Digital Collecting Toolkit

University of Virginia Library

Contents

Welcome to the Digital Collecting Toolkit!

What is this Toolkit?

This toolkit is designed to provide resources and instruction on implementing digital collecting strategies during and after rapidly evolving social events and/or community crises (like campus controversies, natural disasters and public emergencies). Photos, videos, and social media content are major components of these community experiences, and the tools offered here can help organizations, institutions and communities quickly implement an effective digital collecting initiative.

This toolkit was created through a collaborative process by a team at the University of Virginia Library, with funding from the LYRASIS Catalyst Fund. In the immediate aftermath of the events of August 11th and 12th in Charlottesville, VA, UVa Library staff took on the task of quickly launching an online collecting tool and capturing

related social media content. While the UVA Library had some experience documenting and collecting digital content after a major news event, this was the first time we attempted to create a collecting site so quickly after the events occurred. The lessons learned from the site launch of the University of Virginia Library's Digital Collecting site, *"Unite the Right" Rally and Community Response* has led to further workflow and tool development to help future collecting efforts for ourselves and others.

This Toolkit is a result of these efforts, and is intended to help others quickly launch similar sites in times of crisis.

Who is this for?

Our team includes University preservation librarians, digital preservation specialists, archivists, digital content developers, and IT specialists. We developed this toolkit for use by a wide range of cultural institutions and communities with an interest in quickly setting up a digital collection site. Members of University libraries, other educational institutions, and community organizations using this toolkit will need to have some level of access to the tools provided here to implement a digital collection strategy. This could include access to web hosting, server space, and an Omeka installation. However, we have included a range of potential options for these requirements in this toolkit.

How do I get started?

This toolkit is designed so you can get started quickly!

- Read the basics on getting your collection site up and running.
- See the steps for setting up an Omeka Collection Site.
- See the section on Social Media Tools to learn how to set up and use the DocNow tool, Twarc for collecting Twitter data.

Contact us

If you have any questions or comments, you can reach us at digital_collecting@virginia.edu.

Getting Started

The basics for building a rapid response Digital Collection Site

Gathering a Team

In order for your digital collecting strategies to be successful, you may need to gather experts from across your organization. Here are some areas of expertise and types of responsibilities those team members might have:

- **Project lead**
  - Coordinate team, set up meetings and deadlines, document outcomes
- **Metadata**
  - Determine how content should be structured for ease of submission and description, future use/reuse
  - Describe/remediate description of content
- **Preservation**
  - Ensure short and long-term access to content
  - Select appropriate tools for ingest and preservation
- **Legal counsel**
  - Protect organization
  - Ensure institution can preserve and provide access over time
  - Consider implications of collecting on donors, people being documented
- **IT specialists and content developers**
  - Consider security, server capacity, and storage

  - Install and configure Omeka, select templates/plug-ins and tools for capture
- **Communications and user experience**
  - Review language and layout for ease of use by non-specialists
  - Documentation for internal and external users
- **Outreach**
  - Liaison to community from whom you are trying to collect
- **Collections specialist**
  - Define scope of the digital collection (types of materials, collection duration)
  - This might come from your special collections or other subject area

Ethical Considerations

Implementing and maintaining a digital collection site requires ethical, community-centric practices. We've compiled some sources that explore these considerations, and strongly encourage anyone developing an archive from community contributions and/or social media content to read further:

- Michelle Caswell, "Toward a Survivor-Centered Approach to Human Rights Archives: Lessons from Community-Based Archives." Archival Science 14: 3-4 (2014): 307-322.
- GWU Libraries' Social Feed Manager (SFM), Building Social Media Archives: Collection Development Guidelines
- An Invitation Towards Social Justice in the Digital Humanities
- Anti-oppression principles compiled by the Center for Story-Based Strategy

Funding Opportunities

This project was made possible in part by a 2018 award from the Catalyst Fund at Lyrasis. It may be worth exploring grant programs that can help support your digital collections.

Documenting the Now has an annual call for applicants for Community-Based Digital Archives Workshops for Activists. Their workshops focus on helping activists to develop the skills and to use available tools to collect, preserve, and share their web, social media and other types of digital content in their own digital archive.

Technical Considerations

Omeka for Digital Collections

Omeka offers two platforms - Omeka Classic and Omeka S - for publishing and managing your digital collections. Our digital collecting site runs on Omeka Classic, which is at the core of what is currently offered in this toolkit. Follow the materials in our toolkit section on Setting up Omeka Classic to build a site similar to *"Unite the Right" Rally and Community Response*. We've identified plugins and created a custom theme specifically for use with Omeka Classic.

Your project needs may align instead with Omeka S, which offers many of the same options as Omeka Classic. Instead of plugins, Omeka S provides modules to extend the functionality of your site. The Collecting module offers the equivalent to the Contribution plugin for collecting public contributions through your site. The Custom Vocab module is comparable to the Simple Vocab plugin for providing predetermined vocabulary for any metadata elements. See the complete User Manual for Omeka S, and the complete list of modules for that platform.

File Size Limitations

You'll want to consider setting a limit on the file size for individual contributions uploaded through your site's collection form. These settings are server specific, and can be set up by your server administrator when you install Omeka. If you're using Reclaim Hosting, or another shared hosting service, you may not be able to change the file size limitations.

For our digital collecting site, we set the maximum file size for uploads at 250Mb. This is large enough for short video files taken with smartphones, which was an item we were interested in collecting. Creating a limit on file size

may seem restrictive, but it is important for the sustainability of your server, where your files are stored. Your available server space is determined by your institution or your shared hosting service (like Reclaim), so consider how much total space you have for file storage when determining size restrictions.

Large Files and File Sets

Currently, the Contribution plugin for Omeka Classic allows for only individual file uploads, with a maximum file size set by your server administrator. For our site, we encourage contributors to use the contribution form to submit their materials, as it provides the simplest way for us to process the items we receive. However, for files larger than 250Mb or for large file sets we provide an option for contributing a URL link for an album or file location (i.e. Dropbox folder, Flickr album, etc.). When images and video files are uploaded directly through the contribution form, they're processed as Omeka items, and easily managed from the Omeka dashboard. Materials submitted as links have required significantly more time on our end to include them in the collection, and has been a major impediment to making more of our archive public.

Omeka offers the Dropbox plugin, which lets Omeka users batch upload a large quantity of files at once. It allows you to upload multiple files directly into a folder on your server that you can then add in the items admin interface. We have not used this plugin at this time for our site, but it offers one option for handling large file set submissions.

- Dropbox plugin download
- Dropbox plugin User Guide

This page was published: April 25, 2019

Setting Up Omeka Classic

The University of Virginia Library's Digital Collecting site, *"Unite the Right" Rally and Community Response* runs on *Omeka Classic*.

Omeka is a leading open source collections-based web publishing platform developed by the Roy Rosenzweig Center for History and New Media and supports a robust open source developer community. Omeka is standards-based, grounded in a flexible Dublin Core Metadata Schema. Omeka takes a user-centered, access-focused approach to collections, emphasizing approachable, accessible web design.

Technical Specifications

If you are a member of a university or institution with an accessible IT Department or Digital Content Developers, we recommend reaching out to individuals in your organization with experience setting up Omeka.

Omeka Classic has the following system requirements: - Linux operating system - Apache HTTP server (with `mod_rewrite` enabled) - MySQL version 5.0 or greater - PHP scripting language version 5.3.2 or greater (with mysqli and exif extensions installed) - ImageMagick image manipulation software (for resizing images)

If you lack access to a server that meets Omeka's basic requirements, or are looking for a **simpler and less technical set-up process**, we recommend Reclaim Hosting, a low-cost shared web host offering one-click Omeka installation, with a free domain registration included. Their support staff is easy to reach and their Community Forum offers additional resources.

- Reclaim Hosting
- Installing Omeka Classic on Reclaim Hosting
- Uploading Plugins and Themes to Omeka on Reclaim Hosting
- Working with Omeka Classic on Reclaim Hosting
- Omeka on Reclaim: Community Forum

Omeka also provides a list of suggestions for low-cost shared web hosts that offer the server environment required for Omeka, see here for details. - Complete Omeka Classic Installation Guide

Omeka Classic Plugins

One of the primary features of a rapid response digital collecting site is a user-friendly submission form set up to collect photos, videos, stories, website links, audio and other files from the community. Omeka Classic offers the Contribution plugin, which adds a content submission form to your public collecting site.

When using the Contribution plugin, contributors may share and upload content anonymously, and their information will only be available to site administrators. All contributions are private by default and require a site administrator to review and make them public on the Omeka site.

The plugin can also automatically add a reCAPTCHA box at the bottom of each form to prevent bots from spamming your website. Contribution also offers options for users to create guest accounts that make it easier for one user to submit multiple items.

In addition to the Contribution plugin, there are additional Omeka Classic plugins that may be useful to your digital collecting site. See the following page for details on Omeka Classic Plugins.

Custom Omeka Theme

This toolkit also offers a custom Omeka theme that you can download for the styling and content of your public Omeka site. This is not required, but you may find it useful for getting your collecting site up and running quickly with minimal technical knowledge. Our theme offers customizable settings for the appearance and content of your site, easily accessible from the Omeka admin dashboard, no coding knowledge needed! For details on installing and customizing our Digital Collecting Theme for your site, see here.

Administering your Omeka Classic Site

For helpful information on administering your Omeka Classic site, including managing users, security settings, and site navigation, see the related page in this toolkit: Omeka Site Administration.

This page was published: April 25, 2019

Omeka Plugins

Download and install plugins to extend the basic functionality of your Omeka site: build forms for community contributions, customize the collected metadata in your archive, and other useful options for your digital collecting site.

Our Digital Collecting site includes the following plugins. Links to download for your site are included below, as well as details on how we used the plugins. You may choose which plugins, among others offered by Omeka, are most useful for your site.

Please follow the links for related installation and user guides for details on how to set-up and use the individual plugins. For instructions on installing and managing Omeka plugins, see here.

Published: April 25, 2019

Guest User

You must upload and install the Guest User plugin before installing and activating the Contribution Plugin.

- Download Guest User plugin
- Installation and Configuration Guide

Contribution

- Download Contribution plugin
- Installation & Configuration Guide

The Contribution plugin provides a form to collect stories, images, and other files from the public and manage those contributions in your Omeka site as items. Contributors may share and upload content anonymously, and their information will only be available to site administrators. Other plugins can be integrated into the Contribution form, such as Simple Vocab for creating dropdown menu choices, and Geolocation for inviting users to map their submissions or locations.

Installing and configuring the Contribution plugin requires a few steps. Please read through the Omeka documentation carefully.

See Collecting Materials for additional information on our site's Contribution Terms of Service and how we configured this plugin for collecting and managing stories, images, and other files from the public.

Dublin Core Extended

- Download Dublin Core Extended plugin
- Documentation

Dublin Core Extended provides additional metadata elements that you may find useful to collect in your Contribution collection form. We used this plugin specifically for the 'Spatial Coverage' element, which along with the Simple Vocab plugin, allows contributors to submit the location of their contribution.

Simple Vocab

- Download Simple Vocab plugin
- Configuration & Use Guide

Simple Vocab is a useful plugin for providing predetermined vocabulary for any metadata elements. This plugin allows you to create drop-down menus that replace the usual text box for an element.

We used Simple Vocab to replace the element for "Spatial Coverage" with a list of locations around Charlottesville relevant to the events. When a contributor chooses to fill out the 'Location' input in the collection form, they choose from a drop-down list of predetermined terms, set in the Simple Vocab settings. Once the plugin is installed, you can manage your custom vocabulary by clicking on the 'Simple Vocab' tab on the left-hand menu in your Omeka dashboard:

Element Types

- Download Element Types plugin

Used for standardizing date input for contribution submissions (provides a pop out calendar for date selection in the Contribution form).

Derivative Images

- Download Derivative Images plugin
- Documentation

Simple Pages

- Configuration Guide

Used to create our About the Archive page. This plugin is useful for creating additional text rich pages for your site, and comes bundled with your Omeka download.

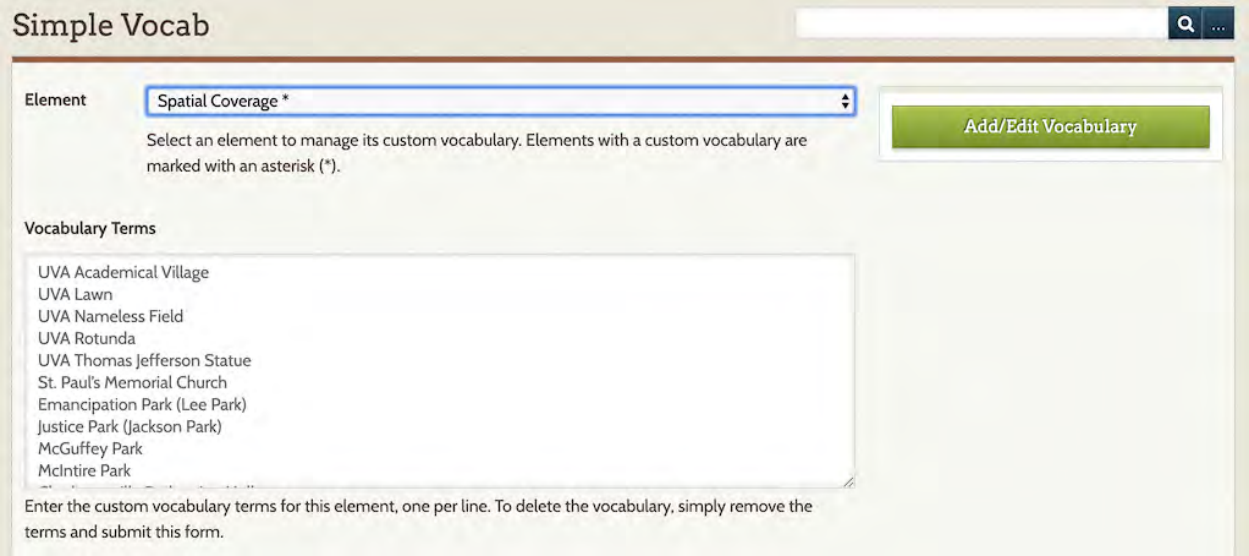


Figure 1: Screenshot of Simple Vocab page in Omeka

Search by Metadata

- Download Search by Metadata plugin
- Installation and Configuration Guide

The Search by Metadata plugin makes it possible for visitors to your public Omeka site to see a browse page with items that share specific metadata (i.e., all items with 'UVA Rotunda' in the Spatial Coverage field). This plugin works very well with the Simple Vocab plugin for controlling metadata vocabulary. See the related Omeka documentation for details on configuring this plugin.

Additional Plugins

We do not currently have the following plugins active on our collecting site, but they may be useful for others.

Geolocation

- Download Geolocation plugin
- Configuration Guide

The Geolocation plugin allows you to assign a location to items in your Omeka site. The locations are displayed on maps on individual items page and on a browsable map of all geolocated items, that you can add to your site's primary navigation. This plugin also allows you to add a map to your Contribution form, allowing contributors to select locations on a map to submit with their materials.

Exhibit Builder

- Configuration Guide

The Exhibit Builder plugin allows you to develop online exhibits, or special web pages, that combine items from your Omeka archive and may include narrative text. You may find it useful for creating exhibit pages out of collected materials. This plugin is included in your downloaded Omeka installation and does not require a separate download.

Dropbox

- Dropbox plugin download

- Dropbox plugin User Guide

The Dropbox plugin allows you to upload multiple files directly into a folder on your server that you can then add in the items admin interface.

Collecting Materials

We've included detailed documentation on how our Digital Collecting site uses the Contribution plugin for Omeka to collect stories, images, videos, and links from the public.

