FRAMES OF MISINFORMATION AND EXTREMISM:

DISCUSSION-FORUMS AND SOCIAL-MEDIA

DOCUMENTATION

Richard Lee Rogers

Youngstown State University

Version 1.0.0

December 20, 2022

**TABLE OF CONTENTS**

[AIMS 3](#_Toc122441538)

[DESCRIPTION 4](#_Toc122441539)

[Background 4](#_Toc122441540)

[Significance 4](#_Toc122441541)

[PROCEDURES 5](#_Toc122441542)

[HUMAN SUBJECTS ETHICS 7](#_Toc122441543)

[Risk to Participants 7](#_Toc122441544)

[Adequacy of Protection Against Risks 8](#_Toc122441545)

[CRITERIA FOR INCLUSION OF A FORUM 9](#_Toc122441546)

[MAXIMUM NUMBER OF FORUMS PER FORUM TYPE 10](#_Toc122441547)

[FRAME MAP 12](#_Toc122441548)

[APPENDIX A: PROGRAM DIRECTORY 13](#_Toc122441549)

[Terminology 13](#_Toc122441550)

[Weekly Exports and Scrapes 13](#_Toc122441551)

[NOT YET INCLUDED: 4chan-r9k (Archived Moe) 13](#_Toc122441552)

[Qkun QResearch 13](#_Toc122441553)

[Reddit Community Health Benchmarks 13](#_Toc122441554)

[Monthly Background Exports 14](#_Toc122441555)

[Telegram 14](#_Toc122441556)

[Background Programs 14](#_Toc122441557)

[Reddit FOME Programs 14](#_Toc122441558)

[Incel subreddits 14](#_Toc122441559)

[Telegram: Python 14](#_Toc122441560)

[Telegram: Telegram Export Tool 15](#_Toc122441561)

[Telegram Adjustments 15](#_Toc122441562)

[Regular Monthly Scrapes 15](#_Toc122441563)

[Win Board: Patriots.Win 15](#_Toc122441564)

[Stormfront 15](#_Toc122441565)

[Incels.is 16](#_Toc122441566)

[Hexbear (current not included in study) 16](#_Toc122441567)

[Ron Paul Forums (Liberty) 16](#_Toc122441568)

[Win Board: GreatAwakening.Win 16](#_Toc122441569)

[Raddle.me 17](#_Toc122441570)

[AboveTopSecret 17](#_Toc122441571)

[8kun Other 17](#_Toc122441572)

[White Biocentrism 17](#_Toc122441573)

[Christogenea 18](#_Toc122441574)

[End Chan: News 18](#_Toc122441575)

[End Chan: Pol 18](#_Toc122441576)

[PinkPill.co (website currently down) 18](#_Toc122441577)

# AIMS

The aims of the FOME project initially included the following objectives regarding the study of:

1. Quantify the volume of public discourse related to misinformation, extremism, and conspiracism.
2. Document trends related to the spread of misinformation, extremism, and conspiracism among potential security threat groups.
3. Identify underlying ideologies used to legitimate misinformation, extremism, and conspiracism.
4. Investigate propositions on the relationship between misinformation, extremism, and conspiracism.
5. Describe the mainstreaming of misinformation, extremism, and conspiracism.
6. Monitoring community health misinformation.

# DESCRIPTION

## Background

Frames of Misinformation and Extremism (FOME) is a big data project identifying the ideological and discursive legitimations underlying the spread of misinformation, extremism, and conspiracist groups in American society. Most of the work up to this point has been tied to material found in publicly available social media and discussion forums known for hosting these points of view. To date, I have recovered 87 million social-media and discussion posts going back to April 2018 from 13 platforms, and my earliest analyses are built around a group of 13.3 million posts with at least 50 words, i.e., enough words to elaborate a point of view. These numbers do not include YouTube transcripts collected by my students and me.

