


The Illustrated Guide to the WE1S Project (rev. 20 Aug. 2022)

Contents

Research Goals	1
Research Materials	2
Innovations in Research Methods for DH.....	6
Research Tools	9
Research Reporting	11
Research Findings.....	15
Recommendations	16
Repository Deposits (for openness, reproducibility, and sustainability).....	17

This document is a descriptive summary of the WhatEvery1Says (WE1S) project. It is more detailed, scholarly, and administrative in its voice (including language about the project originally created as part of grant reports or reports to the universities of WE1S researchers) than the [“Our Story & Our Results”](#) explanation written in public voice on the WE1S home page.

 For full documentation of the project, see the WE1S website (<https://we1s.ucsb.edu/>).

Research Goals

The WhatEvery1Says (WE1S) project uses digital humanities methods to study media discourse about the humanities at big data scales. The project concentrates on, but is not limited to, journalistic articles in the U.S. available in digital textual form beginning circa 1981. Other materials include social media. The hypothesis of the project is that digital methods can help us learn new things about how the humanities have been portrayed in public discourse. For example, are there sub-themes beneath the familiar dominant clichés and memes? Are there hidden connections or mismatches between the “frames” (premises, metaphors, and narratives) of those arguing for and against the humanities? Where do the humanities fit among other topics in public discussion? How do different sources (for example, mainstream versus student newspapers) or different parts of the nation or world compare in the way they talk about the humanities? And in what way—if to any significant degree—does journalistic media position racial, ethnic, gender, first-generation-student and other social groups in relation to the humanities? Instead of concentrating on set debates and well-worn arguments, can we exploit new ideas or surprising commonalities to better grasp the role of the humanities in the 21st century?

WE1S uses a method of computational machine learning called “topic modeling” (complemented by other computational methods) to assist in understanding patterns in large collections of documents and in guiding researchers to representative texts for close reading. WE1S also conducted surveys and focus groups with students and others to help contextualize perceptions of the humanities.

Outputs from the project include findings and recommendations (posted initially in brief “card” format) alongside longer reports, blog posts, and other materials. WE1S also makes data and tools available for

others to use—both for the general purpose of promoting open, replicable research methods and for the specific purpose of inviting others to extend WE1S’s mission of understanding how the world (including other nations) talks about and constructs an understanding of the humanities.

The ultimate goal of the WE1S project is to provide advocates for the humanities with research-based materials and strategies for effective communication about the value of humanistic knowledge in today’s world—with evidence, arguments, narratives, and scenarios that advance, rather than simply react to, public conversation.

Research Materials

► Corpus of Primary Research Materials

The final corpus of research materials that WE1S harvested consists of over 1 million (1,028,629) unique English-language journalistic media articles and related documents mentioning the literal word “humanities” (and for some research purposes also the “liberal arts,” “the arts,” and “sciences”) from 1,053 U.S. and 437 international news and other sources between the 1980s and 2019, though mostly after the year 2000 when news media began producing digital texts en masse. For comparison and other uses, WE1S also gathered about 1.38 million unique documents representing a random sampling of all news articles in the U.S. In addition, the project harvested over 6 million posts mentioning the “humanities” and related terms from social media (about 5 million from Twitter and 1 million from Reddit).

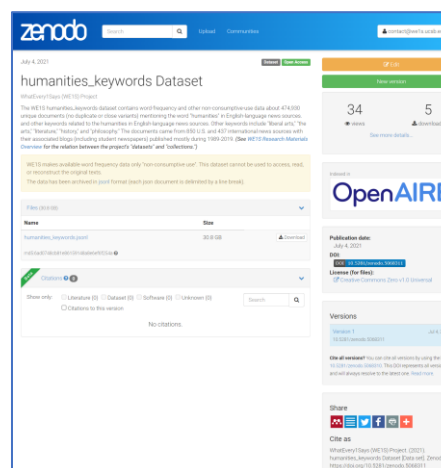
WE1S concentrated on analyzing results from U.S. media at the urging of its Advisory Board, which counseled focusing on what the project could do well in the time frame of the grant and on what project researchers understood best. (By contrast, WE1S researchers based in the U.S., even though they included RAs and others originally from other nations, were not equipped to understand well the audiences and roles in society of English-language media in international locations where English is not the main language.)

WE1S’s primary source for news articles was the LexisNexis Academic database (accessed algorithmically at scale through the paid LexisNexis “Web Services Kit” API), supplemented by other databases such as ProQuest requiring more manual harvesting methods. The project also directly harvested from the web a variety of born-digital news and other texts (using the [“Chomp”](#) web-scraping tool it created) and from social media (using tools it adapted for its [“TweetSuite”](#) and [scripts and other methods for scraping from Reddit](#)).

► Data Derived or Generated (Visualized) from Corpus:

Datasets

From its corpus of research materials, WE1S derived or generated datasets of primarily [“non-consumptive use”](#) materials, including metadata, term-frequency lists (“bags of words”), parts-of-speech, and other natural-language-processing information that can be used for computational modeling but not plain-text human reading. Due to copyright and database licensing constraints, WE1S only makes publicly available as its “datasets” this non-consumptive use data (except for public domain or other open texts the project directly harvested). These datasets (organized in six segments, each assigned a DOI) are deposited for access and sustainability in the international [Zenodo](#) open science data repository. WE1S shares these datasets as sources of rich material for others with their own research questions. There is no other existing dataset designed to make visible at scale the public perception or representation of the humanities in the media.



Example of WE1S dataset deposited in [Zenodo](#).

Collections:

The datasets described above represent data about all documents collected by WE1S. By contrast, what the project calls its “collections” are subsets of datasets representing portions of this corpus filtered by keyword, source, or other criteria to help address particular research questions. For instance, WE1S’s [Collection 1 \(C-1\)](#) is a subset of its *humanities_keywords* dataset of documents mentioning the word “humanities” (and phrases related to the humanities such as “liberal arts” or “the arts”). Focusing on the U.S., C-1 reduces that original dataset of 474,930 articles to 82,324 articles mentioning “humanities” just from 850 U.S. news sources. Other “collections” focus on top-circulation newspapers, student newspapers, articles mentioning either “humanities” or “science(s),” and so on. (See the [metadata tags](#) WE1S added to its data to help analyze groups of publication sources.)

The importance of these “collections” is that they provide research starting points not just for WE1S but for others interested in pursuing particular questions. For example, WE1S’s [Collection 21 \(C-21\)](#) – U.S. Top Newspapers, 2000-2018 (articles mentioning “humanities” or “science[s]”) – could prompt many more research inquiries into how the humanities and sciences are represented to the public than WE1S itself studied for its [“Humanities and Science in the Media”](#) key findings.