Please read through the plugin's Installation & Configuration Guide for full set-up instructions before getting started.

Published: April 25, 2019

Creating Contribution Types

After installing the plugin, find the settings for Contribution by clicking on the "Contributed Items" tab on the left-hand navigation panel of the Omeka admin dashboard. Here you will see four tabs: Getting Started, Contribution Types, Submission Settings, and Contributions.

Select "Contribution Types" to manage the types of items you'd like users to share through the form:

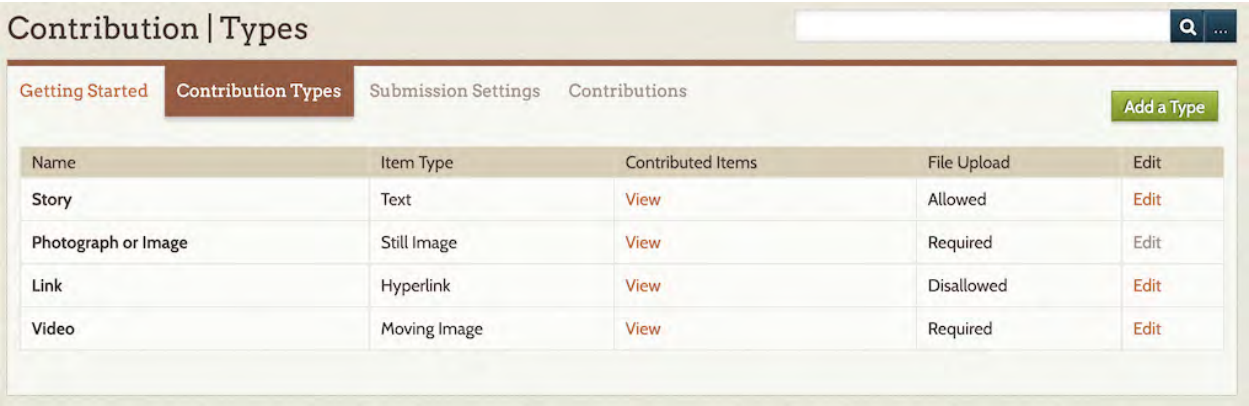


Figure 2: Screenshot of Contribution Types Admin

See in the above image that we have included four contribution types for our site: *Story, Photograph, Link,* and *Video*. The type's *Name* is the label that will appear in your site's Contribution Form:

Figure 3: Screenshot of Contribution Form, Select Type

The *Item Type* is the type of object associated with your Contribution Type, and is selected when you add a new contribution type to your form (see the green button 'Add a Type' on the right). These are not visible to the public, but can help in defining search fields.

Omeka comes with pre-defined item types, found by clicking on the "Item Types" tab on the left-hand navigation panel of the Omeka admin dashboard. You can edit and add to this list if needed (see the related Omeka Documentation). Our site uses the pre-defined types: *Text, Still Image, Hyperlink,* and *Moving Image*.

Contributor Anonymity Settings

Under the tab for "Submission Settings" you can set options for contributor anonymity. Our site allows both Non-Registered Contributions and Anonymous Contributions:

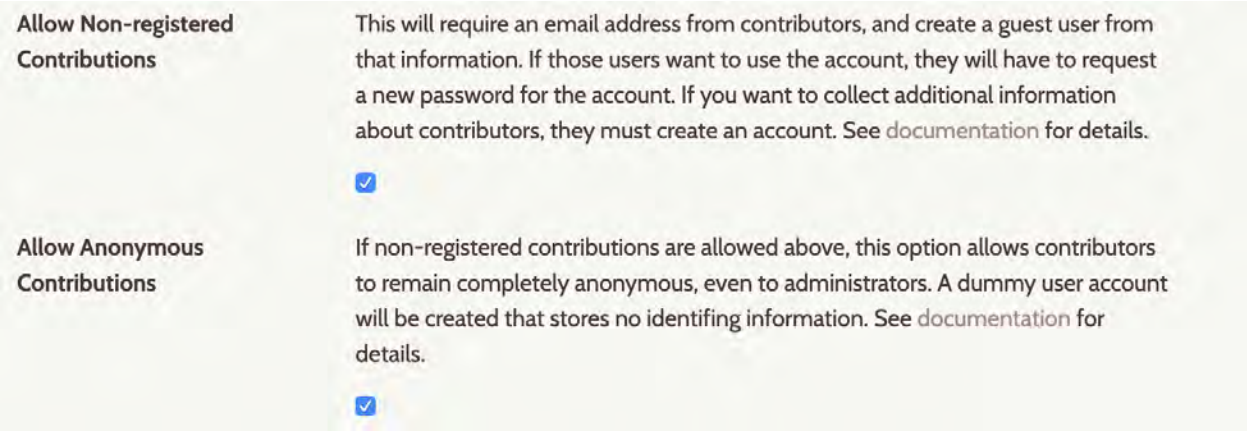


Figure 4: Screenshot of Contribution Settings panel

It is important to make clear in your Terms of Service that user anonymity is conditional, and it is likely that under a lawful subpoena or court order, all submissions and associated data may be required to be provided to federal, state, or local law enforcement or other government agencies. See our Terms of Service section for more details.

For further documentation on Contribution Submission Settings and Contributor anonymity settings, see the related Omeka Documentation.

Contribution Terms of Service

We offer a general Contribution Terms of Service template for others as a model. Our terms were written in collaboration with University of Virginia General Council. We recommend reviewing your terms with relevant parties for your own collecting site.

> **Contribution Terms of Service:**
>
> You are being asked to contribute your recollections, photographic images, video, social media postings or other digital content to [insert institution here], which is creating a digital record of the [describe event, location, and date(s) here]. You may only submit material created entirely by you and not copied from or based, in whole or in part, upon any other photographic, literary, or other material, except to the extent that such material is in the public domain, or you have permission of the copyright owner, or its use is allowed by "Fair Use" as prescribed by the terms of United States copyright law. If you would like to refer or nominate material which you do not own, please contact us at [insert email address or link to google nominating form]. You must be 18 years of age or older to submit material. By submitting content through this form, you are granting [insert institution here] permission to disseminate, preserve, and use that content in connection with its educational and research mission, including promotional purposes, in all media in perpetuity. You retain ownership of and copyright in the material you share. If you indicate on the form that your submission is "public," your material may be published on the web (with or without your name, depending on what you have indicated) as part of [insert institution here] digital collections or exhibits. Otherwise, your material will only be available to [insert institution here]-approved researchers.

> Submitted material must not violate any confidentiality, privacy, security or other laws. Please be aware that all submissions and any information associated with the submissions (email address, descriptive information, etc.) may be provided to federal, state, or local law enforcement or other government agencies pursuant to a lawful subpoena or otherwise as required by law. We reserve the right to discard or mark private any submission that [institution] staff identify as irrelevant or for any other reason within their professional judgment. [Insert institution here] is not obligated to include your content in this project or preserve it in perpetuity.

In addition, our Digital Collecting site's Terms of Service includes a Summary of Terms to help contributors quickly read and understand the main points of the full terms, in simpler language. It is important to note to users that a summary does not replace the full terms, and submissions are governed by the full terms of use.

> **Summary of Terms***
>
> - You must be at least 18 years old.
> - Submitted material must be owned and/or created by you.
> - You have the option of making your contribution public or private. If public, your content may be published as part of the Library's digital collections (with or without your name displayed, depending on what you have indicated).
> - All submissions will be available to Library-approved researchers and can be used by the Library from now on in support of its teaching and research mission.
> - Your submission must not violate any laws. If we receive a lawful subpoena or court order, we may be required to turn over any submissions and related information (email address, descriptive information, etc.). *This summary is to help you read and understand the terms, but does not replace them. Your submission is governed by the full terms of use.

Custom Omeka Theme

Our Digital Collecting site uses a custom Omeka Theme: Charlottesville Rally Theme. We developed this theme specifically for use in digital collecting sites, making it simple to customize the content and appearance of your public site directly from your Omeka Dashboard - no coding knowledge required.

Published: April 25, 2019

Steps for Installing our Custom Theme

1. Download and unzip our custom theme from Github. Download the .zip file to your desktop for easy retrieval.
2. Locate your Omeka installation and login to access your site.
3. Navigate to your Omeka folder (this should have the same name as your Omeka install)
4. Open your Omeka folder and locate the '/themes' folder within.
5. Copy or move the unzipped 'cville_rally_theme' folder from your desktop (or from where you saved this folder) and place it within the 'omeka/themes' folder located in step #4.
6. Log in to your Omeka admin panel (found at the url: 'your-site-url/admin'), and click on the 'Appearance' option in the top navigation bar.
7. The custom theme should now be installed and visible (see screenshot below). If not, double-check that the folder is in the right location ('/themes') and that the folder name for the theme does not start with 'theme-'.
8. Click on "Configure Theme" to customize your site's appearance. See the next section for details on configuration settings.

For more detailed instructions on installing an Omeka theme, see the related Omeka Documentation.

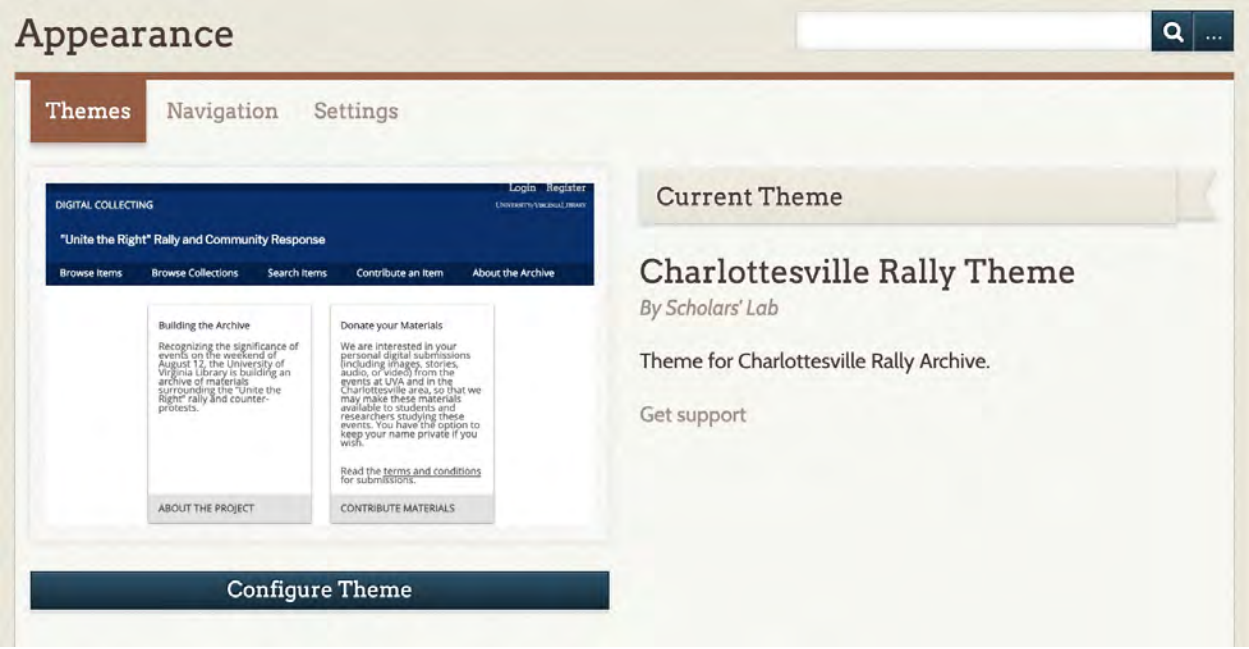


Figure 5: Screenshot of Omeka Theme Admin

Configuration Settings

To add your site-specific content and customize your site's color scheme, begin at step #6 above: Log into your Omeka Admin panel, and click on the 'Appearance' option in the top navigation bar. Next select "Configure Theme" to access the theme settings (see screenshot in step #7 above).

You should now be on the Theme Configuration page:

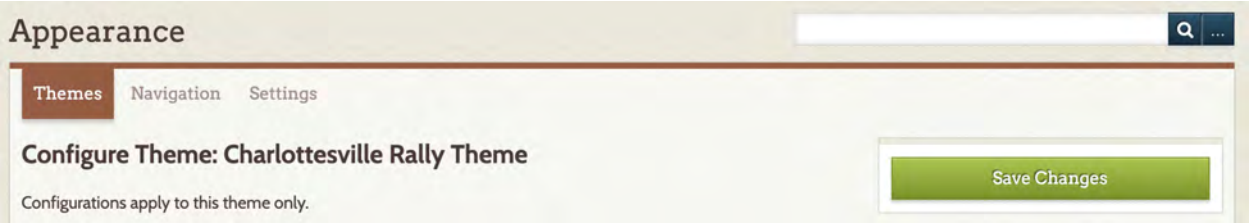


Figure 6: Screen shot of Theme Configuration page

Customize your Site's Color Scheme

Use this section to customize your site's color scheme. These boxes will automatically fill in with our UVA theme colors, you can change them for your site. You must use a six-character hexadecimal color value, including the #.

- **Header Color**: Determines the color of the widest section of the page header, as well as the overlay color of the homepage's banner image
- **Header Text Color**: Determines the color of the text in the widest section of the page header, specifically, the text you provide in the **Header Site Title** and **Header Tagline Text** in the Header and Footer content section, and the Homepage Banner Introduction Text.
- **Navigation Background Color**: Determines the color of the primary navigation bar, the Omeka login bar along the top of the page, and the page footer.
- **Navigation Text Color**: Determines the color of the text in the sections listed above. The color of the navigation links in the footer, and the Omeka login links in top admin bar, are set to be a shade darker than

Figure 7: Screen shot of Theme Configuration page

the color given here, and show full brightness when you hover over these links. For instructions on how to customize your site's navigation, see here

- **Headings Text Color**: Determines the color of the headings in your site (for html tags h1, h2, h3, h4).

Header and Footer Content

Use this section to provide content for your site's header and footer. All of these sections are optional, feel free to pick and choose what works best for your site. We recommend including at least the **Header Tagline Text**, as this is the primary heading for your site.

- **Header Site Title**: This is for an optional secondary title section for your header, this site title is separate from the title of your Omeka site.
- **Header Tagline Text**: This is the primary header text, we recommend filling this in as it provides the primary level heading used by screen readers.
- **Logo File**: Upload an image file for your site's logo, to be placed in the top right of the header. Your image file should be no larger than 150px in height. This is optional.
- **Logo Link URL**: Provide a URL that your **Logo File** will act as a link for. This is optional, if you leave this blank your logo will not link to anywhere.
- **Logo Image Accessibility Alt Text**: Provide descriptive alt text for your logo image for use with screen readers.
- **Footer Text**: Provide some content for your page footer. This custom theme already provides site navigation in the footer, this box is for additional content.
- **Shortcut Icon File**: Provide an image file to appear in your browser tab/address bar. This must be a square image, 16x16 or 32x32 pixels, and must be a PNG file in order to upload through your Omeka theme configuration.

Customize your Site's Homepage Content

Use this section to provide content for your site's homepage. All of these sections are optional, pick and choose what works best for your site.