Misinformation, extremism, and conspiracism can be viewed as forms of stigmatized knowledge within the information ecosystem (Barkun, 2002; 2013).Extremism is defined as groups whose goals, methods, or beliefs fall outside the normal spectrum of political processes and ideologies. Some of the Extreme-Right movements captured in the study are well known in the light of recent events, e.g., the Proud Boys, sovereign citizens, neo-Nazis, MAGA and America First groups, and the Ku Klux Klan. The study also includes Extreme-Left forums promoting ideas related to socialism, anarchism, antifascism, and anarchocapitalism. Conspiracy sites (or more appropriately, conspiracy and fringe sites) promote alternative beliefs about science and history (ancient civilization myths related to Atlantis, Lemuria, and Tartaria; flat-earth and young-earth theories; belief in UFOS and the paranormal; health misinformation; we live in a simulation or matrix; the Galactic Federation is sending us messages) and a variety of political and social conspiracy theories, many of them about elite cabals or coverups (e.g., QAnon, Biden in an actor, Kennedy assassination theories, fake moon landing). By definition, the platform does not examine content formally tied to a major political party with the exception of one platform associated with libertarian Republicans, where over one-third of the content was identified as extremist or conspiratorial in nature.

## Significance

Identifying the ideological underpinnings of misinformation, extremism, and conspiracism is regarded as important to understanding current threats to the stability of Western democracies, active shootings, and the open attacks on expert knowledge. The importance of the study lies in its examination of discourse related to radicalization supporting collective violence (e.g., January 6 Capitol insurrection) and individual criminal acts (e.g., active shooters). This discourse and some of these platforms have been tied directly to activities such as the organization of the January 6 Capitol insurrection and antivaccination protests. Lone wolves also frequent these sites, and several active shooters have placed their “manifestos” in web discussion forums.

# PROCEDURES

Data collection and analysis can be divided into 5 stages:

1. *Extraction:* Downloading the posts, usually into Excel format. For some platforms, this task is a simple download or export, but for others it is a multistage, crawling and scraping process to identify recent additions to the site while avoiding the time and memory it would take to sort through millions of web pages on a single platform. Most platforms are crawled and scraped using Screaming Frog SEO Spider. Telegram material is downloaded using a modified Python program from ISD Global in London. Reddit posts are taken from a Pushshift, a third-party aggregator, using a modification of a program in the public domain. Excel is appropriate for stage because of its capacity to handle large strings, an upgrade from earlier versions of Excel.
2. *Preprocessing:* Conversion of files into Stata data sets for each platform. Stata is an appropriate software package because of its string functions for manipulating text and its capacity to handle unlimited length variables. (I once put the transcript from an entire 4-hour webcast into a single cell of a Stata data set.) Also, Stata is available at the Ohio Supercomputer Center on the Owens cluster, which is relevant to the next stage.
3. *Processing:* Preparing for coding and the coding itself. At this point all processing is done through the Ohio Supercomputer Center due to the size of the data sets and the extensive processing. Data is combined into a series of monthly files. Words are decapitalized, punctuation is removed, and Unicode formatted text is cleaned up. Then the posts are coded in the manner summarized below.
4. *Reporting and Research:* Getting out the results. I am modestly succeeding in getting out topline numbers through my personal website and social media. The area where I am faltering is publishing Research results, which is why this aspect of the project is the focus of this Research Professorship request.

The coding of items uses common text mining techniques. Grounded theory is modified to accommodate the large amount of data processed. Conceptually the methodology is integrated into the language of frame analysis, i.e., social stances are positioned through discursive frames, e.g., fringe beliefs, politics, religion, science, values. Themes are identified using an open-coding technique that constructs binary ngrams (dichotomous variables) indicating discussion of an important theme in the text. I use three types of open codes:

* Simple coding: The theme is identified solely by key words, phrases, and word stems. These words can be “seeds” based on outside sources or emerge from audits of samples of the text.
* Qualified coding: The theme is identified using the key words plus other words in the text. For example, “millennial” has a different meaning depending on whether one is discussing generational cohorts or end-times prophecies.
* Forum coding: The key word is assigned a meaning based on the forum. (This is limited.) For example, the word “Jew” and its derivatives are always treated as antisemitic when they appear on conspiracy boards or extreme-right forums where the mere mention of the word is nearly always intended as a slur, but in some Christian or political forums, they are used in conjunction with non-pejorative historical references (“Jesus was a Jew”), a desire to celebrate “Jewish festivals” (e.g., some Christian groups celebrate Passover or teach that Jesus will return on the Feast of Trumpets), or political support for the nation of Israel (“the Jewish state”).

The open codes are then grouped into subframes (axial coding), and the final frames are built on the aggregation of the subframes (selective coding).

The results are regularly reviewed for untagged posts (tag = at least one frame identified) and to make sure the coding is mostly correct. The tag rate is about 96%, i.e., 96% of posts have at least one frame identified.