WE1S makes available in the [Zenodo](#) open science data repository a total of 19 “collections” for others to download ([example](#)), along with each collection’s topic-model data files and interactive visualization

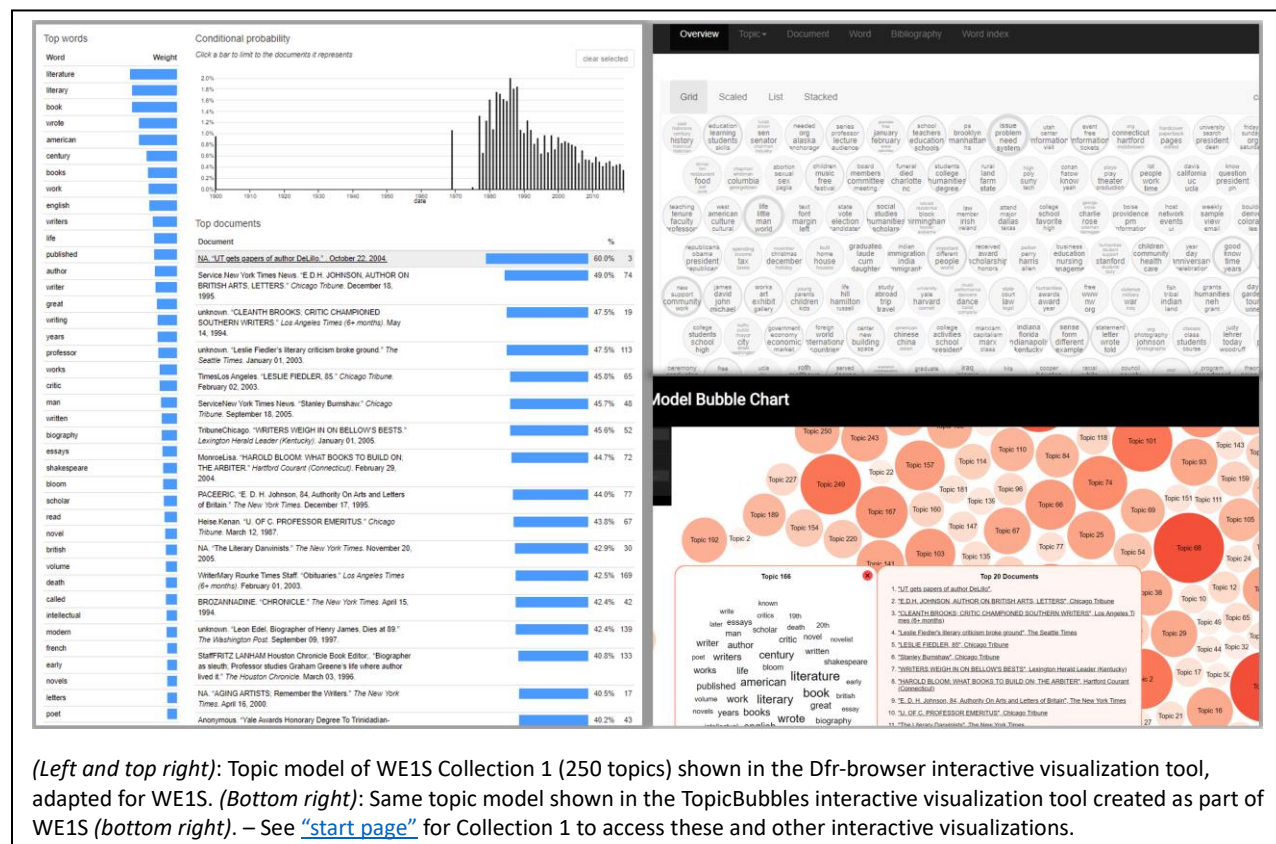


WE1S “collections” described on 1-page cards ([Key Collections](#))

interfaces. (Topic-model data files and the files used to generate visualization interfaces are deposited in Zenodo as the production “project” files of each collection.) In addition, WE1S makes available on its website live, interactive visualizations of the topic models it created of these collections— including a way to access non-consumptive features tables (alphabetized lists of tokens containing annotations of the tokens' part of speech and named entity status) serving as representations of each article that cannot be shared in readable form due to copyright and licensing constraints) ([example](#)). (See *more detailed description of topic models and visualization tools immediately below*.) WE1S also provides “metadata tags” (see [searchable list](#)) that categorize publication sources and make it possible during modeling and analysis to ask such questions as: “is a specific topic in a topic model most associated with news sources from the U.S. West or South, or with public or private universities?”

Topic Models


WE1S created for each of its “collections” a “Start Page” ([example](#)) providing access to topic models of texts in that collection. Using the [MALLET](#) tool for topic modeling, the project generated several models for each collection at different granularities of numbers of topics (typically 25, 50, 100, 150, 200, and 250 topics). WE1S makes available these models as interactive visualizations on its website and also in Zenodo deposits of its production projects containing downloadable MALLET data files that can be used by other researchers to re-create models. (In addition, WE1S makes available in downloadable form its whole algorithmic Workspace so that other researchers can themselves reproduce or vary its procedures for making models.)



Visualizations (and Visualization Tools)

WE1S's topic models come with ready-made interactive visualization interfaces. Typically, four different interfaces are provided for each topic model of a collection: [Dfr-browser](#), [TopicBubbles](#), [pyLDavis](#), and [DendrogramViewer](#). (TopicBubbles and DendrogramViewer were created by WE1S; Dfr-browser, as implemented by WE1S, was adapted in customized form from an existing topic-model exploration tool; and pyLDavis was used as an off-the-shelf tool.) WE1S also incorporated an adaptation of the topic model [Diagnostics visualization](#) tool from the MALLET website. These visualization interfaces allow users to explore topics and their associated words and texts interactively. WE1S also prototyped other experimental visualization tools not yet applied widely to its collections (e.g., [Metadata7D](#), [GeoD](#), and the open source Python [Scattertext](#) library).

Ready-to-use visualization interfaces are available from the “start page” of each WE1S data “collection” (e.g., [start page for Collection 1](#)). This whole suite of visualization interfaces is what WE1S calls its “[Topic Model Observatory](#)”, which is provided to other researchers with downloadable visualization model data and Jupyter notebooks for making visualizations as part of the WE1S Workspace (see below).


[Suggested Citation](#)
[Metadata](#)
[Other WE1S Collections](#)

Collection 1: U.S. News Media, c. 1989-2019

(WE1S core collection of articles mentioning “humanities”)

A collection of word-frequency and other data representing 82,324 unique articles mentioning “humanities” (no duplicate or close-variant documents) published mostly during 1989-2019 in 850 U.S. news sources and their associated blogs. (About 5,000 articles originate from earlier in the 1980s.) The word “humanities” occurs 134,948 times in the collection. WE1S and other researchers use this data to look for broad patterns and to help guide closer study.

WE1S makes available only “non-consumptive use” word frequency, topic model, and other datasets along with their visualizations. Datasets cannot be used to access, read, or reconstruct the original texts.

