the naming convention for clippings on newspapers.com is:

[Newspaper name] [year] endorsements [+office, if not a list]

e.g.s: “New York Times 2010 Endorsements”, or “Wilmington Evening News 1968 Senate”

* + Download a PDF of the page to a newspaper-specific folder on your computer (see Folder Preparation section below for details).
  + In the folder where you saved the PDF, name the PDF file of the clipping according to the newspaper naming instructions (Appendix A). In general, the naming format is:

DELUCA\_[StateAbbreviation]\_[PaperName]\_[IDNumber]\_[DatePublished]\_[Type]\_[Source#].pdf

example: DELUCA\_CA\_BERKELEYGAZETTE\_101400\_19821025\_ARTICLE\_1.pdf

**Step 2: Transcribe Endorsements**

* Type up the endorsements into the correctly structured format. It’s best to use the pre-formatted endorsement collection spreadsheets (ask QA Team for a copy, if you don’t have one). You should type up all endorsements from all clippings that you found.
  + For endorsements of candidates: each row in your data should represent a single year-newspaper-office level endorsement. For example: if a paper endorses multiple offices (e.g., President, Governor, Senator) in a single year, there should be one row for each different candidate endorsed in that year.
    - Note on non-endorsed candidates: you do NOT have to type up the information for candidates who were not endorsed, even if they are listed in the newspaper article
    - See the variable descriptions in Appendix B for details on typing up data and ensuring data is encoded correctly
  + For endorsements of propositions/amendments: each row in your data should represent a single year-newspaper-proposition level endorsement.
    - Note on non-endorsed propositions: unlike the candidate endorsements, we DO need the non-endorsed propositions/questions/amendments
    - In cases where information about the proposition/question/amendment is available in the endorsement newspaper article, write a 1-sentence (or less) description of the proposition in the “prop\_desc” variable field
    - See the variable descriptions in Appendix C for details on typing up data and ensuring data is encoded correctly
* Note: Straight ticket endorsements, or entire party endorsements
  + Sometimes, especially in papers/editions published before 1940s-ish, the newspaper will endorse straight ticket (meaning, they endorse every candidate from a specific party), or they will write an article that says something like “all Americans should vote Republican!” These are “straight ticket” endorsements, and to save time, we can record them in a single line, rather than recording every office.
  + For straight ticket endorsements, there would just be one row (for the candidate spreadsheet – the paper may also make a bunch of proposition endorsements, which would need to be recorded separately).
    - office = “STRAIGHT TICKET”
    - party = [party being endorsed, typically R or D]
    - cand\_name = “STRAIGHT TICKET”
    - leave dist, dname, and incumbency variables blank; fill in notes as necessary
* Double-check for errors, typos, or missing information to ensure the transcription is as complete and accurate as possible. You should spend a (very) short amount of time looking for missing data – if it can’t be found within 1-2 google searches, just leave the information as missing. The most important variables to try to find values for, if they are missing, are: party labels of candidates, district number (if applicable), and incumbency status.

**Step 3: Folder Preparation**

* After collecting and typing up all endorsements for a newspaper, you should prepare a folder that you will share with the QA team with all the clippings, your spreadsheet of typed endorsements, and a metadata text (txt or rtf) file.
* Checklist: Your final folder should have the following items:
  + A spreadsheet named [newspapername]\_[newspaperID]\_candidates.xslx, which has all the typed up *candidate* endorsements finalized.
  + A spreadsheet named [newspapername]\_[newspaperID]\_propositions.xslx, which has all the typed up *proposition*/amendment/question endorsements finalized.
  + A folder called “clippings”, which stores all the pdf clippings, appropriately named according to the naming instructions (see Appendix A)
  + A metadata text file in the folder. The file should include the following, each on a separate line:
    - Newspaper name
    - Alternative names (e.g. if the paper changes its name at some point)
    - Years of publication
    - Newspaper ID number
    - Total number of clippings in archive
    - Years of clippings/endorsement data included
      * If clippings and endorsements years are different, separate these out (see Los Angeles Times example data below)
    - Total number of candidate endorsements
    - Total number of proposition endorsements
    - Notes on any missing or incomplete data. (e.g. years checked where there were no endorsements or where the archive data was missing from newspapers.com)

**Metadata example/template:**

Newspaper Name: Los Angeles Times

Alternative Names: Los Angeles Daily Times (1881-1886), Los Angeles Times (1886-present)

Years Published: 1881-present

Newspaper ID number: 103000

Number of clippings: 102

Clippings years: 1882-1928, 1933, 1936-1988, 2004, 2008-2022

Endorsement data years: 1882-1930, 1940-2006, 2008-2020 (pres only), 2022 (propositions)

Candidate endorsements: 515

Proposition endorsements: 1196

Notes:

-Technically drops the “Daily” from the title in 1886 but they still print it at the top of the page for a bit after.