# HUMAN SUBJECTS ETHICS

To date, FOME has taken material from 13 platforms and collected recovered approximately 93 million posts dating back to April 2018, and analysis is based on 13.7 million posts with at least 50 words.

Collection and use of FOME data complies with CITI Training standards and EU Data Protection Regulations (Hagen et al., 2019; Martinez, 2019). Practically speaking, this means balancing the needs of the researcher with the rights of the content creators and platform, including protection of their copyrights, largely in terms of the following principles:

1. We are permitted to scrape social media posts from social media or web-based locations that are public, which provide a tool for exporting data, or for which we have permission either expressly or through the issuance of an authorization code. Under no circumstances do we use material regarded as private.
2. We do not cite social-media posts directly or reveal the identities of the content creators unless they are public figures.
3. We will provide a deidentified data set to any researcher requesting it for analysis, but we will not share personal information or the text of social media, in the latter case to protect the copyright privileges of others. Because of an agreement with a social media company, the data will not be placed in the public domain but must be requested.
4. We will share publicly our own program code to assist others assessing the credibility of our research or trying to replicate our results.

## Risk to Participants

The research study will be able to access existing data and data processing capabilities of the Frames of Misinformation and Extremism project, an already existing initiative operating separate from this proposed study and led by Richard Rogers, Co-Principal Investigator of this study. This project is approved as exempt research under YSU IRB protocol #2022-191. FOME is a big-data approach to qualitative analysis that uses text-mining techniques to scrape and classify posts gleaned from social media and web-based discussion forums on platforms known to host concentrations of material that could be classified as misinformation, extremism, and conspiracism.

The focus is on the text, not the individual, and there is no direct contact with individuals. Individuals are not assigned to groups in an experimental sense. Text may be classified as part of a “frame” based on key words, phrases, and word stems. To date, the project has recovered

We have recovered 93 million social-media and discussion posts going to April 2018 and for most analyses use a group of 13.7 million that contain at least 50 words of text.

There are no interventions associated with the FOME project. All material has been placed in public view or used in conjunction with permissions. Although individuals have voluntarily placed this information in public view, we follow a generally accepted practice of not speaking about or seeking to identify individuals except for those who are public figures.

This research has been determined to be minimal risk as part of YSU IRB protocol #2022-191. Any personal information about individuals in the social media and discussion posts has already been voluntarily disclosed by the individual in a public forum.

## Adequacy of Protection Against Risks

Informed consent does not apply in the traditional sense because individuals have already put information in public view. Nonetheless, current best practices in the use of social media for research recognize that (1) content creators did not intend for any personal information to be disclosed in other forums and (2) content creators and platforms retain copyrights over this material. CITI Training currently permit the use of this kind of material (Martinez, 2019), and we follow EU Data Protection Regulations (Hagen et al., 2019). Practically speaking, this means balancing the needs of the researcher with the rights of the content creators and platform, including protecting their copyrights, largely in terms of principles already described above.

There is no direct contact with these individuals. There is no direct risk.

# CRITERIA FOR INCLUSION OF A FORUM

It is not realistic to cover all discussion forums with the time and tools available. The study is built around a relatively stable purposive sample of discussion forums that are select based on their volume, influence, and contribution to the ideological diversity of the project. The inclusion of many well known platforms and discussion forums gives credibility to the findings. The relative stability of the sample allows for the ongoing monitoring of trends. For these reasons, historical trends in study results are analogous to stock market indices in that they are a meaningful but imperfect measure of market activity and relatively stable among a group of companies meeting specific criteria.