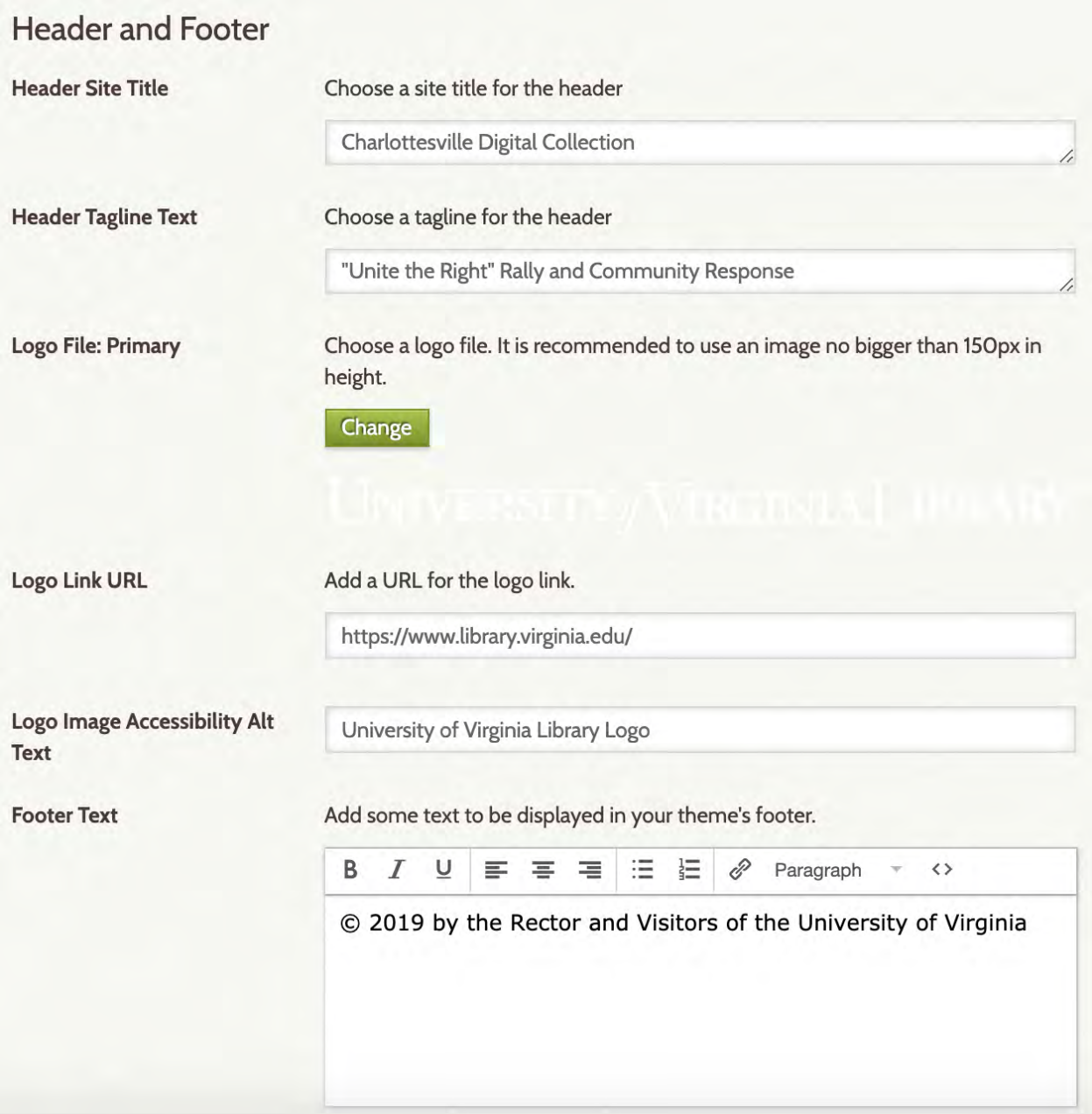


Figure 8: Screen shot of Theme Configuration page

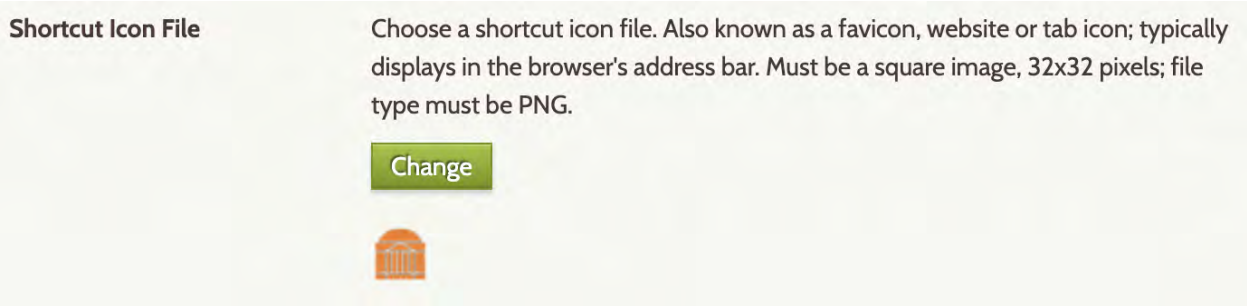


Figure 9: Screen shot of Theme Configuration page

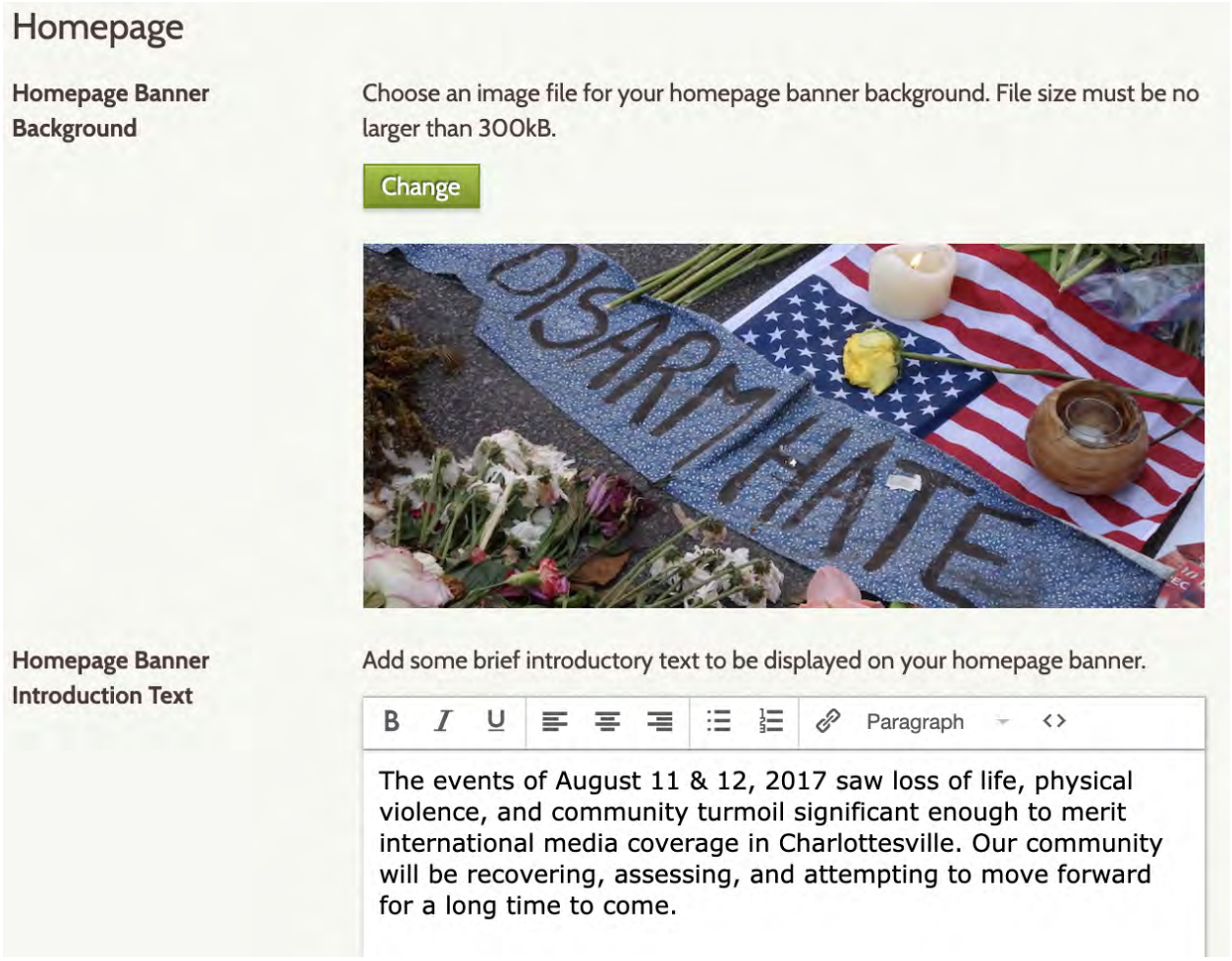
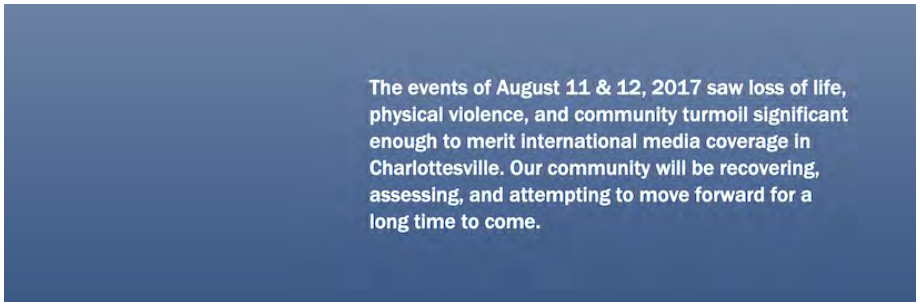


Figure 10: Screen shot of Theme Configuration page

Homepage Banner

- **Homepage Banner Background**: Provide an image file to be used on the homepage introduction banner. File must be no larger than 300kb, and will automatically be centered and scaled to fit. Your image will be overlayed with a color gradient, as selected in **Header Color**.
- **Homepage Banner Introduction Text**: Provide some brief text to display on your banner image. This text will appear in the same color as selected in **Header Text Color**.

  Both the banner image and introduction text are optional, and can be used together or individually. If no image and no text is provided, your site will not have a banner. If you provide introduction text but no image, your text will display over a color gradient banner, the color selected in **Header Color**. If you provide a banner image with no introduction text, your image will not have the color overlay. In this case, if you'd like the color overlay on your image, type in a few blank spaces into the introduction text box (no text, just use your space bar to fill in some blank content).

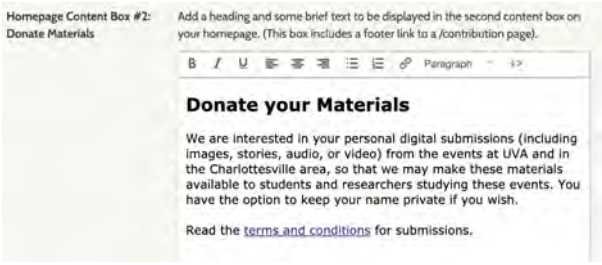
Omeka Admin	Public Site
Homepage Banner Background & Introduction Text	
Introduction Text only	

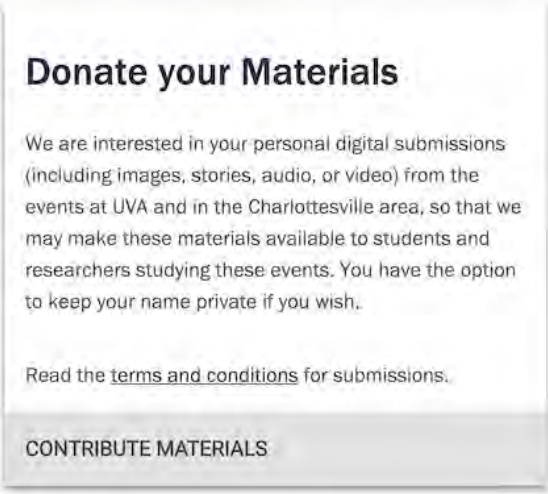
Homepage Content Use the following sections for the primary content of your homepage. All of the following sections are optional, please pick and choose what will work best for your site.

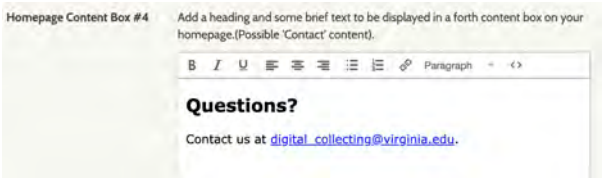
This theme's homepage content can be created with up to four content boxes. The first two are designed for use with a Contribution page (/contribution) and an 'About' page (/about), created with the Simple Pages plugin. If you do not want to use these content boxes with pre-defined footer links, leave sections *Homepage Content Box #1: About the Archive* and *Homepage Content Box #2: Donate Materials* empty, and these will not appear on your site.

The other two content boxes, *Homepage Content Box #3* and *Homepage Content Box #4*, do not have any pre-determined footer links and can be used for any type of content. Pick and choose which content boxes you'd like to use for your site, and leave empty any you do not wish to appear. The boxes will automatically resize to fit the screen, and depending on how many you select to use, may stretch across the whole page.

Omeka Admin	Public Site





Omeka Admin	Public Site
	

Use the toolbox buttons to format your content, add headings, links, and lists. Our site uses 'Heading 2' (or the h2 tag) for these content box headings. By clicking the 'source code' button, you can edit the html source code directly, if you'd like. Below is an example of our HTML source code:

HTML	Public Site
`<h2>About the Archive</h2><p>Recognizing the significance of events on the weekend of August 12, the University of Virginia Library is building an archive of materials surrounding the "Unite the Right" rally and counter-protests.</p>`	

Optional Featured Content Check the boxes for additional content to be displayed on your homepage. All of these are optional. You must have featured items, collections and/or exhibits in your Omeka collection for some options to be relevant.

- **Display Browse Collection by Type**: Check this box if you wish to show 'Browse the Digital Collection' on the homepage. Includes icons for Item Types: Still Image (Label: Photos), Moving Image (Label: Videos), Text (Label: Stories), Hyperlink (Label: Links). To customize what is included in this box, you must edit the file `browse-types.php`, found in the `/common` folder within `themes/cville_rally_theme`.
- **Display Featured Items**: Check this box if you wish to show three featured items on the homepage. If you have more than three featured items in your Omeka collection, these will show random featured items each time the site is refreshed.

Display Browse Collection by Type	Check this box if you wish to show 'Browse the Digital Collection' on the homepage. Includes icons for Item Types: Still Image (Label: Photos), Moving Image (Label: Videos), Text (Label: Stories), Hyperlink (Label: Links). ☑
Display Featured Items	Check this box if you wish to show random featured items on the homepage. ☑
Display Featured Collection	Check this box if you wish to show the featured collection on the homepage. ☐
Display Featured Exhibit	Check this box if you wish to show the featured exhibit on the homepage. ☐
Display Recent Items	Check this box if you wish to show recent items on the homepage. ☐

Figure 11: Screen shot of Theme Configuration page

Figure 12: Screen shot of 'Browse Collection by Type' Content

- **Display Featured Collection**: Check this box if you wish to show a featured collection on the homepage.
- **Display Featured Exhibit**: Check this box if you wish to show a featured exhibit on the homepage. For use with the Exhibit Builder plugin.
- **Display Recent Items**: Check this box if you wish to show recent items to be displayed on the homepage. These will appear in the order in which they were mostly recently added to the archive.

Contribution Page: Form Submission Instructions and Information

Use the following sections to create content for your public Contribution page. This is the page created by the Contribution plugin, and contains the form for public submissions. Provide instructions and additional information for those contributing materials to your collection. See our page as an example, here.