The top 20 news sources in Collection 1 (ordered by number of articles in descending order) are: *New York Times*, *Washington Post*, *Los Angeles Times*, *Chronicle of Higher Education*, *Chicago Tribune*, *News Day*, *Boston Globe*, *Hartford Courant*, *St Louis Post Dispatch*, *Oregonian*, *Dallas Morning News*, *Stanford Daily* (Stanford University), *Advocate* (Baton Rouge), *Birmingham News*, *Star Ledger* (Newark), *Knoxville News Sentinel*, *Chicago Daily Herald*, *Providence Journal*, *Daily Oklahoman* (Oklahoma City), *Charlotte Observer*. (Full sources and counts are available as a csv file in the Collection dataset.)

Kinds of Sources (by Tags)

Sources in Collection 1 are tagged (through metadata) as associated with categories that include those listed on the [Metadata Tags for WE1S Document Sources](#) page. (Sources are assigned to categories based solely on explicit publication information and/or self-identification.)

(See also a [visualization](#) in an experimental “GeoD” interface created by WE1S researchers of a version of Collection 1 enhanced through “wikification” to create a topic model that includes geotagged and named entity tokens. See [GeoD](#) in the *WE1S Topic Model Observatory Guide*.)

Suggested Citation for Collection

WhatEverySays (WE1S) Project. (June 20, 2019). Collection 1: U.S. News Media, c. 1989-2019 (WE1S core collection of articles mentioning “humanities”). Zenodo. DOI 10.5281/zenodo.4902187.

Collection Metadata

- Created by: Lindsay Thomas
- Created on: June 20th 2019, 12:00:00 am
- WE1S Collection Registry ID: 20190620_2238_us-humanities-all-no-reddit
- Data sources: LexisNexis (via LN Web Services Kit), ProQuest, and direct scraping from the Web.
- Collection dataset (“non-consumptive use” data derived from the original texts): DOI: 10.5281/zenodo.4902187

Topic Models of This Collection

Model Family 1 (created June 20, 2019): models for 25, 50, 100, 150, 200, 250 topics

Visualizations for this model family.
(Click on radio buttons to open the interactive visualizations. See [Topic Model Observatory Guide](#) for how-to for visualization interfaces.)

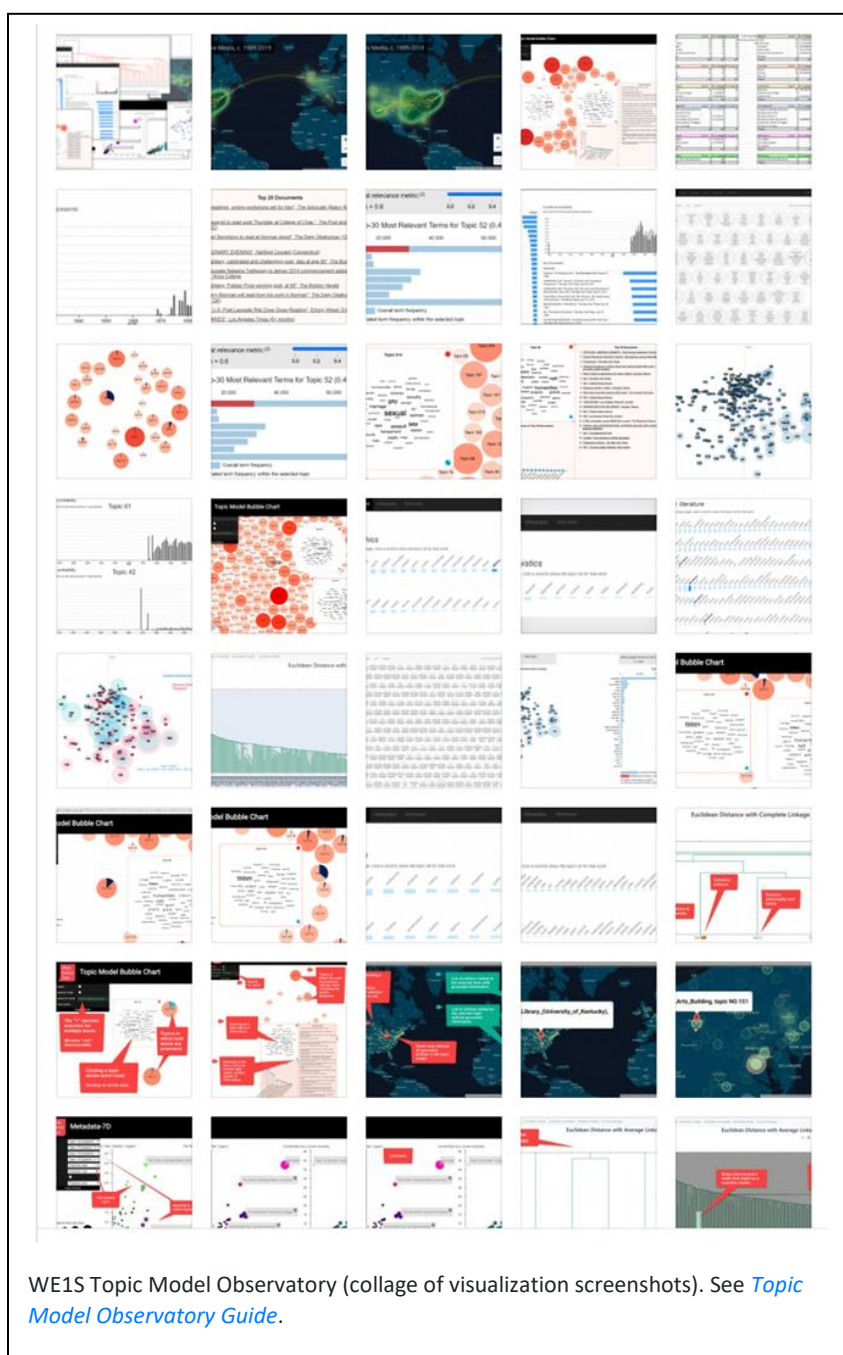
	25 topics	50 topics	100 topics	150 topics	200 topics	250 topics
Dfr-browser	✓	✓	✓	✓	✓	✓
TopicBubbles	✓	✓	✓	✓	✓	✓
pyLDavis	✓	✓	✓	✓	✓	✓
DendrogramViewer	✓	✓	✓	✓	✓	✓
Diagnostics	✓	✓	✓	✓	✓	✓

Available Visualizations

Ready-made interactive visualizations of topic models available from “[start page](#)” of Collection 1.

Innovations in Research Methods for DH

Below are some research methods in the digital humanities that are innovative in the WE1S project (some inspired by the STEM sciences and social sciences). These and other, more widely shared methods are described in more detail in the “cards” on the WE1S [Key Methods](#) page.