# MAXIMUM NUMBER OF FORUMS PER FORUM TYPE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Conspiracy/Fringe & QAnon Forums** | **Patriot / Military Forums** | **Niche Forums** | **Other Forums** |
| Forum Type | Conspiracy/Fringe, QAnon | Patriot / Military | Extreme Right, Religion | Extreme Left, Health Misinformation, Incels/Femcels/r9k, News/Politics, Political Right |
| Volume Minimums | 10,000 subscribers or  100 posts per month | 10,000 subscribers or 100 pots per month | 1,000 subscribers or 30 posts per month | 1,000 subscribers or 30 posts per month |
| Usability (50 words or more) | 100 posts per month (a little more than 3 posts per day) | 100 posts per month (a little more than 3 posts per day | 30 posts per month (about 1 post per day) | 30 posts per month (about 1 post per day) |
| History Minimum | 13 months | 13 months | 13 months | 13 months |
|  |  |  |  |  |
| Maximum Number of Forums  (broadcast and chat combined to be a single forum) | 30 | 20 | Religion Christian Prophecy: 25  Extreme-Right White Aryan: 15  Extreme Right Personalities: 10  Others: 10 per niche | Health Misinfo: 20  Incels/Femcels/r9k: 5 |
| Exceptions to Volume and Usability | 5 exceptions: Forums must exceed 250,000 subscribers | 5 exceptions: Forums must exceed 250,000 subscribers | Any exception permitted, but no more than 50% of niche maximum | Extreme Left, Health Misinfo, Incels/Femcels/r9k: No more than 50% of niche maximum  Other forum types: Must exceed 100,000 subscribers |
| Niches | None | None | Extreme-Right niches: Kingdom of America, National Justice Party, Personalities, Proud Boys, Regionalist & Secessionist, SovCit, White Aryan (inc. KKK and neo-Nazi), Other  Religion niches: General, Protestant/Nondenom,Prophecy, Catholic, New Age/Esoteric/Eastern, Other |  |

Discontinuation: A channel is unless its average usability falls below minimums the channel is a designed exception.

Replacement: A larger channel may replace a smaller channel or an exception. When this occurs, the two time lines will be compared on a monthly basis, and the history of the smaller channel will be included until the larger channel surpasses it.

Review date

Notes:

Religion references that convey personal belief but do not define channel content are not counted as religious.

Religion (Prophetic) references must be Christian.

Spiritual warfare references not counted as military chat.

End times news is not regarded as news unless in title.

# FRAME MAP

# APPENDIX A: PROGRAM DIRECTORY

## Terminology

Export: Text downloaded through third-party tool.

Scrape: Text read from a web site or social media application.

Parsing: Exported and scraped text cleaned of attached html.

Preprocessing: Text converted to the format from which the code will be processed.

## Weekly Exports and Scrapes

### NOT YET INCLUDED: 4chan-r9k (Archived Moe)

1. Scrape a manually created list of urls for a selected time period. Check for the need to do repeats, which if done, should be added to the Output File on completion on repeating the program. Software: Screaming Frog. Configuration File: …\Trend\_Studies\Programs\1-Extraction\achived-moe\_r9k\_1\_scrape.seospiderconfig. Output File:

### Qkun QResearch

A complete, historical page catalog including dates was made for May 2022. The page catalog and scrape for May are complete. A quality audit of the history files has not been conducted.

1. Catalog users. Update Custom Search before running. Note Exclude. Software: Screaming Frog. Config File 8kun-qresearch\_1\_user-catalog.seospiderconfig. Save Custom Search.
2. Scrape the QResearch site. Because of large number of posts on some pages, it is necessary to break the scrape into parts and assemble in Step 3.
   1. Software: Screaming Frog. Config File: 8kun-qresearch\_2\_scrape\_author.xlsx.
   2. Software: Screaming Frog. Config File: 8kun-qresearch\_2\_scrape\_datetime.
   3. Software: Screaming Frog. Config File: 8kun-qresearch\_2\_scrape\_text.
3. Transfer scraped results to OSC and then parse. Software: Stata. Program: 8kun-qresearch\_3\_parsed.do.
4. Finalize data set for time period. Software: Stata. Program: 8kun-qresearch\_4\_month.do.

### Reddit Community Health Benchmarks

1. Extract from Pushshift. Execute from Extraction folder. Software: Python. Program: reddit\_1\_chealth.py. Transfer results to OSC.
2. Parse results. Software: Stata. Program: reddit\_chealth\_2\_parse.do.
3. Create monthly file. Software: Stata. Program: reddit\_chealth\_3\_month.do.
4. Clean up for coding. Software: Stata. Program: reddit\_chealth\_4\_working.do.
5. Run health frame on benchmark files. Software: Stata. Program: reddit\_chealth\_5\_program.do.

## Monthly Background Exports

### Telegram

Approach #1; Python

This crawl is a single program taking a Telegraph export and converting it to a preprocessed Stata file. The use of Python to export through the Telegram API is still experimental. The difficulty is the slow speed, which can produce processing errors, especially with large databases.