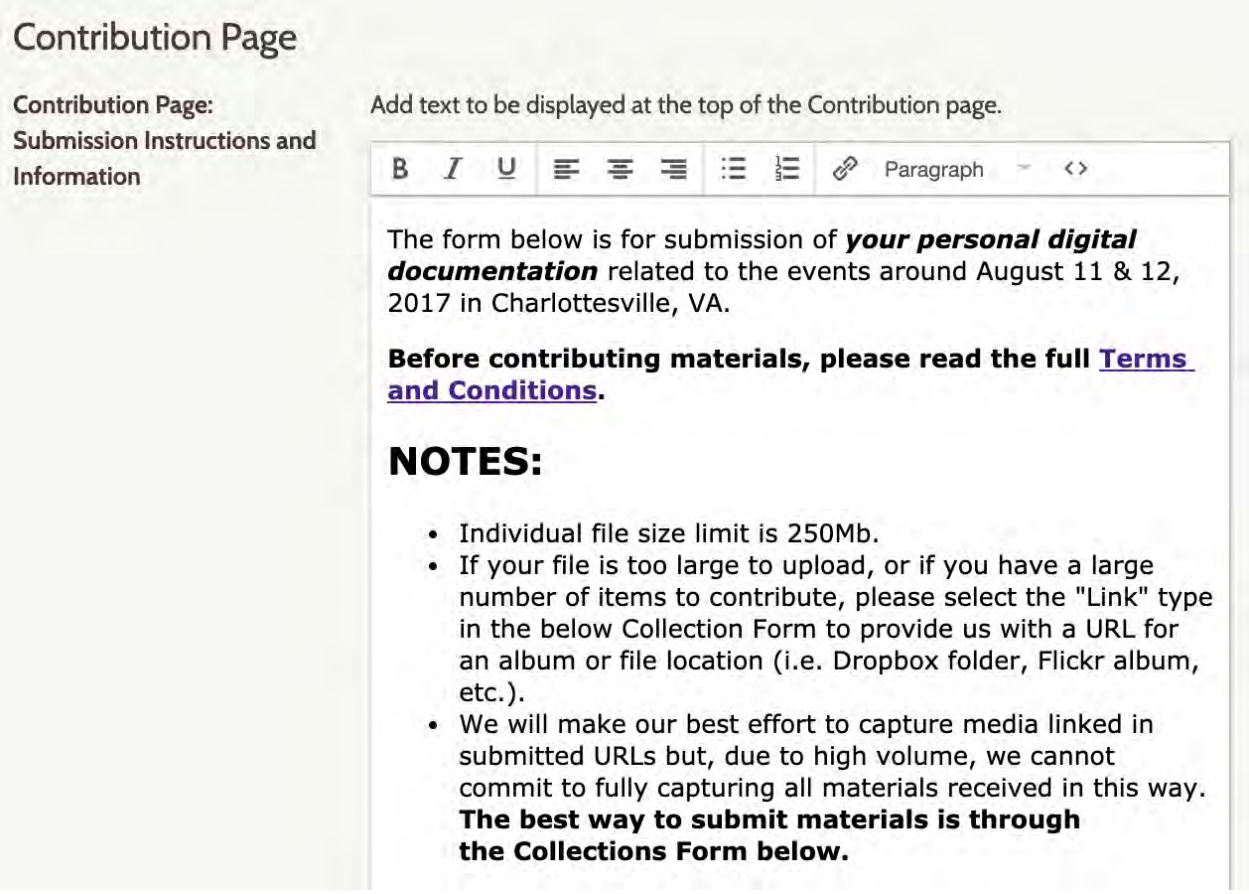
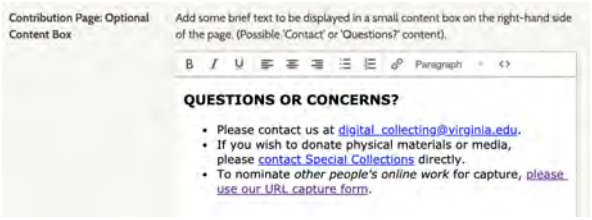


Figure 13: Screen shot of Theme Configuration page

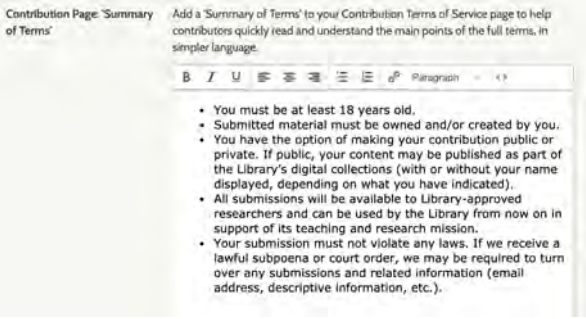
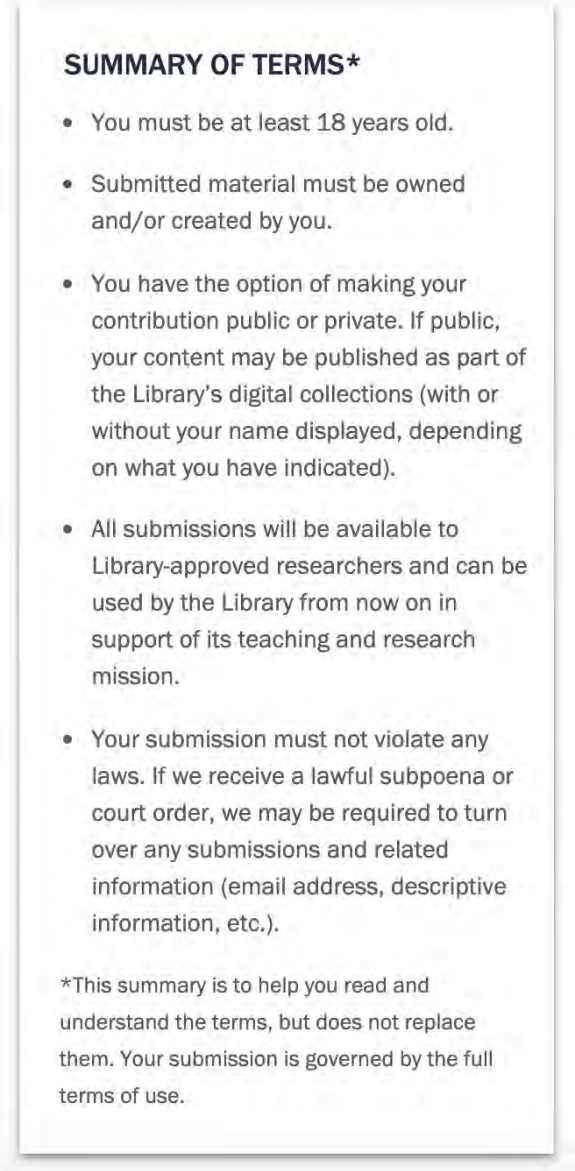
In addition to the primary content, you have the option of including a small information box, set on the right-hand side of the page. This is useful for contact information or anything else you'd like to include. Leave this empty if you do not want this content box to appear. See below:

Omeka Admin	Public Site
	

Contribution Terms of Service: Optional 'Summary of Terms' Box On your Contribution Terms of Service page, this theme has an option to include a 'Summary of Terms' content box on the right-hand side of the page, to help contributors quickly read and understand the main points of the full terms, in simpler language. This is optional, leave this field empty if you do not want this to appear.

The complete Terms of Service text is given within the Contribution plugin settings. Contribution is managed from the plugin's tab on the left-hand navigation of the admin dashboard. You can find the field for **Text of Terms of Service** under the 'Submission Settings' tab. See the complete Contribution plugin user guide, here.

To use the 'Summary of Terms' box, provide a bullet-pointed list in the field. This box automatically features a heading and (*) footnote. See below:

Omeka Admin	Public Site
	

Administering your Omeka Classic Site

Published: April 25, 2019

Managing Users

Manage users by clicking on the Users tab in the top navigation of your Omeka admin dashboard. The Users section gives the site Administrator control over who may access the admin section of the site and what they can do.

User Levels and Access

Omeka allows you to give different backend users different levels of access to your archive. Read through the following list of actions available to users to determine what works best for your project team members.

All logged in site Super, Admin, Contributor, and Researcher users can view non-public content (items, collections, Simple Pages, Exhibits, etc) on the site.

Super Users

Can do anything and everything in Omeka. Supers are the only users with access to the top navigation tabs for Plugins, Appearance, Users, and Settings.

Admin Users

Admin users do not have access to the tabs for managing plugins, appearance, users, or site settings. Admin users can:

- Add, edit, tag, and delete items, both their own and created by other users.
- Make items, collections, exhibits, and other content public or not public.
- Make items, collections, exhibits, and other content features or not featured.
- Add, edit, and delete Item Types.
- Add, edit, and delete files.
- Interact with plugins installed and activated by a SuperUser.
- Add, edit, and delete tags.

Contributor Users

Contributor users have control over their own content but can only view content created by others. They cannot make their own content public. **Contributor Users are different from 'Guest' and 'Contribution anonymous', the user types added when using the Contribution and Guest User plugins. See below for details on these user types**. Contributor users can:

- add, edit, tag, and delete items which they created.
- cannot make their own items public.
- create their own exhibits from items that are public.

Researcher Users

Researchers can log in to the admin side of an Omeka site and see the content, but cannot interact with it in any way. They cannot add, edit, delete, or tag any items.

For complete documentation on managing users, see the related page in the Omeka Classic documentation, here.

Guest & 'Contribution Anonymous' Users

When using the Contribution and Guest User plugins, two additional user types will become available - Guest & 'Contribution Anonymous'. Depending on how you configure these two plugins, when users make submissions through your contribution form, they will potentially have the option of including their email to set up a 'Guest' user account, or they can choose to make their submission anonymous.

For details on configuring these settings for user contributions, see the related section in Collecting Materials.

Site Security Settings

Access your site's Security Settings by clicking on the 'Settings' tab in the top navigation of your Omeka admin dashboard. Then select the 'Security' tab in the Settings toolbar:

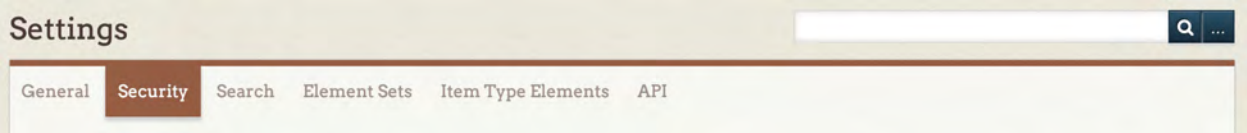


Figure 14: Screen shot of Security settings bar

Scroll down to the Captcha section. A captcha is a program that can help ensure that only actual people, not spammers or robots, are using public forms on your Omeka site, including the Contribution plugin content submission form.

In order to use reCAPTCHA, you will need to sign up for Captcha key to reduce spam on your site. Sign up by following the link to the google developer's reCaptcha site, provided in the Captcha section (see below screenshot). Once you sign up, you can enter your site and secret keys in their respective fields:

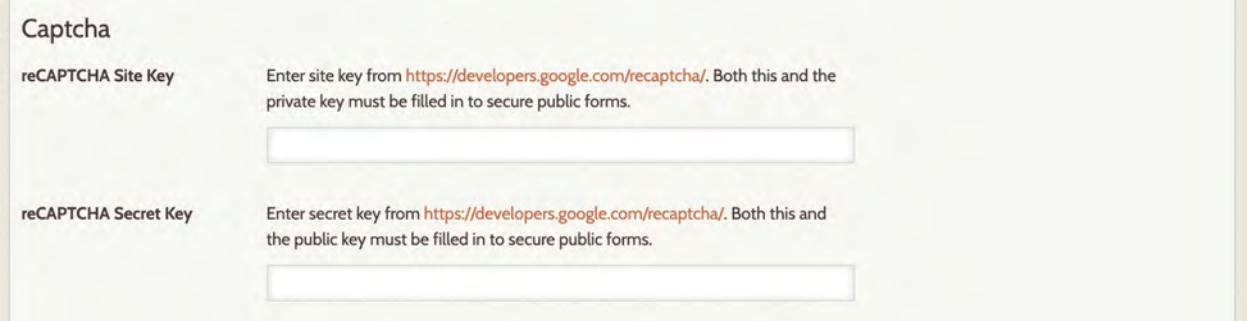


Figure 15: Screen shot of Security settings bar

For more details on reCaptcha settings, see the related Omeka Documentation.

For more information on additional site security settings, see here.

Customizing Site Navigation

From your Omeka admin dashboard, click on the 'Appearance' option in the top navigation bar. Select the 'Navigation' tab in the Appearance toolbar to see your site's navigation settings.

These settings determine your site's main navigation, see our site as an example:

Figure 16: Screen shot of Navigation bar

1. To change the default label for a navigation link, click on the arrow to its right and change the text under 'Label' (see the screen shot below, with open editor for 'About the Archive'):
2. Add a link to your navigation bar using the box 'Add a Link to the Navigation'. To create a link for a search page, fill in the 'Label' and 'URL' as seen below, and click the 'Add Link' button:

   Use this box to add any additional links to your navigation bar as needed. You can include URLs from outside your Omeka site as well.
3. Select the check box for links you want to appear in you navigation bar. Click and drag the links to change to your preferred order.

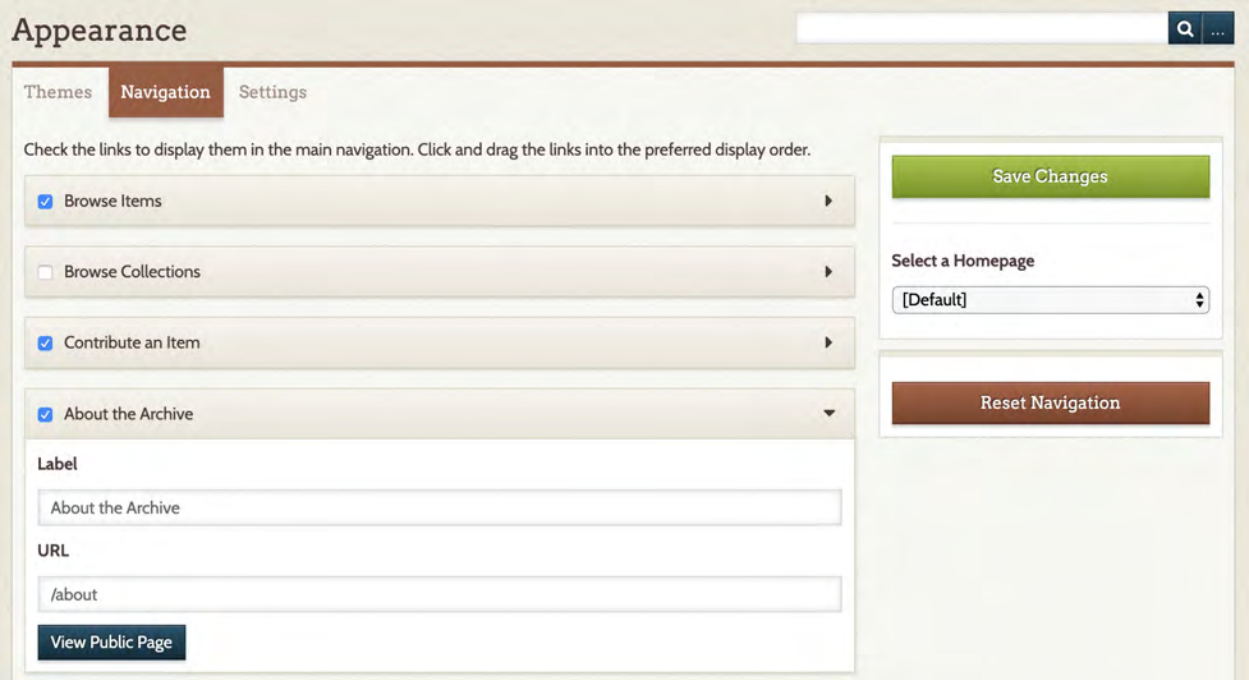


Figure 17: Screen shot of Navigation settings

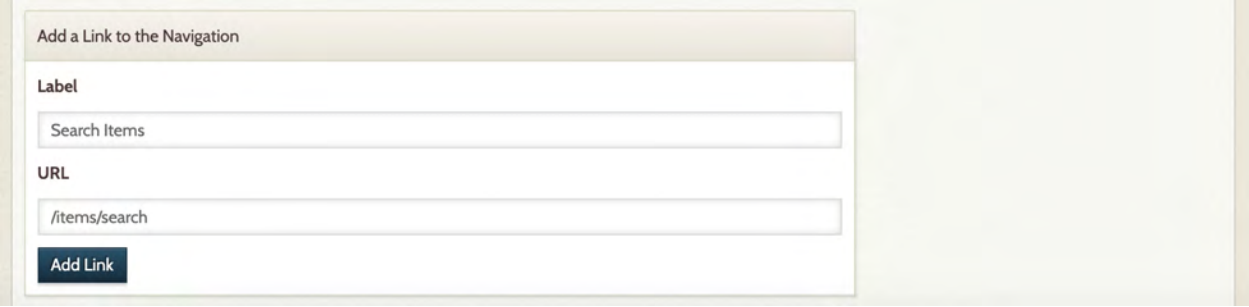


Figure 18: Screen shot of Navigation settings

Figure 19: Screen shot of Navigation settings

Tools for Archiving Social Media

This guide presents methods of collecting twitter data using tools built by DocNow, a collaborative effort between Shift Design, Inc., the University of Maryland, and the University of Virginia, with funding from the Andrew W. Mellon Foundation. We use the tools developed by DocNow for collecting twitter data because of their strong commitment to prioritizing ethical practices in collection, use, and preservation of social media content.