-In early years they endorse party line but sometimes also write editorials about specific candidates - these are coded as “straight ticket” in the data even though you could also look up editorial for specific candidates if needed in some cases.

-Spreadsheets are missing some endorsements for local and other non-statewide or non-federal offices/amendments/propositions in many years

**Step 4: Submit for Quality Assurance Check**

* Once you folder is prepared and all data has been collected and cleaned, share the completed folder (clippings, typed endorsements, metadata) with the QA team.
* QA team member will check over work and inform you of anything in need of correction before archiving occurs.

**Step 5: Address Quality Assurance Feedback**

* If QA team identifies errors or omissions, fix them and resubmit the corrected folder for review.
* After completion, you will be assigned a new newspaper’s endorsements to collect.

**PART 2: Quality Assurance Process**

**Step 1: Review Submitted Data**

The main purpose of the quality assurance (QA) process is to ensure that the data is correctly compiled, stored, and recorded before we catalog it in the official archive. You should receive a newspaper-specific folder from an RA who is done collecting a newspaper’s endorsement data (sometimes, assigned to a specific time period). Here is a checklist of what needs to be reviewed:

* Is the folder organized properly? The folder should have the following:
  + The folder name should be the newspaper name
  + A spreadsheet named [newspapername]\_[newspaperID]\_candidates.xslx, which has all the typed up *candidate* endorsements finalized.
  + A spreadsheet named [newspapername]\_[newspaperID]\_propositions.xslx, which has all the typed up *proposition*/*amendment/question* endorsements finalized.
  + A folder called “clippings”, which stores all the pdf clippings, appropriately named according to the naming instructions (see Appendix A)
* Are the clipping file names accurately named? (see Appendix A)
  + Also, make sure you are able to open the files and that the clippings were done correctly (e.g. that it is the pdf download of the newspapers.com clipping info, if the clipping is from newspapers.com, the clipping itself has newspaper endorsements and the endorsements are from the newspaper (not ads from candidates), etc.)
* Are the variables for the candidate endorsements correctly recorded?
  + Make sure all the right variables (for candidates) are there, that text fields are text and number fields are numbers, that the variables were encoded correctly (1s and 0s in their right places), and that there is no significant missing data (e.g. shouldn’t have candidate names missing, shouldn’t have the “endorsed” variable missing, etc.)
* Are the variables for the proposition endorsements correctly recorded?
  + Make sure all the right variables (for propositions) are there, that text fields are text and number fields are numbers, that the variables were encoded correctly (1s and 0s in their right places), and that there is no significant missing data (e.g. shouldn’t have proposition names missing, shouldn’t have the “endorsed” variable missing, need to record BOTH endorsed and not-endorsed propositions/question etc.)
* Are all the endorsements in the clippings recorded in the spreadsheets?
  + To check this, open up a sample of the clippings (pick 5 or more, randomly) and see if you can find the endorsements that are made in the clipping in the spreadsheet(s).
* Is there a metadata text file with the appropriate and correct information about the data in the folder?
  + See metadata text info in part 1, data collection process

**Step 2: Provide Feedback**

If you find problems with the folder or data, you have two options:

1. *Send back to the RA to correct the errors*
   1. This is better for ‘bigger’ problems, like if there is an issue with the way variables were recorded, if the clipping file names are not correctly recorded, or if not all of the endorsements in the clippings are contained within the endorsement spreadsheets
   2. In this case, clearly document the issue(s) and (nicely) inform the RA of what needs to be fixed before the newspaper’s data is complete. A good practice would be to give numbered bullet points of everything that needs to be addressed. Answer any RA questions and assist them in figuring out the best way to correct their data.
   3. Check over *all* the data first, and then send back to RA if there are problems. We want to minimize the number of times that folders go back and forth for corrections, so if we can catch all/most of the errors in the first check, that will be more efficient.
2. *Correct the errors yourself before archiving*
   1. This is best for smaller errors, like incorrect metadata, which can be easily fixed by a member of the QA team without much hassle.
   2. Make your best call on this. Even if you can quickly fix the error, you should still inform the RA of the issue so that they can do it correctly next time.
   3. In general, if the error involves the data collection itself (clippings + spreadsheet data), then the RA should go back and fix the problem(s). Your job here is to check and correct the data, not to go back and redo the data collection itself.