1. Telegram software

## Background Programs

### Reddit FOME Programs

Includes alt-reality, conspiracy, extreme left, and extreme right. Tartaria folded into alt-reality starting 202211.

1. Run extract. Execute from Extraction folder. Software: Python. Program: reddit\_1\_fome.py. Copy results from Program to Download folder.
2. Parse. Software: Stata.
   1. Alt-reality. Program: reddit\_2\_altreality.do.
   2. Conspiracy. Program: reddit\_2\_conspiracy.do. Transfer results to OSC.
   3. Extreme Left. Program: reddit\_2\_extremeleft.do.
   4. Extreme Right: Program: reddit\_2\_extremeright.do.
3. Preprocessing. Software: State. Program: reddit\_3\_fome.py. Copy to
   1. Software: Python. Program: reddit\_3\_altreality. Transfer results to OSC.
   2. Software: Python. Program: reddit\_3\_extremeleft. Transfer results to OSC.
   3. Software: Python: Program: reddit\_3\_extremeright. Transfer results to OSC.

Notes: Reddit extracts use the Pushshift API, which occasionally has errors.

### Incel subreddits

1. Extract from Pushshift. Execute from Extraction folder. Software: Python. Program: reddit\_1\_incels.py. Move results to Download folder.
2. Parse. Software: Stata. Program: reddit\_2\_incels.do.
3. Preprocessing. Software: Stata. Program: reddit\_3\_incels.do. Transfer results to OSC.

### Telegram: Python

1. Download files marked “App” through Telegram application. Software: Python. Program: telegram\_1\_parse\_yyyymm.do. Transfer results to OSC.
2. Preprocessing. Software: State. Program: telegram\_2\_month.do.

### Telegram: Telegram Export Tool

1. Export files marked “Export” through Telegram Export tool.
2. Combine html files and convert to text file. Software: txtcollector.
3. Parse. Comment out “App” files if that section completely separately. Software: Python. Program: telegram\_1\_parse\_yyyymm.do. Transfer results to OSC.
4. Preprocessing. Software: State. Program: telegram\_2\_month.do.

### Telegram Adjustments

Depending on file, follow steps for Telegram: Python or Telegram: Telegram Export Tool as appropriate.

## Regular Monthly Scrapes

### Win Board: Patriots.Win

1. Catalog pages. Software: Screaming Frog. Config File: patriots-win\_1\_page-catalog
2. Select URLs to be used. Software: Screaming Frog.
   1. Create list of comments to scrape. Program: patriots-win\_2\_comments-href.seospiderconfig.
   2. Expand list for users who recent month of comments exceeds six pages. Program: patriots-win\_2\_comments\_href\_big.seospiderconfig. Combine with Step 2a to create composite file for comments.
   3. Create list of urls containing posts. patriots-win\_2\_posts-href.seospiderconfig.
3. Create list of post urls to scrape. Software: Stata. Program: patriots-win\_3\_posts-masterfile.do. Save Custom Extract.
4. Scrape posts
   1. Type=comment. Software: Screaming Frog. Config File: patriots-win\_4\_comments\_scrape.seospiderconfig. Save as csv file and convert to xlsx.
   2. Type=post. Software: Screaming Frog. Config File: patriots-win\_4\_posts\_scrape.seospiderconfig. Save as csv file and convert to .xlsx.
5. Transfer scraped results to OSC. Parsing and preprocessing. Software: Stata. Program File: patriots-win\_5\_parse.

Notes: Patriots.Win, also known as TheDonald.win, is a sprawling site because there are no folders. Nonethless, the page links are stable, so an exclude statement based on previous searches can be used to limit the search. History: A complete user list was created for 202206.

### Stormfront

1. Manually prepare folder in Excel file with a list of the most recent pages for each forum. Save as manual-catalog.xlsx.
2. Catalog forums. Software: Screaming Frog. Config File: stormfront\_2\_forum-catalog.seospiderconfig. Save Custom Search and Custom Extract as forum-catalog\_date.xlsx and forum-catalog\_href.xlsx.
3. Catalog pages from folders within activity in the most recent month. Software: Stata. Program: stormfront\_3\_page-catalog.do. Do not need to save file.
4. Scrape. Software: Screaming Frog: Config File: stormfront\_4\_scrape.
5. Preprocessing. Software: Screaming Frog. Program: stormfront\_5\_parse. Transfer results to OSC.