The following documentation is intended to assist in setting up Twarc, a DocNow tool for archiving twitter data. For more information on Twarc or other DocNow tools, please visit their site.

Twarc for Twitter Data Collection

Twarc is a command line tool that downloads tweets using Twitter's API. API's, or application programming interfaces, are simply ways that different organizations, whether it is Twitter or the Census bureau, provide more direct access to data. API's also oftentimes provide limits to how much data you can gather. Twarc will handle Twitter API's rate limits for you.

The following pages provide instructions on installing and using Twarc. Parts of this guide are subject to change with updates to Twitter's developers site, so please use this guide as a general guideline. For troubleshooting with Twarc, please contact the developers of DocNow, and join in conversation with the DocNow community of scholars, students, and archivists.

To get started you will need a twitter account to register a twitter application. Twarc also requires Python.

This page was published: April 25, 2019

Setting up your Twitter Account for Collecting

Published: April 25, 2019

Steps for Creating a Twitter Application

Please note that these instructions may be subject to change with updates to Twitter. If the following steps and screenshots do not match precisely, please use this as a general guide, or contact DocNow.

1. If you do not have a twitter account, create one at twitter.com. A twitter account is required for access to twitter data.
2. Log into you twitter account to set-up and and authorize a twitter application. A Twitter application will let you download twitter data using Python.

   Create an app at developer.twitter.com/en/apps

   Following the above link should bring you to this page:

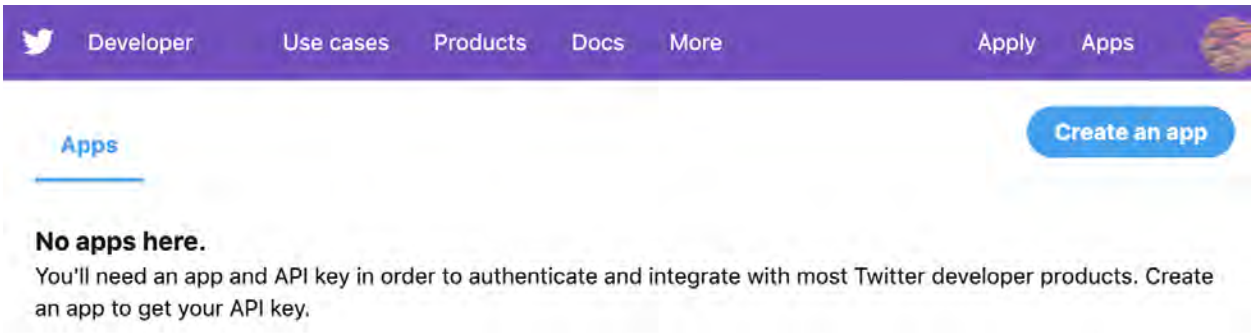


Figure 20: Screen shot of Twitter Developers Site

3. Click on 'Create an App'

   You may be prompted to create a Twitter developer account. Select 'Apply' and continue:

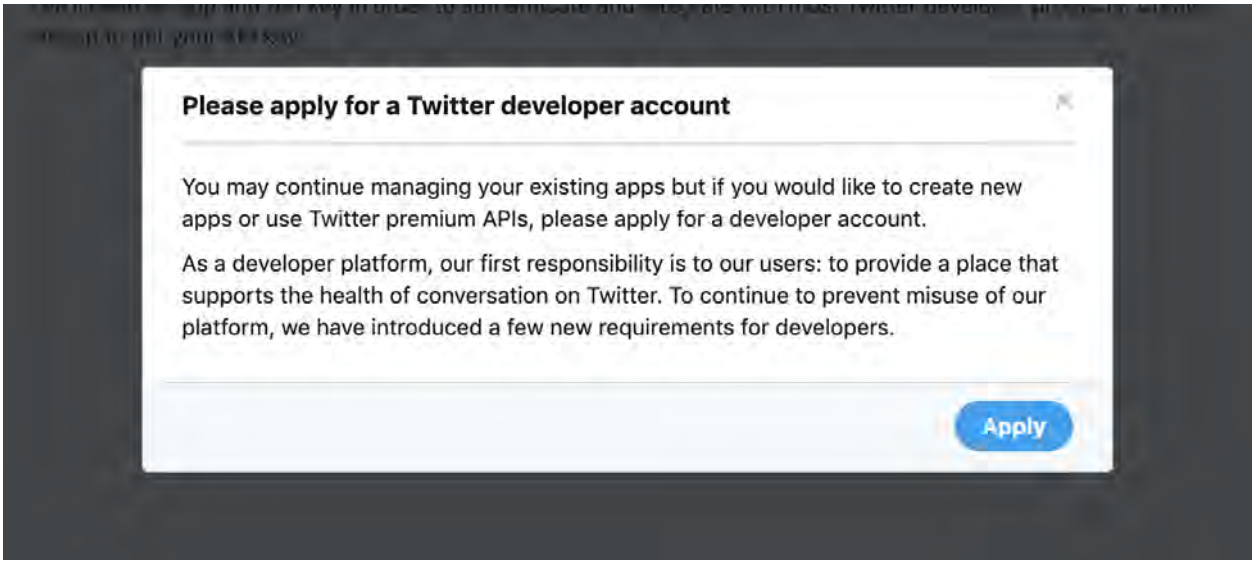


Figure 21: Screen shot of Twitter Developers Site popup

Setting up a Twitter Developer Account

If you already have a developer account set up, skip to Creating a Twitter App.

1. Select a user profile to associate with the developer account, and select 'Continue'.

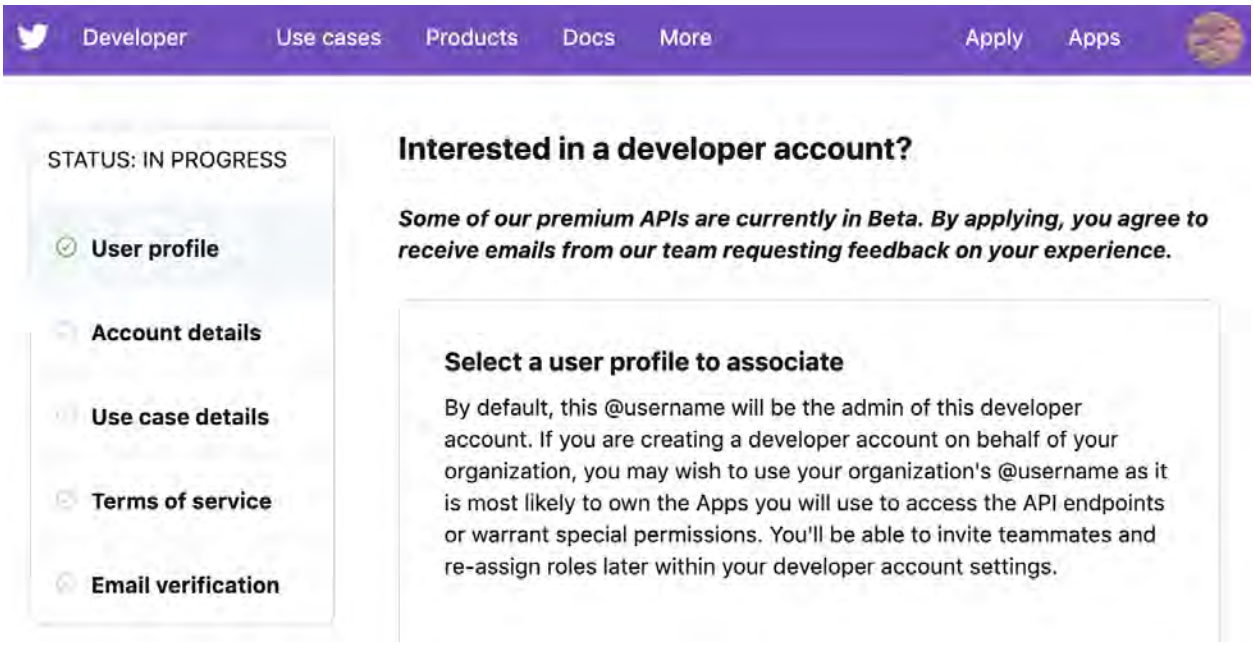


Figure 22: Screen shot of Twitter Developers Site

2. In the 'Account details' section, select your request for access for either your organization or for personal use.

   Note: Selecting "for my own personal use" offers the simplest and quickest method of setting up a developer account. If you are following this guide for research or teaching purposes, an individual, personal account is suggested. Twitter users can only have a single development account, and cannot change account types. An 'Organization' account differs in that it allows for additional twitter users to share access on a single dev account, however, this is mainly geared toward commercial use and is intended for development for premium APIs ($), and is outside the scope of this guide. For more, see Twitter's FAQ.

   - If you select "for my own personal use," you will be prompted to fill in an Account Name and to select your Primary Country of Operation. Both are required.
   - If you select "for my organization," you are required to fill out the following:
     - Organization Name (the name of your account)
     - Legal Entity Name (may be the same as Organization Name)
     - Organization Twitter @username
     - Organization Primary Country of Operation
     - Customer Location (select 'Not Applicable(we do not have customers)')
     - Categorize your Organization (select 'Academic')
     - Industries served (select Academic)

3. In the next section on 'Use Case Details', answer several questions about your project and why you are building an app to gather twitter data. You can provide answers similar to the ones below if relevant to your project. **Most importantly note that this is for academic purposes and that you will not be Tweeting, Retweeting, or liking content.**

   If your project does involve analysis of twitter content, an answer for question #2 might look something like "Yes, my project will analyze tweets using text analysis, word clouds, word frequency, and word association using R."