► Open and Reproducible Research Practices for the Digital Humanities

While the digital humanities have increasingly supported open-access policies and practices for the *reception* (providing access to readers) of scholarship, they are behind the sciences in supporting equivalent policies and practices for scholarship *production*. Few digital humanists follow “open science,” “open lab,” or “reproducible” principles of data research by documenting their data processing and analysis workflow in the manner of STEM-science data workflow systems such as [WINGS](#) or [Apache Taverna](#); documenting their corpora and models along the lines of “[data sheets](#)” or “[model cards](#)”; publishing their datasets under open licenses; or providing open access to their lab notes.

WE1S implemented a paradigm of “open, shareable, and reproducible methods in the digital humanities” in its work, and shared its policies and protocols for such. Specifically, the following are the elements of the open scholarship paradigm that WE1S demonstrates:

- Open-access publication on the WE1S website of datasets, models, visualizations, and documentary materials (e.g., reports) under a Creative Commons [Attribution-ShareAlike 4.0 International \(CC BY-SA 4.0\)](#) license.
- Open-access publication of WE1S lab notes and reports from its research teams (in “as is” form) in the Zenodo open science repository under a CC BY-SA 4.0 license ([example](#)).
- Open-source publication of WE1S’s technical platform under the [MIT License](#) through the project’s GitHub repositories as well as Zenodo open repository deposits.
- Open-access publication of the WE1S Topic Model Interpretation Protocol on the project website and also in the Zenodo repository, accompanied by open-access publication of results from WE1S’s use of the protocol in its research team lab notes deposited in Zenodo.

Production Environment

WE1S Collector – system for algorithmically collecting materials for analysis
(<https://github.com/whatevery1says/we1s-collector>)

- **Purpose: search / parsing.** Third-party dependencies:
 - BeautifulSoup4 4.5.1 (MIT) <https://pypi.org/project/beautifulsoup4/>
 - wsk 0.1.9, lexis-nexis-wsk by Yale DHLab (MIT) <https://github.com/YaleDHlab/lexis-nexis-wsk/blob/master/LICENSE>
- **Purpose: data / utilities.** Third-party dependencies:
 - get-docker-secret 1.0.0 (MIT) <https://pypi.org/project/get-docker-secret/>
 - lxml 3.6.4 (BSD) <https://pypi.org/project/lxml/>
 - pymongo 3.3.1 (Apache 2.0) <https://pypi.org/project/pymongo/>
 - python-dateutil 2.7.3 (Apache/BSD Dual) <https://pypi.org/project/python-dateutil/>
 - requests 2.20.0 (Apache 2.0) <https://pypi.org/project/requests/>
 - unicode 1.0.22 (GLP-2) <https://pypi.org/project/Unicodeode/>

WE1S Preprocessor – Python library that performs natural language processing (NLP) operations on collected data to prepare it for analysis (<https://github.com/whatevery1says/preprocessing>). The library contains methods that can be implemented on its own or within the Virtual Workspace described below.

- **Purpose: feature extraction.** Third-party dependencies:
 - ftty (MIT) <https://pypi.org/project/ftty/>
 - Illinois Wikifier (slightly modified MIT and request for citation)
https://cogcomp.seas.upenn.edu/page/download_view/Wikifier,
https://cogcomp.seas.upenn.edu/page/demo_view/Wikifier
 - nltk (Apache 2.0) <https://www.nltk.org/>
 - pandas (BSD) <https://pypi.org/project/pandas/>
 - pql (BSD 3-Clause) <https://github.com/alonho/pql>
 - python-fire (Apache 2.0) <https://github.com/google/python-fire>
 - python-ssdeep (LGPLv3+ and GPLv2) <https://github.com/DinoTools/python-ssdeep>
 - scikit-learn (BSD-3) <https://pypi.org/project/scikit-learn/>

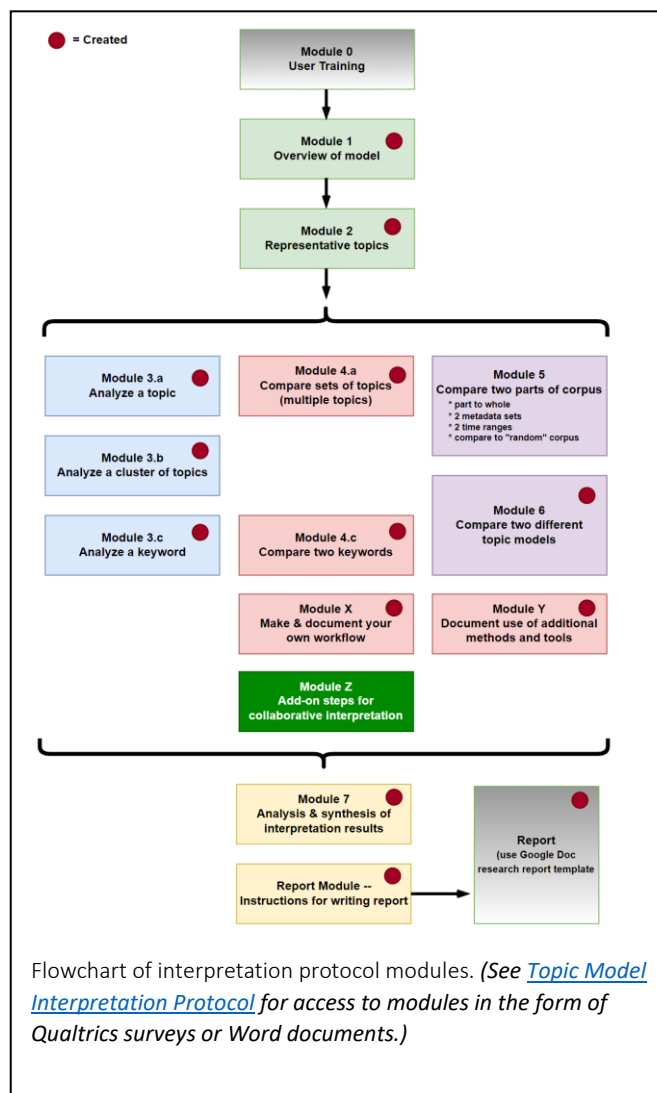
Part of documentation of open-source and open-access licenses for WE1S’s software and its underlying dependencies on open software by others.

► Topic Model Interpretation Protocol

Because complex data analysis can have a “black box” effect, researchers using machine-learning methods need not only to document technical workflows but also to make humanly understandable the steps in their whole workflow of finding understandable results from machine learning. Digital humanities research is rooted not just in data science but long-standing traditions of humanistic interpretative methods (hermeneutics). Digital humanists thus carry an even heavier burden of explaining for their audiences the combination of machine-to-human and human-to-human interpretive steps usually hidden in reports on the results of computational processing—e.g., the steps that allow researchers to read a topic model and reach credible conclusions from them. Yet there are currently no best practices in the digital humanities for explaining the workflow of data interpretation, let alone with attention to the act of humanistic interpretation.