### Incels.is

1. Catalog forums. Software: Screaming Frog. Config File: incels-is\_1\_forum-catalog. Save Custom Search and Custom Extract files as forum-catalog\_date and forum-catalog\_href.
2. Create list of pages to scrape. Software: Stata. Program: incels-is\_2\_threads.do.
3. Scrape. Software: Screaming Frog. Config File: incels-is\_3\_scrape.seospiderconfig.
4. Parse. Software: Stata. Program: incels-is\_4\_parse. Also produces monthly file for upload to OSC.

Notes: The historical timeline was completed back through April 2018 in May 2022.

### Hexbear (current not included in study)

1. List communities currently on Hexbear. Software: Screaming Frog. Program: hexbear\_1\_community-catalog.seospiderconfig.
2. Scrape. Software: Screaming Frog. Program: hexbear\_2\_scrape.seospiderconfig.
3. Parse. Software: Stata. Program: hexbear\_3\_parse.do. PROGRAM CURRENTLY WORKS WELL FOR INITIAL SUBMISSION BUT NOT FOR RESPONSE POSTS.

### Ron Paul Forums (Liberty)

1. Catalog forum display. Software: Screaming Frog. Config File: ronpaul\_1\_forumdisplay-catalog.seospiderconfig.
2. Catalog forums. Software: Screaming Frog. Config File: ronpaul\_2\_forum-catalog.seospiderconfig.
3. Create list of urls to scrape. Software: Stata. Program: ronpaul\_3\_threads.do.
4. Scrape. Software: Screaming Frog. Config File: ronpaul\_4\_scrape\_seospiderconfig.
5. Parsing and preprocessing. Software: Screaming Frog. Config File: ronpaul\_5\_parse.

### Win Board: GreatAwakening.Win

1. Software: Screaming Frog. Config File: greatawakening-win\_1\_user-catalog.
2. Create list of URLs.
   1. Software: Screaming Frog. Config File: greatawakening-win\_2\_comments-href.
   2. Software: Screaming Frog. Config File: greatawakening-win\_2\_posts-href.
3. Create post masterlist. Software: Stata. Program: greatawakening-win\_3\_post-masterlist.

4. Scrape.

a. Software: Screaming Frog. Config File: greatawakening-win\_4\_comments\_scrape.

b. Software: Screaming Frog. Config File: greatawakening-win\_4\_posts\_scrape.

5. Transfer scraped results to OSC. Parse. Software: Stata. Program: greatawakening-win\_5\_parse.

The catalogs of user and pages for 202205 are possibly complete histories and need verification.

### Raddle.me

1. Catalog forums. Software: Screaming Frog. Config File: raddle-me\_1\_forum-catalog.seospiderconfig. Save Custom Search.
2. Catalog pages. Software: Screaming Frog. Config File: raddle-me\_2\_page-catalog.seospiderconfig. Save Custom Search.
3. Select urls for time period from Step #3 results and scrape. Software: Screaming Frog. Config File: raddle-me\_3\_scrape.seospiderconfig.
4. Parse. Software: Stata. Config File: raddle-me\_4\_parse.
5. Preprocessing. Software: Stata. Config File: raddle-me\_5\_month. Transfer results to OSC.

### AboveTopSecret

1. Catalog forums. Software: Screaming Frog. Config File: abovetopsecret\_1\_forums\_catalog.seospiderconfig. Save Custom Search and Custom Extract as forum-catalog\_date.xlsx and forum-catalog\_href.xlsx.
2. First attempt to catalog pages. Software: Screaming Frog. Config File: abovetopsecret\_2\_catalog.
3. Final page catalog. Software: Screaming Frog. Config File: abovetopsecret\_3\_page-catalog.
4. Create list of pages in preparation for scrape. Software: Stata. Program: abovetopsecret\_4\_threads.do.
5. Scrape. Software: Screaming Frog. Program: abovetopsecret\_5\_scrape.
6. Parse. Software: Stata. Program: abovetopsecret \_6\_parse.do.
7. Preprocessing. Software: Stata. Program: abovetopsecret\_7\_month.do. Transfer results to OSC.

A history file was created in 202205. Only 202205 was run using the history data.

.