**Step 3: Recheck Data**

If you send the data back to the RA for corrections, you will need to recheck for accuracy. Go through the whole checklist again, not just the parts that needed to be fixed!

* Repeat Steps 1-3 as necessary until the folder meets quality standards.

**Step 4: Finalize and Archive**

When the newspaper data is complete and has been checked (and potentially corrected and re-checked), then that means it is ready to be cataloged in the github archive

There are two cases the new data could fall under, and the processes from archiving the completed data are slightly different for each:

* 1. *The endorsements are from a new newspaper that does not currently have a folder in the archive*
     + First, “pull” from the github archive to make sure your archive is up to date before “pushing” any changes
     + Second, move the completed folder into the correct part of the archive. Newspaper-specific folders should go under the appropriate state folder within the “newspaper endorsement archive” folder
       - The archive is organized: Endorsement Archive -> [State Name] -> Newspaper
     + Third, “push” the changes to the archive. For the description of the change, you should make a note of the newspaper data being added to the archive
       - E.g. “added New York Times endorsements, 1950-1986”
     + Last, check that the data was successfully added to the archive
  2. *The endorsements are additional endorsements for a newspaper that already has a folder and data in the archive*
     + First, “pull” from the github archive to make sure your archive is up to date before “pushing” any changes
     + Second, you will have to (manually) merge the newly collected data into the existing folder for the newspapers:
       - Add all clippings from the new clippings folder into the clippings folder that is already in the archive
       - Update the metadata file in the existing folder, update the new clippings and endorsement data that was added (years, total number, etc.)
       - Combine the newly collected endorsements and already-archived endorsements into a single spreadsheet (one for candidates and one for propositions)
         * Be extremely careful not to overwrite or leave off any endorsements!
         * Check that: number of endorsements in the newly collected data PLUS the number of endorsements in the already-archived data is equal to the number of endorsements in the data you’re about to “push” to github *–* check the number of rows on the spreadsheet(s)
         * Sort the endorsements in the spreadsheet(s) by year, to make sure that you have not accidently excluded particular years of data
     + Third, once you are very sure that you have updated the data in the archive correctly without deleting any data, you should “push” the changes to the archive. For the description of the change, you should make a note of the newspaper data being updated within the archive
       - E.g. “updated New York Times endorsements, added 1986-2020”
     + Last, check that the data was successfully added to the archive

**Appendices**

1. **File Naming Conventions**

*A.1 PDF Files of Newspaper Clippings*

All PDFs of clippings should be named according to the following conventions. Please follow the instructions closely and make sure each clipping is appropriately named before submitting a folder to the QA team.

The basic format of PDF file names of clippings is:

DELUCA\_[StateAbbreviation]\_[PaperName]\_[IDNumber]\_[DatePublished]\_[Type]\_[Source]

**Example name:** DELUCA\_CA\_BERKELEYGAZETTE\_101400\_19821025\_ARTICLE\_1

NOTE: MAKE SURE ALL CLIPPING FILE NAMES ARE IN ALL CAPS

**Part 1/2: Researcher name/State abbreviation**

* All clippings should start with DELUCA and the two-letter abbreviation of the state the newspaper is from. If unsure about a state’s abbreviation, refer to: <https://www.faa.gov/air_traffic/publications/atpubs/cnt_html/appendix_a.html>

**Part 3: Newspaper name**

* When writing the newspaper’s name, if it has a common name (The Post, The Daily News, The Gazette, etc), add the location to distinguish it (examples: New York Post, Corona Daily News, Berkeley Gazette). Typically, this will be the main city where the paper is published (you can usually find the city where the paper is located in newspapers.com). If unsure, ask a member of the QA team.

**Part 4: ID number**

* ID numbers should be given to you when you get your assignment, and are typically 6 digits long – ask a QA team member if you are unsure what the paper ID number is.

**Part 5: Date published**

* The day the article was published
* Format: YYYYMMDD (Ex: November 5, 1964 is 19641105)
* If there are two clippings from the same day, number them *after the date* (not at the end of the clipping!)
  + Ex: if there are three different clippings from the Berkeley Gazette on November 2, 1964, the names would look like this:
    - DELUCA\_CA\_BERKELEYGAZETTE\_101400\_19641102\_1\_ARTICLE\_1
    - DELUCA\_CA\_BERKELEYGAZETTE\_101400\_19641102\_2\_ARTICLE\_1
    - DELUCA\_CA\_BERKELEYGAZETTE\_101400\_19641102\_3\_LIST\_1

**Part 6: Endorsement format**

* Was the endorsement an article or a list? For article, put “ARTICLE”; for list, put “LIST”
* Even if an article mentions multiple candidates/propositions, generally anything that has description of who the candidates are and why they should or shouldn’t be endorsed should count as an article. Lists should be cases where only the candidate names or propositions are listed, and there is no justification for the endorsement mentioned.
* Examples of lists vs articles are below at the end of these instructions.