WE1S developed a topic-model “interpretation protocol” that declares standard instructions and observation steps for researchers using topic models—a transparent, documented, and understandable process for the interaction between machine learning and human interpretation. The goal is not to assert a definitive topic-model interpretation process (because this will vary for different projects and materials), but to publish a paradigm that can be adapted, improved, and varied by others in the digital humanities. The WE1S Topic Model Interpretation Protocol consists of a modular set of survey-like questionnaires with instructions, required observation methods and waypoints, and reporting methods (utilizing the principles of iterative annotation derived from “[grounded theory](#)” in the social sciences). Used in flexible sequences or combinations, these modules step a researcher through an interpretation and documentation process that results in “grounded” discursive notes that can be used to produce research reports.

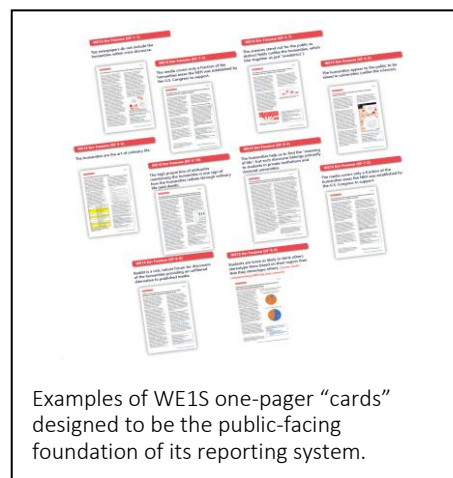
WE1S makes available its Topic Model Interpretation Protocol “as is” (customized for the project) in its original Qualtrics survey format (exported as QSF files for users who can import them into Qualtrics) and also in Word .docx format (using customized versions of Word’s “document properties” in each file to re-create the editable, repeated “running notes” of the original Qualtrics surveys). These files include instructions and references that are specific to the WE1S project. But they can be further developed and adapted by other projects to evolve a consensus practice of open, reproducible digital humanities research.



► “Cards” Reporting System

Inspired by the “cards” model used to explain nutrition, medical, and data science ([see explanation](#)), WE1S innovates for the digital humanities the method of initially reporting its materials, methods, tools, findings, and recommendations in one-page “cards” (previous to more extensive reporting formats). WE1S has 175 cards explaining:

- [Collections](#) (what the project collected from media discourse to study)
- [Methods](#) (how the project studied its collected materials)
- [Tools & Software](#) (platforms, interfaces, and scripts that the project created or borrowed)



- [Key Findings](#) (high-impact observations from WE1S’s research)
- [Calls-to-Action](#), [Calls-for-Communication](#), and [Research-to-Action Toolkits](#) (preliminary, sample recommendations for broadening and deepening the engagement of society with the humanities).

Research Tools

WE1S created (and/or adapted) a set of software tools for processing, analyzing, and visualizing topic models of large collections of texts. These tools are assembled into an open-source workflow platform called the WE1S “Workspace,” whose modules of related Jupyter notebooks and associated tools can be downloaded for deployment on other users’ computers. Other researchers can run these tools on WE1S’s collections of data about media coverage of the humanities (part of the way WE1S supports open and reproducible digital humanities). Or they can run the tools on their own texts by starting with the project’s Jupyter notebooks for creating a project and importing their materials.

Specifically, the platform (and component tools) that WE1S developed are as follows:

► WE1S “Workspace”

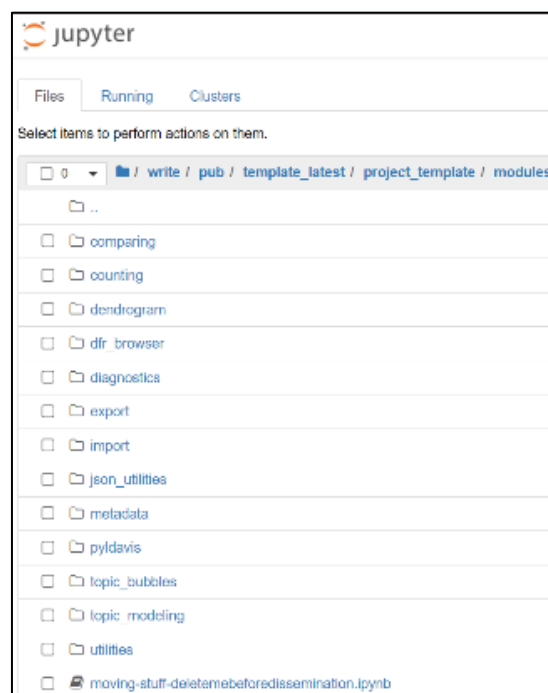
The WE1S Workspace is the ensemble of Jupyter notebooks and related scripts and resources that can be downloaded from the project’s GitHub repository or Zenodo repository. The ensemble can be used modularly or in a workflow series to collect, manage, analyze, topic model, visualize, and perform other operations on texts. (See “[Getting Started with the WE1S Workspace](#).”)

When initially downloaded as part of the WE1S Computing Environment, the Workspace includes a Jupyter notebook for initiating a project and installing modules of other Jupyter notebooks with supporting scripts and files. (See [Glossary](#) on *projects* and *modules*.) WE1S explains in brief cards the major parts of the Workspace on its [Key Tools & Software](#) page. Fuller explanations and step-by-step instructions are available in “[Getting Started with the WE1S Workspace](#)” and in the Jupyter notebooks themselves.

Specifically, the Workspace includes major modules for the following purposes:

Topic Modeling Tools (and related pre-processing and analysis tools)

Important modules in the WE1S Workspace include those for creating and running a topic modeling project—setting up the project; importing, exporting, or managing texts; pre-processing texts; performing various analyses (such as counting documents or terms); topic modeling; and conducting topic model diagnostics.



Example of a project template with modules of Jupyter notebooks for topic modeling, visualization, and other analysis functions.

Visualization Tools

The WE1S Workspace also includes Jupyter notebook modules for generating interactive visualizations of topic models. As mentioned above, WE1S calls its suite of visualization interfaces its [Topic Model Observatory](#). Available visualization modules include: [Dfr-browser](#), [TopicBubbles](#), [pyLDAvis](#), [DendrogramViewer](#), and [Diagnostics](#). (For descriptions and screenshots of many of these, see the [WE1S Topic Model Observatory Guide](#).)

Tools for Collecting from the Web & Social Media

WE1S also makes available [Chomp](#)—a set of Python tools (created for the project by Sean Gilleran) to find and collect text from webpages on specified sites that contain search terms of interest. Unlike other web scraping tools, Chomp is designed first and foremost to take a wide sweep, working at scale and across a variety of different platforms to gather material.