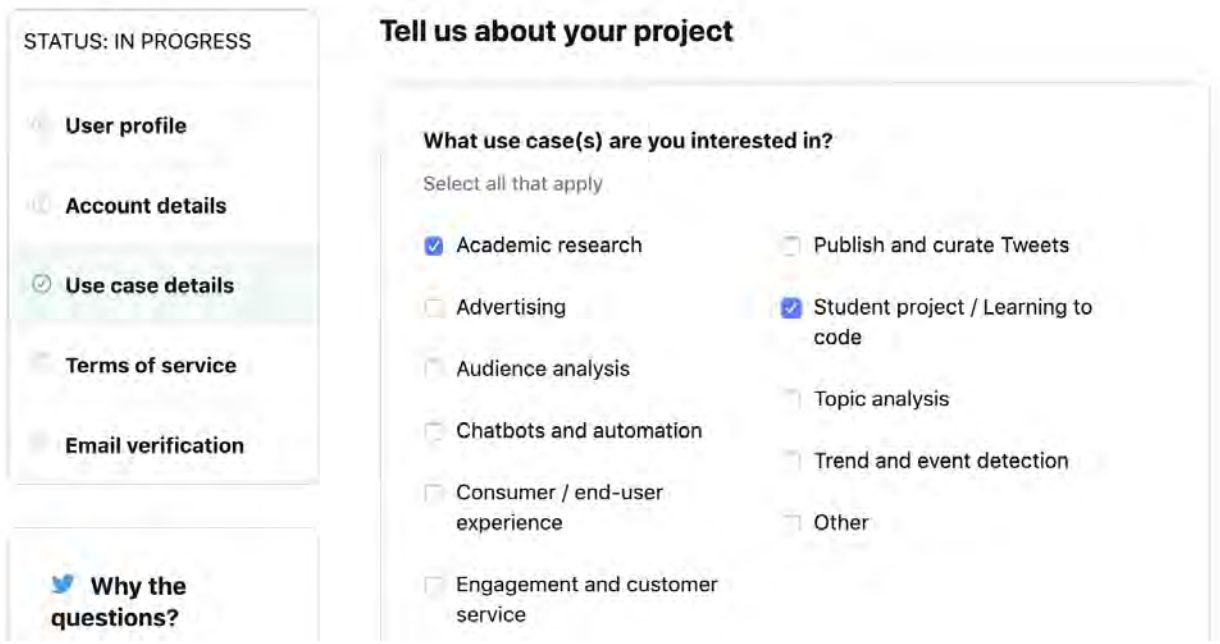


Figure 23: Screen shot of Twitter Developers Site

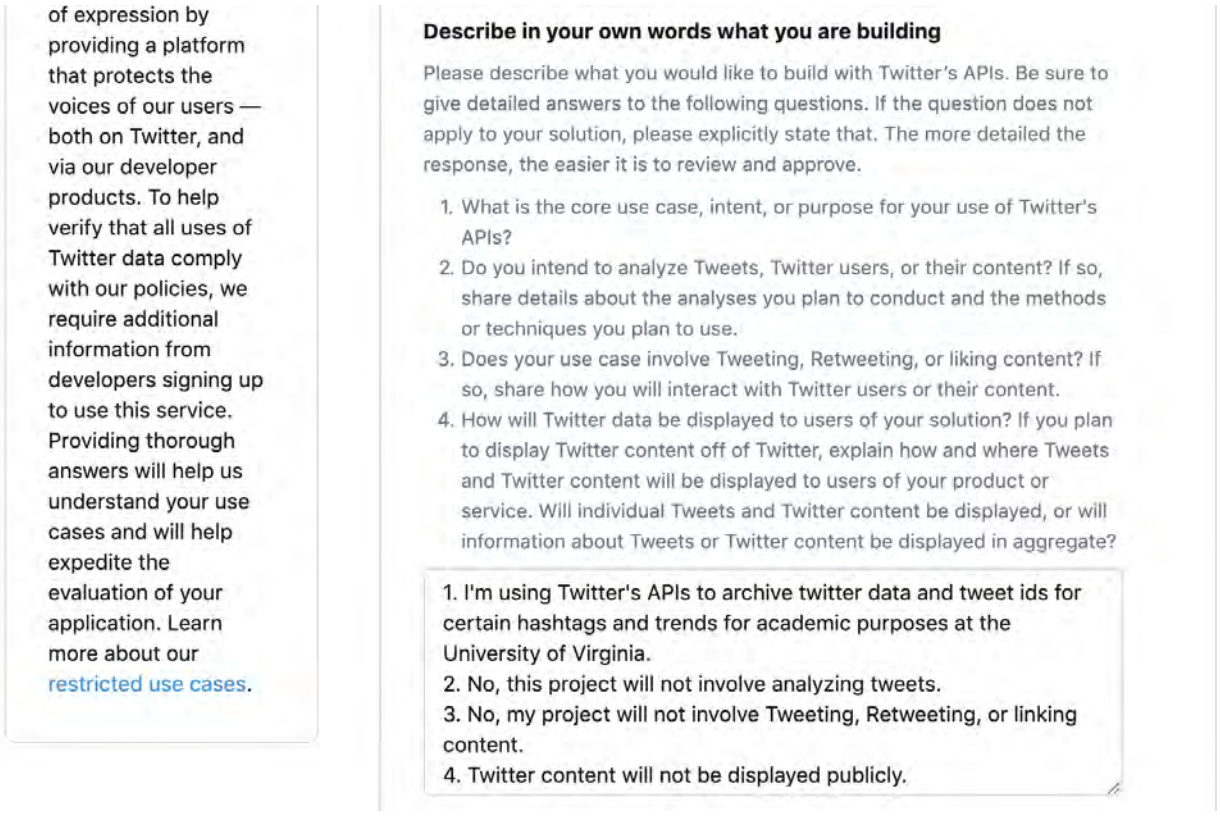


Figure 24: Screen shot of Twitter Developers Site

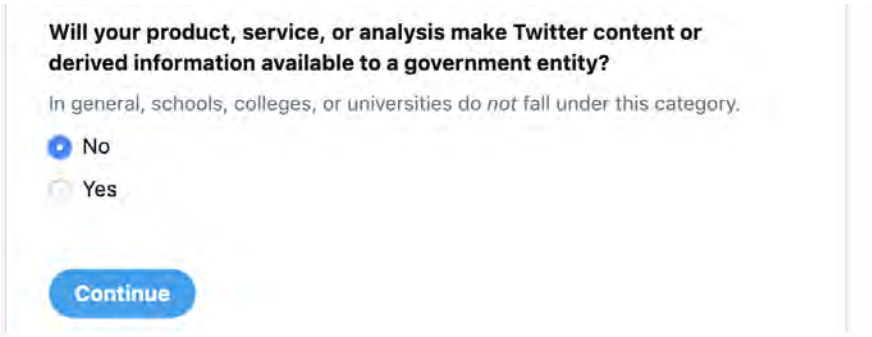


Figure 25: Screen shot of Twitter Developers Site

4. Read and Accept the Twitter Terms of Service.
5. Verify your Twitter Development Account via the email associated with your Twitter account.

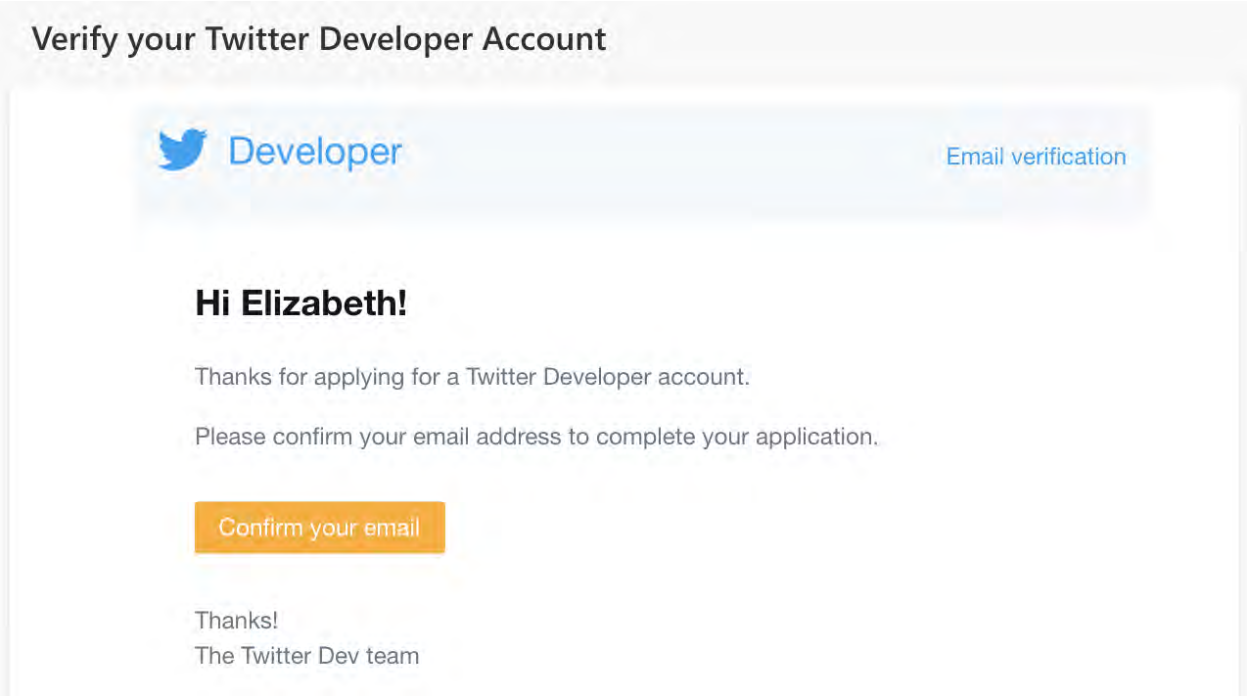


Figure 26: Screen shot of Twitter Developers Site

Creating a Twitter App

1. Once you have your Twitter Developer Account set up, you can register an application at developer.twitter.com/en/apps. Click 'Create an App' to begin:
2. Fill out the required parts of the form for App Details. For the Website URL, you can simply put the URL for your twitter account, or any website you are affiliated with:
3. Click 'Create'. A pop-up may appear for reviewing the Twitter Developer Terms, click 'Create' to continue.
4. You now have a registered Twitter app! You can edit any of these fields later from your Developer Account Apps page.

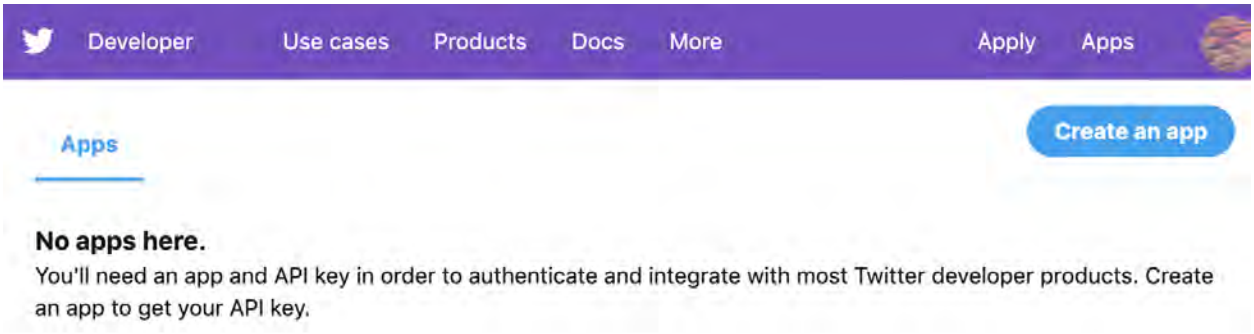


Figure 27: Screen shot of Twitter Developers Site

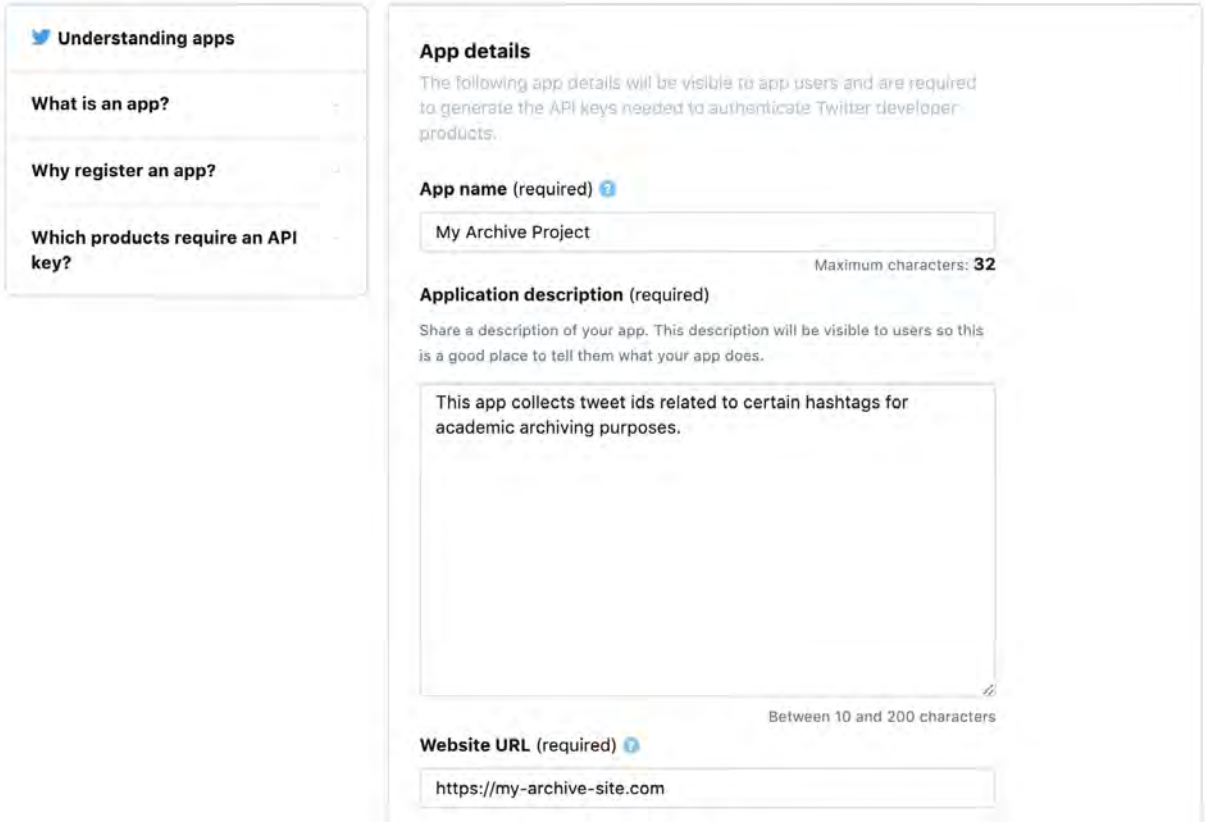


Figure 28: Screen shot of Twitter Developers Site

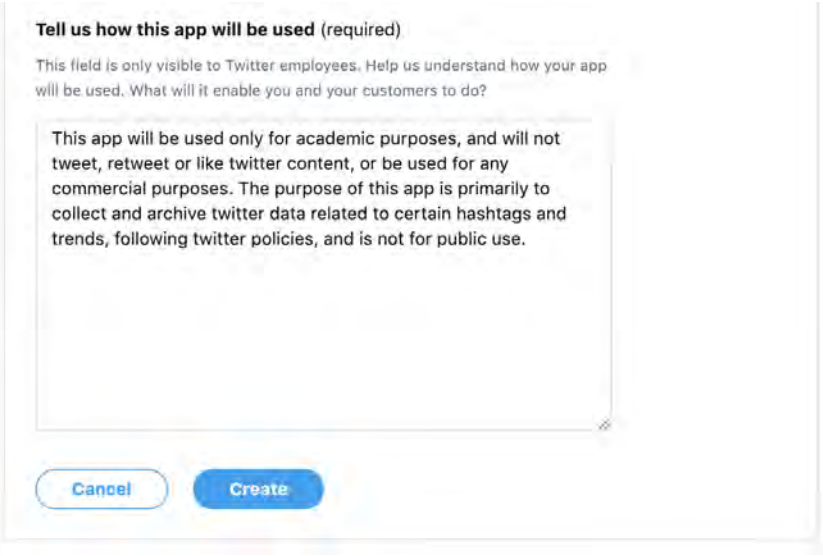


Figure 29: Screen shot of Twitter Developers Site

Accessing Keys and Tokens

1. From your Developer Account Apps page, find your app and click 'Details'.

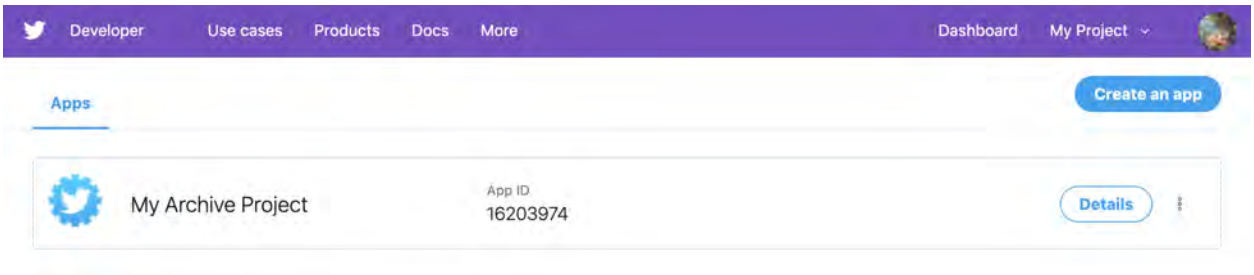


Figure 30: Screen shot of Twitter Developers Site

2. Select the option for 'Keys and Tokens'. On this page you will find your Consumer API keys. Under 'Access token & access token secret', click 'create' to generate.

   Note down for use with Twarc these four alphanumeric values:

   - Consumer API key
   - Consumer API secret
   - Access token
   - Access token secret

Next, get Twarc up and running: Installing and Configuring Twarc

---

Installing and Configuring Twarc

Published: April 25, 2019

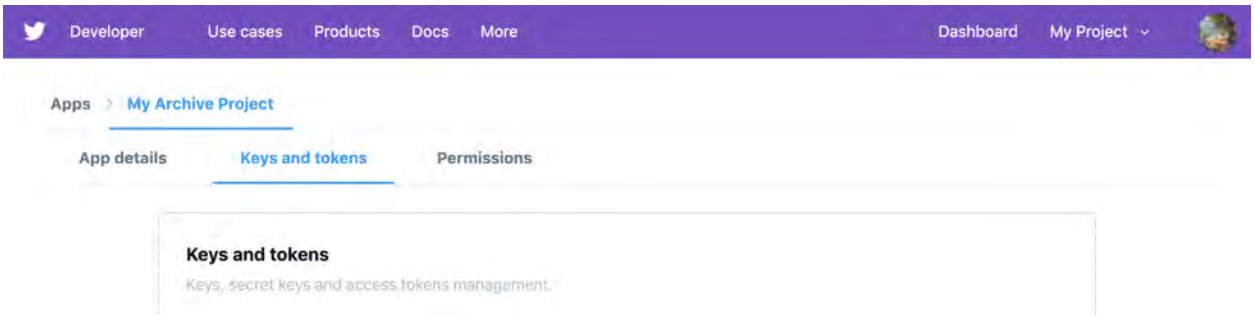


Figure 31: Screen shot of Twitter Developers Site

Before you Begin

1. Before using Twarc, you will need to register a Twitter application, and have your consumer key, consumer secret, access token, and access token secret on hand.
2. Twarc requires familiarity with using the command line to navigate your file system, configure Twarc, and run queries. For introductory lessons on using the command line, see these tutorials for Mac and Windows users from the UNLV Libraries, here.

   The Command Line tutorial is part of UNLV Libraries' Twitter Data Tutorial Series, which features ten tutorials that take you step-by-step through the design, collection, and documentation process of curating a collection of Twitter data, as well as tutorials on tools for data analysis. This series is focused on using Twarc and offers another resource, in addition to this toolkit.
3. Download and install the latest version of Python, here. Twarc works with with either version 2 or 3.