**Part 7: Source**

* Where the clipping came from, denoted by numbers
* So far, here are the categories you can choose from:

1: Newspapers.com

2: Newspaperarchive.com

3: Other

4: News Bank Inc.

5: ProQuest Historical Newspapers

* We will mostly be collecting articles from newspapers.com (so the clippings will end with a “1” in most cases)

*A.1.1 Examples of List and Article Formats*

List examples:

A document with text on it

AI-generated content may be incorrect.A black and white paper with black text

AI-generated content may be incorrect.A close-up of a legal document

AI-generated content may be incorrect.

Article examples:

A newspaper article with text

AI-generated content may be incorrect.A close-up of a newspaper

AI-generated content may be incorrect.A close-up of a newspaper

AI-generated content may be incorrect.

*A.2 Endorsement Spreadsheets*

The endorsement spreadsheets should be .csv files, and should be named:

* Candidate endorsements: [newspapername]\_[newspaperID]\_candidates.csv
* Proposition endorsements: [newspapername]\_[newspaperID]\_propositions.csv

*A.3 Metadata Text Files*

And finally, the metadata text (txt) file should be named:

* [newspapername]\_metadata.txt

1. **Variables – CANDIDATE ENDORSEMENTS**

For the candidate endorsements spreadsheets, all of the following variables should be included in the data:

* state\_newspaper: state where the paper is located (two-letter abbreviation)
* newspaper: name of newspaper
* newspaper\_id: newspaper ID number
* year: year of election when the endorsement is occurring
* office: office that is being endorsed
  + Suggested ways to record offices is at the end of this document, in Appendix D
  + If an office is not on the suggested list, record as it is reported in the newspaper
  + Avoid too-generic labels like “Congress” (is this US House or US Senate?), or labeling things like “SS” (is this secretary of state or state senate?), etc.
  + Make sure to code the same offices consistently within your own spreadsheets. E.g. don’t label things as “State House” in one year of data and then “ST House” in the next year for the same office. Make notes of any labels that may be confusing in the metadata notes.
* dist: district number of the office for which the endorsement was made (if applicable)
* dname: If the district is just a number, this variable should be the same as the “dist” variable. For weird district names, like letters or town/city names, you can type the name here. For positions like “County Supervisor”, you should put in county name here.
* state\_election: state of the office/election. Some papers located in one state will endorse candidates in another state. Usually, this is the same as state\_newspaper but not always. Newspapers will make it clear if it is a cross-state endorsement.
* cand\_name: name of the candidate endorsed. Try to do a consistent format:
  + I use “Last Name, First Name “Nickname”, Suffix”.
  + As a last resort, just record this as it is listed in the endorsement (e.g. if they only list a nickname, or if they list something like “Mr. DeLuca” instead of the candidate’s full name).
* party: party of the candidate being endorsed
  + “D” = Democratic party
  + “R” = Republican party
  + “LBT” = Libertarian party
  + “PROG” = Progressive party
  + “IND” = Independent
  + “NP” = Non-partisan race
  + Use your own judgement for other parties (record the party as listed in the endorsement). Avoid putting “Other” as the party – should be more specific than this.
* endorsed: endorsed=1 if the candidate is endorsed, endorsed=0 if candidate is NOT endorsed. For now, we are collecting only the endorsed candidate info so this should basically always = 1 for manually recorded endorsements that we collect. The exception is if a candidate is explicitly not endorsed – for example, if they recommend voting *against* a candidate (but not necessarily for voting *for* the other candidate).
  + For “No Endorsement” or “Endorsed Both” cases, make a note of this in the notes
  + For “No Endorsement”, set endorsed = 0 (ideally for both candidates as separate rows, or can leave the name variable blank and have two blank rows with D and R in the columns – ask about these or take note of them).
  + For confusing cases, ask a member of the QA team.
* d\_inc: is equal to 1 if the democrat is the incumbent, 0 otherwise
* r\_inc: is equal to 1 if the republican is the incumbent, 0 otherwise
* o\_inc: is equal to 1 if a candidate from a different party (or an “independent”) is the incumbent, 0 otherwise
  + If there is no incumbent, then d\_inc, r\_inc, and o\_inc should all be = 0
* notes\_endorse: any additional notes about the endorsement. Try to add in “incumbent” when possible (i.e. when indicated on the endorsement page, especially if there is missing party info), or “neither endorsed” or “both endorsed” when those cases arise.