For collecting from Twitter, WE1S offers its [TweetSuite](#), a set of tools used to collect data from Twitter and prepare it for topic modeling. See also WE1S’s research blog post on the project’s methodology for collecting materials from Reddit: Raymond Steding, [“A Digital Humanities Study of Reddit Student Discourse about the Humanities.”](#)

► **Manifest Schema**

To document resources, tools, and workflow in a way that is both transparent to humans and computationally tractable, WE1S created a “manifest” schema for its work that can be adapted by other digital humanities projects. The WE1S manifest schema is a set of recommendations, examples, and validation tools for the construction of manifest documents for the WE1S project. WE1S uses the manifest schema to define metadata for individual documents, collections, sources, and corpora, as well as topic modeling projects.

WE1S manifests are [JSON](#) documents that describe resources. They can be used as data storage and configuration files for a variety of scripted processes and tools that read the JSON format. Manifests may include metadata describing a publication, a process, a set of data, or an output of some procedure. Manifests can also describe software tools, processes, and workflows, as well outputs such as result data, information visualizations, and interactive interfaces. Their primary intent is to help humans document and keep track of their workflow. (See [documentation](#) for the Manifest Schema and the schema’s [GitHub repository](#).)

Research Reporting

► WE1S Website (we1s.ucsb.edu)

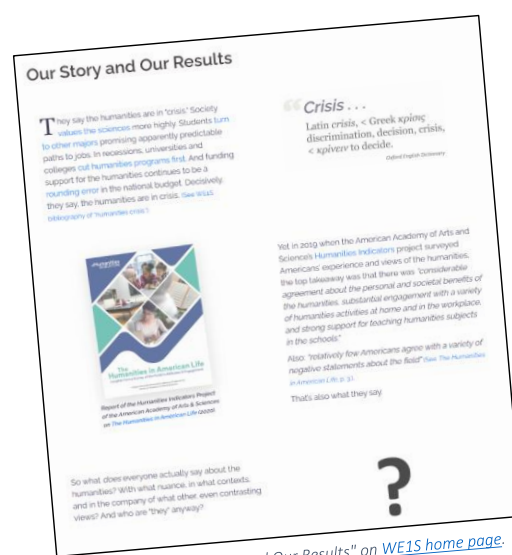
WE1S's primary platform for reporting on its project is its website at <https://we1s.ucsb.edu>. The website is maintained in a WordPress installation at UCSB (currently on an English Department server in a virtualized server environment, and soon to be migrated to the Pantheon.io platform). The main WordPress website is a richly developed resource containing, at the time of this writing, 271 web pages on its main site and 50 additional posts in the site's Research Blog. The website explains the project and provides access to WE1S's research (materials, methods, tools, reports and similar documents), recommendations, and other resources and outputs, along with extensive framing or explanatory documentation.

WE1S took care to write the top-level pages of the site in a public voice. The [“Our Story & Our Results” on the project home page](#) is thus cast in a voice for a public audience (and also a general scholarly audience). In essence, the project's voice says, “we” are talking to “you” in a shared frame of understanding. Similarly, each of the overview pages for major branches of the site has the same direct public voice—for example, the overview pages for [Research](#), [Research Findings](#), [Research Materials](#), [Research Methods](#), [Research Tools](#), and [WE1S Reports](#). Navigational orientation on the site is provided through drop-down menus under the tabs at the top of each page as well as through a [site map](#).

Within the platform of the WE1S website is a whole ecosystem of reporting — beginning with one-page “cards” and escalating to higher-order, larger reporting formats. Specifically, the parts of the WE1S reporting ecosystem nested in its website are as follows:

► Cards

As described above under “Innovative Research Methods,” WE1S created a system for reporting on the project beginning with one-page, plain-language “cards” (175 in total). See cards documenting the project's [collections](#), [methods](#), [tools](#), [findings](#), and recommendations ([Calls-to-Action](#) and [Calls-for-Communication](#)). Cards link to further resources, examples, and evidence, and are themselves referenced in higher-order reporting formats such as reports and overview pages on the project's website.



Beginning of “Our Story and Our Results” on [WE1S home page](#).



WE1S reporting system, starting with “cards” and moving up to long-form reports, posts, and articles.

► Area of Focus reports

WE1S researchers produced a series of reports during various phases of the project. Early on, the project produced [“Area of Focus” reports](#) identifying publication and database sources for journalistic materials for different nations and regions, alternative/indie sources, and news sources related to various racial, ethnic, and gender groups. Area of Focus reports also discuss sociopolitical and other contexts of journalism in an area as well as challenges for accessing a representative sample of materials.

Specific reports cover African News Sources, Canadian News Sources, Central and South American News Sources, Diverse Populations News Sources, European News Sources, Hispanophone Caribbean News Sources, Mexican News Sources, Middle Eastern News Sources, Political Orientation News Sources, U.S. Online News Sources, U.S. Spanish Language Sources, and United Kingdom News Sources.

► Scoping Reports

WE1S [“Scoping Reports”](#) explore relevant scholarly and methodological contexts for the project by reviewing research on journalism and media as well as on concepts such as the “edition,” “canon,” and “corpus linguistics.” These extensive reports provide the project with thoughtful analogies for creating a “representative” corpus of journalistic materials. Individual reports include those on The Edition, The Canon, Corpus Linguistics, Media Impact, Newspaper Studies, Newspaper Corpus Design and Representativeness.

► Human Subjects Research reports

Besides producing “cards” on surveys of students’ and others’ views of the humanities (see [examples](#) on the human subjects research overview page and [similar cards](#) on the Key Findings page), WE1S’s human subjects research teams also created an extensive set of “mini-reports” on the results of their surveys ([see mini-reports](#)). An example is mini-report [MR-9-1, “U. Miami Undergraduate Survey: Would You Advise a Friend to Major in the Humanities?”](#) There are 32 such mini-reports published on the WE1S website.

► Research Blog

WE1S researchers at all levels—faculty, postdoctoral scholars, graduate students, and undergraduates—published posts on the [Research Blog](#) on the project website, which currently includes 50 posts. Many of these posts are extensive research reports and/or reflections on the project’s methods, tools, and curricula. (See a curated, selected list of about 20 of the most substantial and representative such posts on this page of the Bibliography on the website: [Research Blog Posts \(selected\)](#).) In retrospect, the WE1S principals realized that providing research assistants and other researchers-in-training or early career scholars with a research blog platform was essential to the quality of the project. As is true of digital scholarship in general, much of WE1S’s data and technical work occupies a lower tier of intellectual activity equivalent (in an older media analogy) to filling a notebook with observations and winnowing notes reflecting on those observations. WE1S’s researchers found a productive, higher-level intellectual counterbalance to such work in writing for the project’s blog, which served as a “drafting board” for analytical, interpretive, and cultural commentary.