   If you are a Mac user, you may already have Python installed. To check, open your terminal and run the command:

   ```
   python -V
   ```

   or

   ```
   python --version
   ```

Twarc for Mac OS

For Windows users, click here

Installing Twarc

1. Open the Terminal application (located in the applications folder)
2. Pip install Twarc by entering the following command:

   ```
   pip install twarc
   ```

   Note: Pip is already installed if you are using Python 2 at least version 2.7.9 or Python 3 at least version 3.4

   macOS users also have the option of installing Twarc via homebrew using the command:

   ```
   brew install twarc
   ```

Configuring Twarc

To get started, you will need to tell Twarc about your application API keys and grant access to one or more Twitter accounts. Follow these directions to configure Twarc:

1. Enter the following command in Terminal:

   ```
   twarc configure
   ```

2. Twarc will ask you to enter several keys. You should have these keys ready to go after registering an application.

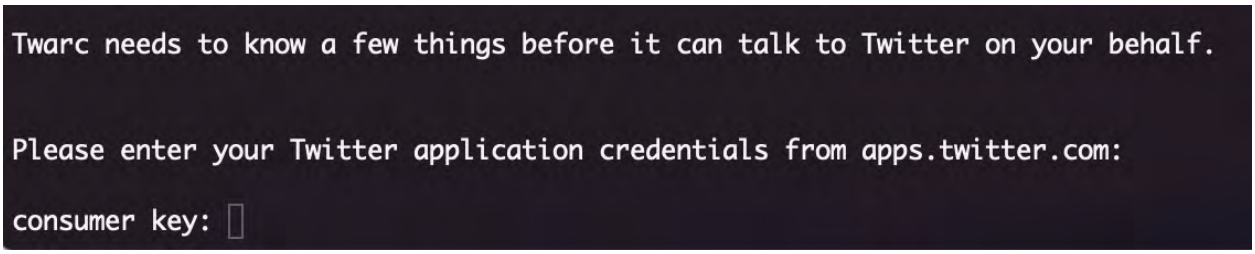


Figure 32: Screen shot of Terminal

   Type or copy/paste your consumer key, then press 'enter'. Next, Twarc will ask for your consumer secret. Type or copy/paste your consumer secret, then press 'enter'.

3. Next, Twarc will ask you to log into your twitter account. Copy/paste the provided URL into your browser to authorize your application.

4. Following the provided URL will bring you to a page that looks something like this:

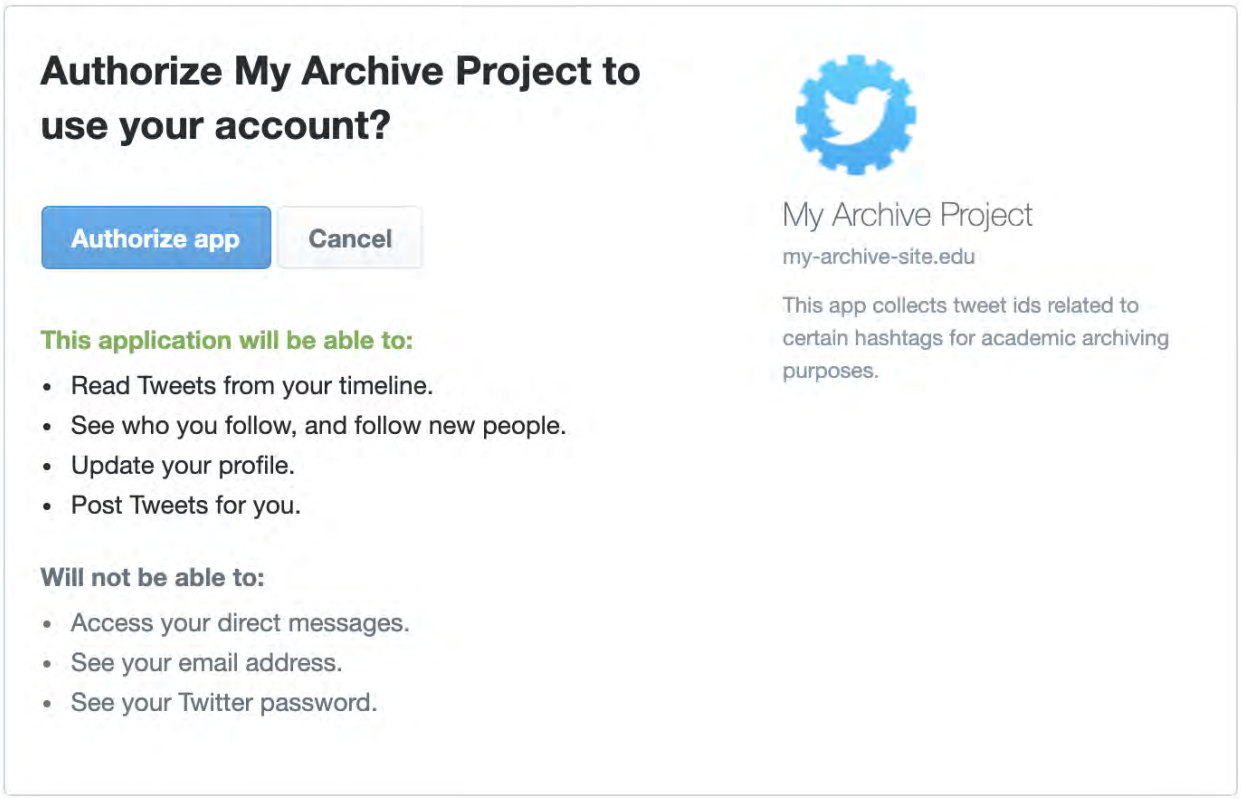


Figure 33: Screen shot of twitter authorize page

   Click 'Authorize app'. A pin number will be provided, Type this pin into your terminal, as seen in step 3, then press 'enter'.

5. After entering your pin, your terminal should give a message telling you where your twitter credentials are saved, followed by "Happy twarcing!" This means your twarc installation is configured.

Next, learn the basics of collecting twitter data: Twarc Introductory Lesson

Twarc for Windows

Installing Twarc

1. Open PowerShell (to open PowerShell, use the taskbar to search for PowerShell and select 'Windows PowerShell')
2. Pip install Twarc by entering the following command:

   ```
   pip install twarc
   ```

   Note: Pip is already installed if you are using Python 2 at least version 2.7.9 or Python 3 at least version 3.4

Configuring Twarc

To get started, you will need to tell Twarc about your application API keys and grant access to one or more Twitter accounts. Follow these directions to configure Twarc:

1. Enter the following command in PowerShell:

   ```
   twarc configure
   ```

2. Twarc will ask you to enter several keys. You should have these keys ready to go after registering an application.

   Type or copy/paste your consumer key, then press 'enter'. Next, Twarc will ask for your consumer secret. Type or copy/paste your consumer secret, then press 'enter'.

Next, learn the basics of collecting twitter data: Twarc Introductory Lesson

---

Introductory Lesson on Twarc for Twitter Data Collection

Published: April 25, 2019

Why Collect Tweets?

Twitter is undeniably a part of the cultural landscape of the modern world, and its content represents a new form of the historical record, one that archivists around the world are actively working to preserve.

Particularly relevant to researchers interested in exploring popular movements, the dynamics of fast-moving socio-political events, and the digital footprint of contemporary culture, Twitter archiving is becoming increasingly more prominent in the catalog of activities taking place under the umbrella of digital curation.

Like other digital archiving efforts, the collection and curation of Twitter data does involve some unique challenges, not least of which is the functionally infinite size of the dataset. For this reason, a large part of collecting, archiving, and providing access to Twitter data involves deciding what to collect and why. Any Twitter archive will reflect the intention of its collector - from the tools used to gather the data to the kind of search used to collect, Twitter archives are always a directed slice of the ever-expanding pie that is social media.

Data Collection Process

The University of Virginia currently archives collections of Twitter data using Twarc, a command line tool and Python library developed as part of the Documenting the Now project. Twarc provides several different methods for collecting Twitter data, which are outlined below. Note that each collection method has a distinct goal in mind, and will have slightly different outputs depending on how the data is collected. It is important to note the kind of collection process used to generate a specific Twitter archive - the table below outlines some of these variations and the effects on their output. The type of collection method should be noted in the description of each Twitter archive, as well as the date and size of the collection, along with the dates of collection and dataset size.

Common Twarc Collection Methods Used in Archiving

For more detailed usage documentation, see the following page on Twarc Commands. You can also skip ahead to the beginner's tutorial on using Twarc. ### Search Collects pre-existing tweets from up to seven days ago that match the given query. This method allows you to gather tweets that are already published, but will not collect any tweets that have been made private or deleted by the time you run your search. ### Filter Initiates a collection process that will gather tweets matching the given query as they are published. This process does not gather tweets that were already published by the time the filter is initialized, but will ensure a high-fidelity capture of all material between the start and end times of the filter. Unlike a search, it is also not limited to a single week, and can be run as long as you want.
Sample This method returns a random sampling of tweets. ### Timeline The timeline command will collect the most recent tweets from a single user. Like search, it can only go back seven days.

Side by Side: Search vs Filter

The following example illustrates some of the differences between using search and filter methods with the same arguments and output. In this case, we are looking for all tweets containing the keyword "charlottesville".

|| Search | Filter | || ——— | ——— | |**Twarc Command**| twarc search charlottesville > charlottesville-tweets.jsonl | twarc filter charlottesville > charlottesville-tweets.jsonl | |**Description** | Searches all existing tweets from the present to seven days earlier (this time limit is set by Twitter).| Initiates a collection process that will gather tweets, as they are published, until you tell it to stop.| |**Output** | Returns results as a JSONL file, where each JSON object = a single tweet.| Returns results as a JSONL file, where each JSON object = a single tweet.| |**Notes** | Captures all relevant tweets up to a week prior to the search date, but will not capture any tweets that have been either privatized or deleted by the time the search is run.| A higher-fidelity collecting method that archives tweets in real time. Twarc filtering requires a dedicated machine that stays on for the duration of the process.|

The blue text in the twarc command field is modifiable: this is where you tell Twarc what you want to search for, and where you want to name the JSONL file that will contain the results.

Dehydrated and Rehydrated Data Sets

Twitter API's Terms of Service does not allow for making large amounts of raw Twitter data available on the Web. The fully-hydrated Twitter data you collect using Twarc can be used for research and archived for local use, but cannot be shared publicly. However, Twitter does allow files of tweet identifiers to be publicly shared. This is useful for when you would like to make a dataset of tweets available, for example, in DocNow's Tweet Catalogue. This is referred to as a "dehydrated" data set - each tweet is reduced to its unique ID number, and a list of these IDs is saved as a text document. You can use Twarc to dehydrate your data set.

A dehydrated data set can be "rehydrated" using Twarc or another program of your choice - Twitter will take each ID and search the current Twitterverse for a corresponding tweet. If it locates a tweet that matches the ID, it will return the original JSON code for that tweet. This process is called "rehydration", and will return a JSON file containing the data for every tweet that it was able to locate. It's important to note that any tweet rehydrator can only check for tweets that are currently live - they cannot find tweets that have been deleted or made private since the original collection event.

Start Collecting: Twarc Command Basics

Collecting Tweets

Let's start collecting tweets! This introductory lesson will use Twarc's search command to return tweets containing 'BlackLivesMatter' occurring within the past 7 days. For further instructions on using other commands, see the following page on Twarc Commands. This lesson assumes some basic familiarity with the command line (Terminal for Mac users, PowerShell for Windows users).

1. Create a new folder on your desktop titled 'BlackLivesMatter_Tweets'.
2. Open Terminal (for Mac) or PowerShell (for Windows)
3. Change directories to navigate into the 'BlackLivesMatter_Tweets' folder by typing into the command line `cd` followed by the path of your folder:

   ```
   cd desktop/BlackLivesMatter_Tweets
   ```

   Hit 'return' to complete the command.

   You can also navigate to your folder by typing `cd` followed by a space, then drag your folder into the Terminal or PowerShell window. Hit return to complete the command

   To check if you are in the correct place, use the command `pwd` to display the path of your current directory. It should end with `/BlackLivesMatter_Tweets`
4. Now that you are in the right directory, enter the following command to start collecting tweets:

   ```
   twarc search blacklivesmatter > blacklivesmatter_tweets.jsonl
   ```

   Tip: You can copy and paste these commands into Terminal or PowerShell to avoid errors

   Note: Your collection may take some time to return all Tweets. You can tell when the process is complete when it returns to the shell ($) (or PowerShell (PS)) prompt:

```
:BlackLivesMatter_Tweets eam5hc$ twarc search blacklivesmatter > blacklivesmatter_tweets.jsonl
:BlackLivesMatter_Tweets eam5hc$
```

Figure 34: Screen shot of terminal twarc search

5. While the prompt is running, check to make sure your command was successful by clicking on your 'BlackLivesMatter_Tweets' folder. Inside you should see your 'blacklivesmatter_tweets.jsonl' file and a 'twarc.log' file.

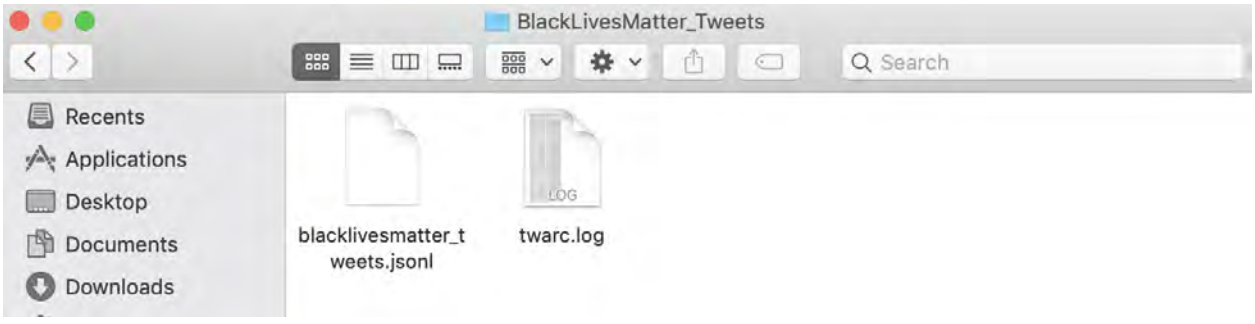


Figure 35: Screen shot of 'BlackLivesMatter_Tweets' folder contents

6. For the purposes of this tutorial, you can cut the search short. After entering the initial search command, wait 5 minutes and then enter Ctrl + C to stop the search.

Note: If you wanted to collect the full set of tweets, you would wait until the process was finished. You can tell when a process is complete when it returns to the shell ($) prompt.

7. You're done! Your tweets are in JSON format in your 'blacklivesmatter_tweets.jsonl' file.

Dehydrate your Dataset

Each Tweet in your dataset has a unique identifier. Twarc's dehydrate command will generate a list of tweet ids from a file of tweets. This lesson will show you how to dehydrate your 'blacklivesmatter_tweets.jsonl' file so that you can share your dataset while keeping to Twitter API's Terms of Service.

1. Navigate into you 'BlackLivesMatter_Tweets' folder using the `cd` command, or if you are already there, check you directory location using the `pwd` command.
2. Enter the command `ls` to list all files in the directory. You should see the files 'blacklivesmatter_tweets.jsonl' and 'twarc.log'. It will look something like this:

```
'your-computer-name':BlackLivesMatter_Tweets 'your-user-name'$ ls
blacklivesmatter_tweets.jsonl   twarc.log
```

3. To dehydrate your tweets, enter the following command:

```
twarc dehydrate blacklivesmatter_tweets.jsonl > blacklivesmatter_tweet_ids.txt
```

4. You should now have a text file containing the unique tweet ids of all tweets in your dataset. Your tweet ids are located in your 'BlackLivesMatter_Tweets' folder in the 'blacklivesmatter_tweet_ids.txt' file.