1. **Variables – PROPOSITION/AMENDMENT ENDORSEMENTS**

* state\_newspaper: state where the paper is located (two-letter abbreviation)
* newspaper: name of newspaper
* newspaper\_type: “N” = newspaper, “S” = “special” newspaper, like “alt press” papers, weekly papers, magazines, student newspapers, etc.
* newspaper\_id: newspaper ID number
* year: year of election when the endorsement is occurring
* state\_election: state of the proposition. Some papers located in one state will endorse propositions in another state. Usually, this is the same as state\_newspaper but not always.
* prop\_type: the “type” of proposition/amendment/question.
  + Examples are “proposition”, “amendment”, “question”, etc. (record however it is reported in the newspaper). This may vary by state or may vary based on whether the proposition/amendment/question is at the state vs. local level.
* prop\_num: number (or letter) of the proposition, e.g. for recording proposition 101 this variable would be equal to “101”; for Question 2B, this variable would be equal to “2B”
* prop\_desc: a short description (1 sentence or less) of the what the proposition/amendment is about
  + Some examples of acceptable descriptions:
    - Ex1: “RELATING TO SAME DAY AIRBORNE HUNTING OF CERTAIN ANIMALS”
    - Ex2: “$315 MILLION BONDS FOR ROADS AND PORTS”
    - Ex3: “1/8 OF 1% SALES AND USE TAX FOR GAME AND FISH”
  + This field can be left blank if there is no info about the proposition/amendment in the newspaper article endorsement
* endorsed: endorsed=1 if the proposition is endorsed/recommended, endorsed=0 if the proposition is NOT endorsed/recommended. You should record and collect **both** the recommended and not-recommended proposition endorsements \*note: this is different from the candidate endorsements, where you can record only the endorsed candidate.
  + For “No Endorsement” or “No Recommendation” cases, make a note of this in the notes\_endorse variable – these are different than recommending against an endorsement. In these cases, leave the endorsed variable blank (rather than 0 or 1).
* notes\_endorse: any additional notes about the proposition or endorsement.

**D. Office Variable – Suggested Coding**

*How you record the office in the data: the office as reported in the paper*

* PRESIDENT: President
* VP: Vice President
* S: US Senate
* H: US House
* ST H: State House (lower chamber, includes “State Assembly”)
* ST S: State Senate (upper chamber)
* G: Governor
* LT GOV: Lieutenant Governor
* SEC STATE: Secretary of State
* ATTY GENL: Attorney General
* AGRIC: Agricultural Commissioner
* INSURANCE: Insurance Commissioner
* TREASURER: State Treasurer
* MINES: State Mine Inspector
* LABOR: Commissioner of Labor
* RR COMM: Railroad Commissioner
* PUB SRV C: Public Service Commissioner
* LANDS: Land Commissioner
* TAX COMM: Tax Commissioner
* COMPTROLL: State Comptroller
* AUDITOR: State Auditor
* ST SUPR CT: State Supreme Court
* SBOE: State Board of Education
* CORP COMM: Corporation Commissioner
* CRIM APP: Court of Criminal Appeals
* SEC OF INTERNAL AFFAIRS: Secretary of Internal Affairs
* MAYOR: Mayor
* SUPERIOR COURT JUDGE: Superior Court Judge
* DIST COURT JUDGE: District Court Judge
* CIV APP: Civil Appeals Court
* CNTY COMM: County Commissioner
* CNTY EXEC: County Executive
* CNTY SUP: County Supervisor
* CNTY LEG: County Legislature
* CNTY TREASURER: County Treasurer
* CNTY ASSESSOR: County Assessor
* CNTY SUPT OF SCHOOLS: County Superintendent of Schools
* DIST ATTY: District Attorney
* SHERIFF: Sheriff
* CONSTABLE: Constable
* CNTY CORONER: County Coroner
* CNTY COUNCIL: County Council
* JP: Justice of the Peace
* STRAIGHT TICKET: Straight ticket endorsement

1. Corresponding author and maintainer. Contact: [kevin.deluca@yale.edu](mailto:kevin.deluca@yale.edu) [↑](#footnote-ref-1)