► “Lab” Reports and Notes

During its research, WE1S organized its participants into multiple teams working on different research and technical goals. From the summer of the project’s second year onwards, eight of these teams focused on using WE1S’s materials, methods, and tools to address the following aspects of how the

humanities appear in the media (while other teams concentrated on materials collection and technical development):

1. The humanities “crisis”
2. The humanities and social groups
3. The value of the humanities
4. The broader profile of the humanities in society (including the comparison of the humanities and sciences in the media)
5. The humanities and other kinds of media (including social media)
6. The impact of government and foundation humanities funding agencies
7. Students and the humanities (UCSB human subjects research team)
8. Students and the humanities (U. Miami human subjects research team)

Each of these teams produced notes and interim reports during its progress, which fed into the project’s final “cards” and other kinds of reporting.

To provide an evidentiary foundation for cards and other reporting, and to demonstrate an “open science” and “open lab” paradigm for the digital humanities, WE1S makes available the research materials of each team “as is” in a “lab” deposit in the Zenodo open-science repository. (*See below on WE1S deposits.*) For example, the “Lab-5” deposit (DOI [10.5281/zenodo.4831113](https://doi.org/10.5281/zenodo.4831113)) includes a folder of “Reports” and another folder named “Lab Notes” holding original notes, spreadsheets, datasets, images, and other material in their raw state. (If WE1S were to do its project over again, it would have learned from this process and created more standardized ways of keeping lab notes by borrowing from the “electronic lab notebooks” paradigm in the sciences.)

► Publications and Talks

Publications and talks to date by WE1S project participants that are centrally or partially about the project are cited in the [WE1S Bibliography](#). Three key articles are the following:

- Alan Liu et al., “What Everyone Says: Public Perceptions of the Humanities in the Media,” *Daedalus* 151, no. 3 (2022): 19–39, <https://www.amacad.org/publication/what-everyone-says-public-perceptions-humanities-media>.
- Lindsay Thomas and Abigail Droge, “The Humanities in Public: A Computational Analysis of US National and Campus Newspapers,” *Journal of Cultural Analytics* 7, no. 1 (2022): 36–80, <https://doi.org/10.22148/001c.32036>.
- Lindsay Thomas and Abigail Droge, “What We Learned About the Humanities from a Study of Thousands of Newspaper Articles,” *Journal of Cultural Analytics*, May 24, 2022, 139–44, <https://doi.org/10.22148/001c.35907>.

► WE1S Bibliography

To support its research and reporting, WE1S created an extensive online bibliography of 1,205 citations (at the time of this writing) in the form of a [Zotero](#) group library that is pulled automatically onto the project’s website through the [Zotpress](#) plugin for WordPress. The [WE1S Bibliography](#) became a substantial output in its own right, covering many categories and subcategories.

Below is the menu for the bibliography on the WE1S website. (See also the searchable [Zotero library](#) underlying the website bibliography.)

All works**[Works on the Humanities \(& Liberal Arts\)](#)** (show/hide)

(all) [Global Humanities](#) | [History of Humanities](#) | [Liberal Arts](#) | [Humanities and Higher Education](#) | [Humanities as Research Activity](#) | [Humanities Teaching & Curricula](#) | [Humanities and the Sciences](#) | [Medical Humanities](#) | [Public Humanities](#) | [Humanities Advocacy](#) | [Humanities and Social Groups](#) | [Value of Humanities](#) | [Humanities and Economic Value](#) | [Humanities Funding](#) | [Humanities Statistics](#) | [Humanities Surveys](#) | ["Crisis" of the Humanities](#)

HUMANITIES ORGANIZATIONS: [Humanities Councils \(U.S.\)](#) | [Government Agencies](#) | [Foundations](#) | [Scholarly Associations](#)

HUMANITIES IN: [Africa](#) | [Asia \(East\)](#) | [Asia \(South\)](#) | [Australasia](#) | [Europe](#) | [Latin America](#) | [Middle East](#) | *NORTH AMERICA:* [Canada - Mexico](#) - [United States](#) | [Scandinavia](#) | [United Kingdom](#)

[Works on Journalism & Media](#) (show/hide)

(all) [Lists of News Sources](#) | [Databases with News Archives](#) | [History of Journalism](#) | [Journalism Studies](#) | [Journalism Statistics](#) | [Journalism Organizations](#) | [Student Journalism](#) | [Data Journalism](#) | [Media Frames \(analyzing & changing media narratives using "frame theory"\)](#) | [Media Bias](#) | [Fake News](#) | [Journalism and Minorities](#) | [Journalism and Women](#) | [Press Freedom](#) | [News & Social Media](#)

[Works on Corpus Collection](#) (show/hide)

(all) [Corpus Representativeness](#)

COMPARISON PARADIGMS FOR IDEA OF A CORPUS: [Archives as Paradigm](#) | [Canons as Paradigm](#) | [Editions as Paradigm](#) | [Corpus Linguistics as Paradigm](#)

[Works on Data Science & Machine Learning \(including Topic Modeling\)](#) (show/hide)

(all) [Artificial Intelligence](#) | [Big Data](#) | [Data Mining](#) | [Data Notebooks \(Jupyter Notebooks\)](#) | [Data Visualization](#) (see also [Topic Model Visualizations](#)) | [Hierarchical Clustering](#) | [Interpretability & Explainability](#) (see also [Topic Model Interpretation](#)) | [Mapping](#) | [Natural Language Processing](#) | [Network Analysis](#) | [Open Science](#) | [Reporting & Documentation Methods](#) | [Reproducibility](#) | [Sentiment Analysis](#) | [Social Media Analysis](#) | [Statistical Methods](#) | [Text Analysis](#) (see also [Topic Modeling](#)) | [Text Classification](#) | [Wikification](#) | [Word Embedding & Vector Semantics](#)

Topic Modeling (all)

[Algorithms & Application Software](#) | [Applied Examples](#) | [Topic Model Interpretation](#) (see also [Interpretation & Interpretability](#)) | [Introductions & Tutorials](#) | [Labeling](#) | [Longitudinal over Time](#) | [Multilingual](#) | [Optimization](#) | [Topic Modeling & Sentiment Analysis](#) | [Topic Clusters](#) | [Visualizations](#)

[Works on Digital Humanities](#) (show/hide)

Selected DH research and resources bearing on, or utilized by, the WE1S project.

(all) [Distant Reading](#) | [Cultural Analytics](#) | [Sociocultural Approaches](#) | [Topic Modeling in DH](#) | [Non-consumptive Use](#)

[Works on Qualitative Analysis Methods](#) (show/hide)

(all) [Grounded Theory](#) | [Human Subjects Research](#)

[Searchable version of bibliography \(and developer resources\)](#)

Menu of the WE1S Bibliography.