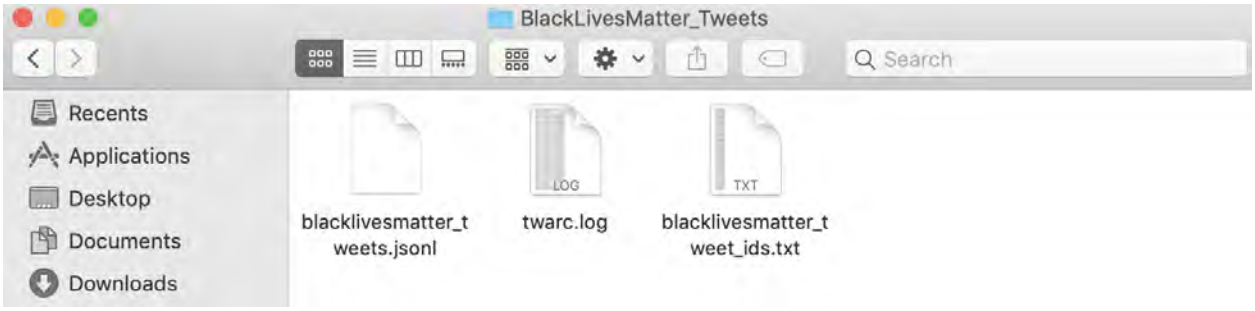


Figure 36: Screen shot of 'BlackLivesMatter_Tweets' folder contents

5. Open the blacklivesmatter_tweet_ids.txt file to see the list of unique tweet ids:

Figure 37: Screen shot of txt file contents

Rehydrate a Dataset

Twarc's hydrate command will read your file of unique identifiers and write out the tweet JSON for them using Twitter's status/lookup API. This is useful if you have a set of Twitter ids from another institution and would like to view the full dataset. Documenting the Now has a collection of Tweet ids that you can explore and rehydrate

here, but for this tutorial we will use the 'blacklivesmatter_tweet_ids.txt' file you created when you dehydrated your dataset.

1. Navigate into you 'BlackLivesMatter_Tweets' folder using the `cd` command, or if you are already there, check you directory location using the `pwd` command.
2. Rehydrate your dataset by entering the following command:

   ```
   twarc hydrate blacklivesmatter_tweet_ids.txt > blacklivesmatter_tweets_hydrated.jsonl
   ```

3. You now have your tweets in JSON format ready to go in your 'BlackLivesMatter_Tweets' folder. Check your folder to confirm your .jsonl file is there.

Now you're ready for more complex Twarc commands that will allow you to create a collection to fit your research needs: Twarc Commands.

Twarc Commands

After you've become familiar with the basics of using Twarc to perform a search command to collect Twitter data, dehydrate an existing dataset, and hydrate a list of unique Tweet ids, you can move on to more complex commands to tailor your search to your research needs.

Below you will find a list of commands that will allow you to create a more targeted collection. This documentation is from the Twarc Usage Guide.

Published: April 25, 2019

Search

This uses Twitter's search/tweets to download *pre-existing* tweets matching a given query.

```
twarc search blacklivesmatter > tweets.jsonl
```

It's important to note that `search` will return tweets that are found within a 7 day window that Twitter's search API imposes. If this seems like a small window, it is, but you may be interested in collecting tweets as they happen using the `filter` and `sample` commands below.

The best way to get familiar with Twitter's search syntax is to experiment with Twitter's Advanced Search and copy and pasting the resulting query from the search box. For example here is a more complicated query that searches for tweets containing either the #blacklivesmatter or #blm hashtags that were sent to deray.

```
twarc search '#blacklivesmatter OR #blm to:deray' > tweets.jsonl
```

Twitter attempts to code the language of a tweet, and you can limit your search to a particular language if you want:

```
twarc search '#blacklivesmatter' --lang fr > tweets.jsonl
```

You can also search for tweets with a given location, for example tweets mentioning *blacklivesmatter* that are 1 mile from the center of Ferguson, Missouri:

```
twarc search blacklivesmatter --geocode 38.7442,-90.3054,1mi > tweets.jsonl
```

If a search query isn't supplied when using `--geocode` you will get all tweets relevant for that location and radius:

```
twarc search --geocode 38.7442,-90.3054,1mi > tweets.jsonl
```

Filter

The `filter` command will use Twitter's statuses/filter API to collect tweets as they happen.

```
twarc filter blacklivesmatter,blm > tweets.jsonl
```

Please note that the syntax for the Twitter's track queries is slightly different than what queries in their search API. So please consult the documentation on how best to express the filter option you are using.

Use the `follow` command line argument if you would like to collect tweets from a given user id as they happen. This includes retweets. For example this will collect tweets and retweets from CNN:

```
twarc filter --follow 759251 > tweets.jsonl
```

You can also collect tweets using a bounding box. Note: the leading dash needs to be escaped in the bounding box or else it will be interpreted as a command line argument!

```
twarc filter --locations "\-74,40,-73,41" > tweets.jsonl
```

If you combine options they are OR'ed together. For example this will collect tweets that use the blacklivesmatter or blm hashtags and also tweets from user CNN:

```
twarc filter blacklivesmatter,blm --follow 759251 > tweets.jsonl
```

Sample

Use the `sample` command to listen to Twitter's statuses/sample API for a "random" sample of recent public statuses.

```
twarc sample > tweets.jsonl
```

Dehydrate

The `dehydrate` command generates an id list from a file of tweets:

```
twarc dehydrate tweets.jsonl > tweet-ids.txt
```

Hydrate

Twarc's `hydrate` command will read a file of tweet identifiers and write out the tweet JSON for them using Twitter's status/lookup API.

```
twarc hydrate ids.txt > tweets.jsonl
```

Twitter API's Terms of Service discourage people from making large amounts of raw Twitter data available on the Web. The data can be used for research and archived for local use, but not shared with the world. Twitter does allow files of tweet identifiers to be shared, which can be useful when you would like to make a dataset of tweets available. You can then use Twitter's API to *hydrate* the data, or to retrieve the full JSON for each identifier. This is particularly important for verification of social media research.

Users

The `users` command will return User metadata for the given screen names.

```
twarc users deray,Nettaaaaaaaa > users.jsonl
```

You can also give it user ids:

```
twarc users 1232134,1413213 > users.jsonl
```

If you want you can also use a file of user ids, which can be useful if you are using the `followers` and `friends` commands below:

```
twarc users ids.txt > users.jsonl
```

Followers

The `followers` command will use Twitter's follower id API to collect the follower user ids for exactly one user screen name per request as specified as an argument:

```
twarc followers deray > follower_ids.txt
```

The result will include exactly one user id per line. The response order is reverse chronological, or most recent followers first.

Friends

Like the `followers` command, the `friends` command will use Twitter's friend id API to collect the friend user ids for exactly one user screen name per request as specified as an argument:

```
twarc friends deray > friend_ids.txt
```

Trends

The `trends` command lets you retrieve information from Twitter's API about trending hashtags. You need to supply a Where On Earth identifier (`woeid`) to indicate what trends you are interested in. For example here's how you can get the current trends for St Louis:

```
twarc trends 2486982
```

Using a `woeid` of 1 will return trends for the entire planet:

```
twarc trends 1
```

If you aren't sure what to use as a `woeid` just omit it and you will get a list of all the places for which Twitter tracks trends:

```
twarc trends
```

If you have a geo-location you can use it instead of the `woedid`.

```
twarc trends 39.9062,-79.4679
```

Behind the scenes twarc will lookup the location using Twitter's trends/closest API to find the nearest `woeid`.

Timeline

The `timeline` command will use Twitter's user timeline API to collect the most recent tweets posted by the user indicated by screen_name.

```
twarc timeline deray > tweets.jsonl
```

You can also look up users using a user id:

```
twarc timeline 12345 > tweets.jsonl
```

Retweets

You can get retweets for a given tweet id like so:

```
twarc retweets 824077910927691778 > retweets.jsonl
```

Replies

Unfortunately Twitter's API does not currently support getting replies to a tweet. So twarc approximates it by using the search API. Since the search API does not support getting tweets older than a week twarc can only get all the replies to a tweet that have been sent in the last week.

If you want to get the replies to a given tweet you can:

```
twarc replies 824077910927691778 > replies.jsonl
```

Using the `--recursive` option will also fetch replies to the replies as well as quotes. This can take a long time to complete for a large thread because of rate limiting by the search API.

```
twarc replies 824077910927691778 --recursive
```

Lists

To get the users that are on a list you can use the list URL with the `listmembers` command:

```
twarc listmembers https://twitter.com/edsu/lists/bots
```

Additional Resources for Social Media Data Collection

Twarc

- UNLV Libraries' Twitter Data Tutorial Series, which features ten tutorials that take you step-by-step through the design, collection, and documentation process of curating a collection of Twitter data, as well as tutorials on tools for data analysis. This tutorial series focuses on using Twarc for twitter data collection.

Hydrator

Hydrator is a downloadable app for your computer developed by DocNow for hydrating Twitter ID datasets. Hydrator does not require a Twitter developer account or registered app, and requires no coding knowledge. It streamlines the process of linking your Twitter account, hydrating tweet ID lists, and exporting the raw JSON data into csv for easy access to the dataset.

This tool is great for researchers who want to do a quick check of the contents of a dataset in Excel - the csv export function in the Hydrator makes it easy to build a spreadsheet of tweets.

For access to a step-by-step guide to installing, configuring and using the Hydrator, click here.

Tweet ID Datasets Catalog

DocNow maintains a catalog of open-access, dehydrated Twitter datasets, with information for each including size, collection date, topic tags and a brief description. This is a great place to start if you want to explore the range of Twitter data currently being collected by scholars active in the field. To browse the catalogue, click here.

About the Toolkit

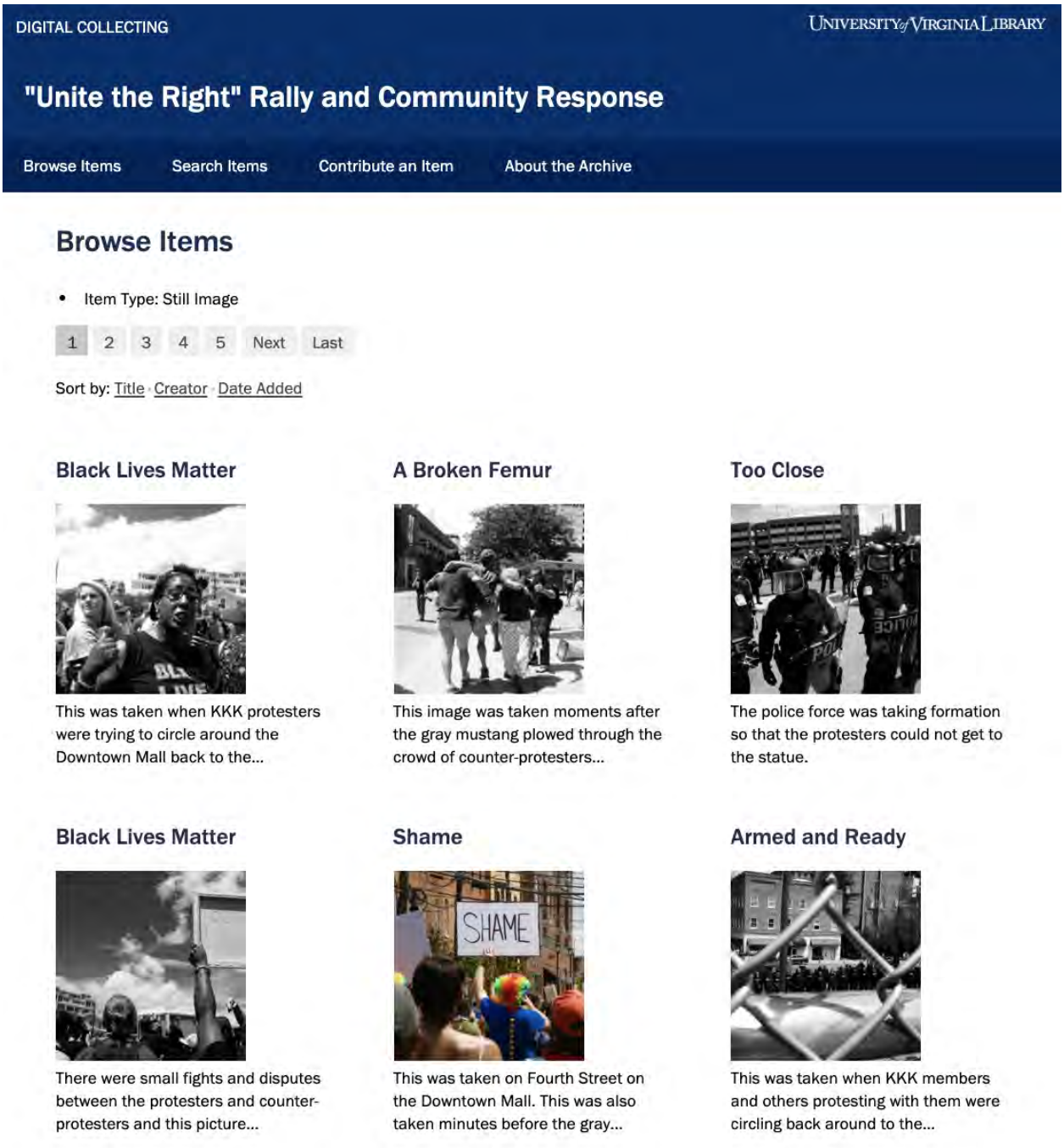
As the events of Friday August 11th and Saturday August 12th, 2017 in Charlottesville unfolded, Library staff at the University of Virginia actively watched the news and began capturing information from websites and social media. When Library administration met the following Tuesday morning, we were asked if we could create a site that allowed community members to contribute their photos and videos online. While the UVA Library had

some experience documenting and collecting digital content after a major news event, this was the first time we attempted to create a collecting site so quickly after the events occurred.

We consulted with other institutions who had created similar sites, and built on their previous work. We set up workflows to ensure that no malware or viruses were upload in the process. It took us a little over three weeks to launch our online collecting tool, far longer than we had initially hoped. The lessons learned from the site launch of the University of Virginia Library's Digital Collecting site, *"Unite the Right" Rally and Community Response* has led to the development of this toolkit to help future rapid response collecting efforts for ourselves and others.

If you have any questions or comments, please feel free to contact us at digital_collecting@virginia.edu.

Visit our digital collection site:



Credits

This toolkit was a collaborative effort, and made possible in part by a 2018 award from the Catalyst Fund at Lyrasis. We are grateful to everyone who brought their expertise to this project.

Special thanks to:

- Kara M. McClurken, Director, Preservation Services, UVA Library
- Jeremy Boggs, Head of Research and Development, Scholars' Lab, UVA Library
- Elizabeth A. Mitchell, Community Advocate, Scholars' Lab, UVA Library
- Spalding Lewis, Research Intern, UVA Library
- Lauren Work, Digital Preservation Librarian, UVA Library

Additional Resources from:

- Omeka, developed by the Roy Rosenzweig Center for History and New Media
- Documenting the Now
- UNLV Libraries' Twitter Data Tutorial Series
- GWU Libraries' Social Feed Manager (SFM), Building Social Media Archives: Collection Development Guidelines
- An Invitation Towards Social Justice in the Digital Humanities
- Michelle Caswell, "Toward a Survivor-Centered Approach to Human Rights Archives: Lessons from Community-Based Archives." Archival Science 14: 3-4 (2014): 307-322.
- Anti-oppression principles compiled by the Center for Story-Based Strategy