### 8kun Other

1. Catalog pages. Software: Screaming Frog. Config File: 8kun-other\_1\_page-catalog.seospiderconfig.
2. Scrape. Software: Screaming Frog. Config File: 8kun-other\_2\_scrape.seospiderconfig. Transfer results to OSC.
3. Parsing and preprocessing. Software: Stata. Program: 8kun-other\_3\_month.

### White Biocentrism

1. Catalog forums and check for recent activity. Software: Screaming Frog. Config File: whitebiocentrism\_1\_forum-catalog.seospidercatalog.
2. Unpack links on selected pages. Software: Stata. Config File: white-biocentrism\_2\_href.do.
3. Scrape. Software: Screaming Frog. Config File: whitebiocentrism\_3\_scrape.seospiderconfig.
4. Parse. Software: Stata. Config File: white-biocentrism\_4\_parse.do.
5. Preprocessing. Software: Stata. Config File: white-biocentrism\_5\_month.do.

### Christogenea

1. Catalog pages. Software: Screaming Frog. Program: christogenea\_1\_catalog.seospiderconfig.
2. Extract html pages. Software: Screaming Frog. Program: christogenea\_2\_extract.seospiderconfig. Download results to Extract file.
3. Convert html pages to text. Software: htmlastext.
4. Combine text pages into single document. Software: txtcollector.
5. Parse. Software: Stata. Program: christogenea\_5\_posts.do.
6. Preprocessing. Software: Stata. Program: christogenea\_6\_monthly.do.

### End Chan: News

1. Scrape. Software: Screaming Frog. Program: endchan\_news\_1\_extract.seospiderconfig.
2. Parse and preprocessing. Software: Stata. Program: endchan\_2\_news.do.

### End Chan: Pol

1. Scrape. Software: Screaming Frog. Program: endchan\_pol\_1\_extract.seospiderconfig.
2. Parse and preprocessing. Software: Stata. Program: endchan\_2\_pol.do.

### PinkPill.co (website currently down)

1. Software: Screaming Frog. Config File: pinkpill\_1\_forum-catalog.seospideconfig.
2. Software: Screaming Frog. Config File: pinkpill\_2\_page-catalog.seospiderconfig.
3. Software: Stata. Program: pinkpill\_3\_scrape.do.
4. Software: Stata. Program: pinkpill\_4\_month.do.

**APPENDIX B.**

**MONTHLY PROCESSING CHECKLIST**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Site** | **Folder Catalog** | **Page**  **Catalog** | **Scrape /**  **Extract** | **Parse** | **Preprocess** | **OSC** |
| **Weekly Flash** |  |  |  |  |  |  |
| 4chan /r9k/ | n/a |  |  |  |  |  |
| Incel-is |  |  |  |  |  |  |
| Incel subreddits | n/a | n/a |  |  |  |  |
| 8kun QResearch | n/a |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **Weekly Public Health** |  |  |  |  |  |  |
| Public health subreddits (benchmarks) | n/a | n/a |  |  |  | n/a |
| Telegraph public health channels | n/a | n/a |  |  |  |  |
|  |  |  |  |  |  |  |
| **Monthly Background** |  |  |  |  |  |  |
| 8kun QResearch | n/a |  |  |  |  | n/a |
| Reddit Alt-Right | n/a | n/a |  |  |  |  |
| Reddit Conspiracy | n/a | n/a |  |  |  |  |
| Reddit Extreme Left | n/a | n/a |  |  |  |  |
| Reddit Extreme Right | n/a | n/a |  |  |  |  |
| Telegram – Python | n/a | n/a |  |  |  |  |
| Telegram – Export | n/a | n/a |  |  |  |  |
| Telegram -- Adjustment | n/a | n/a |  |  |  |  |
|  |  |  |  |  |  |  |
| **High Time-Intensity** |  |  |  |  |  |  |
| Patriots.win | n/a |  |  |  |  |  |
| Stormfront.org |  |  |  |  |  |  |
| RonPaulForums.com |  |  |  |  |  |  |
| GreatAwakening.win |  |  |  |  |  |  |
| Raddle.me |  |  |  |  |  |  |
| AboveTopSecret |  |  |  |  |  |  |
| 8kun Other Sites | n/a |  |  |  |  |  |
| White Biocentrism | n/a |  |  |  |  |  |
| Cbristogenea | n/a |  |  |  |  |  |
| EndChan: News | n/a |  |  |  |  |  |
| End Chan: Pol | n/a |  |  |  |  |  |