Research Findings

► Overview

WE1S first wrote up its findings in the form of [Key Findings](#) cards published on its website. These are organized in the following thematic categories:

- **The Humanities “Crisis”** (How, where, and when do people think the humanities are in “crisis”?)
- **The Value of the Humanities** (How do people assign worth to the humanities in society?)
- **Broader Profile of the Humanities in Society** (How do the humanities bridge from disciplines in the academy to broader society?)
- **Humanities and Social Groups** (What is the relation of the humanities to racial, ethnic, and other social groups as understood in the media?)
- **Humanities and Science in Media** (How do the humanities and sciences compare in their public profile in the media?)
- **Students and the Humanities** (What do students think about the humanities in relation other fields, society, and life?)
- **The Humanities and Social Media** (How do the humanities appear on social media (Twitter and Reddit?)
- **The Humanities and Ordinary Life** (How pervasive are mentions of the humanities in the media as part of the common milieu of individual, social, and cultural existence?)
- **Humanities Funding** (What is the profile in the media of government agencies and private foundations that support the humanities?)

Building on these key findings, WE1S then wrote higher-level, synthetic essays. See, for example, Alan Liu et al., “What Everyone Says: Public Perceptions of the Humanities in the Media”; and Lindsay Thomas and Abigail Droge, “The Humanities in Public: A Computational Analysis of US National and Campus Newspapers” (*cited above*).

In summary form, WE1S’s primary findings about how the humanities appear to the public in the media are the following (described in language adapted from the *Daedalus* article by Liu et al.). These are important takeaways for communicating about and advocating the place of the humanities in society:

- a. The mindshare of the humanities in the media is very small.
- b. This small mindshare means that the “humanities crisis” is not a crisis in the view, or even awareness, of larger society.

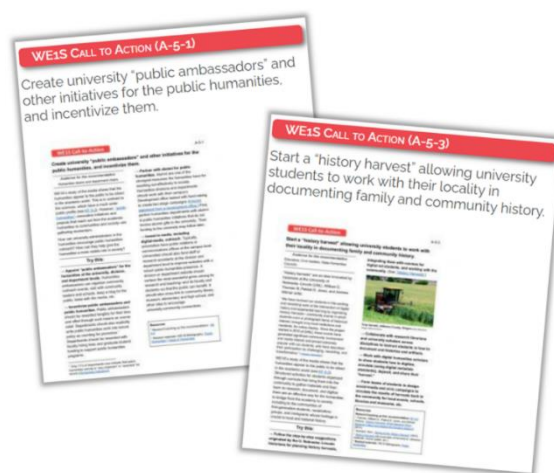
- c. The humanities (by contrast with the sciences) appear to be siloed in the academy.
- d. The humanities seem to be all about institutional infrastructure.
- e. Social media siloes the humanities in the academy too.
- f. The humanities communicate poorly about what they do because they appear in the media at a “double communicative” remove from the public (stories about “Professor giving talk on,” with the actual research hidden in the background).
- g. The humanities radiate through everyday and ordinary life.
- h. There is little public discussion of underrepresented social groups in relation to the humanities.
- i. The humanities are captive to proprietary aggregators who black-box or throttle the materials needed to conduct big-data analytics of the humanities in public.

Recommendations

Based on its research into the representation of the humanities in public media, WE1S created example recommendations for how different groups can put such research to use in actions and communications that reframe society’s engagement with the humanities (and the humanities’ engagement with society). This effort is at an early stage — just recommendations that are proofs-of-concept to start with. WE1S hopes that others will use the recommendations it has so far made as paradigms for ones they can create for their university, community, state, and nation.

WE1S presents its example recommendations in the form of [Call-to-Action](#) and [Call-for-Communication](#) “cards.” These are short, everyday-language recommendations or provocations somewhat similar to the “[one-pagers](#)” created by the American Academy of Arts & Science’s Humanities Indicators project or the cards sparking public engagement with the humanities designed by some U.S. state Humanities Councils ([examples from Humanities Montana](#)).

WE1S’s recommendation cards can be combined modularly with the research cards for its “Key Findings,” “Key Collections,” “Key Methods,” and “Key Tools” to create what the project calls complete “[Research-to-Action Toolkits](#).” Though WE1S only just started creating whole kits, it thinks they are an important paradigm because they demonstrate the core principle of the project: that strong advocacy for the humanities in public life needs to be driven by strong research, and vice versa. WE1S hopes that others will combine the project’s research (findings, collections, methods, tools) with action and communication plans they themselves create to make their own “Research-to-Action Toolkits” for engaging combinations of audiences (such as students, educators, journalists, community leaders, or legislators) in rethinking the role of the humanities in society.



Examples of Call-to-Action cards — “Create university public ambassadors” and “Start a history harvest”

Repository Deposits

(for openness, reproducibility, and sustainability)

WE1S practices principles of research sustainability and openness by depositing its data (datasets and “collections” with their “project” files), tools, and lab notes in the international [Zenodo open-science repository](#). Zenodo tracks deposit versions, assigns a DOI to each, and also provides a “conceptual DOI” for version series (resolving on request to the latest version). WE1S’s deposits are in the form of compressed files. These materials are assigned open licenses—either the [Creative Commons Attribution Share Alike 4.0 International](#) license (for data or documents) or [MIT License](#) (for code).

WE1S also distributes its code resources — for the WE1S computing “Workspace” (tools and workflow)— in GitHub repositories (<https://github.com/whatevery1says>).

The [“WE1S Repositories & Deposits”](#) page lists the project’s data, documents, and tools deposits.

The screenshot shows the Zenodo interface for a specific dataset. The header includes the Zenodo logo, a search bar, and navigation links for 'Upload' and 'Communities'. The user profile 'contact@we1s.ucsb.edu' is visible in the top right. The main content area displays the title 'Collection 1: U.S. News Media, c. 1989-2019 (WE1S core collection of articles mentioning humanities)' and the creator 'WhatEvery1Says (WE1S) Project'. A detailed description follows, explaining the collection's scope and the word frequency analysis. A list of the top 20 news sources is provided, including the New York Times, Washington Post, and Los Angeles Times. The page also shows the 'Kinds of Sources (by Tags)', 'Collection Metadata' (creation date, registry ID, data sources), and a 'Suggested Citation for Collection'. On the right sidebar, there are statistics for '18 views' and '1 download', an 'OpenAIRE' badge, and metadata including the 'Publication date' (June 20, 2019), 'DOI' (10.5281/zenodo.4902187), 'Keywords' (Humanities, Journalism), and 'License' (Creative Commons Attribution Share Alike 4.0 International). A 'Versions' section at the bottom right shows 'Version 1' as the only entry. At the bottom, a table lists the files, showing a single file named 'home/joyan/view/exports/20190620_2238_us-humanities-all-no-rediff.tar.gz' with a size of 691.5 MB.

Example of WE1S deposit in Zenodo.