

```
{
  "created_at": "Thu Apr 06 15:24:15 +0000 2017",
  "id_str": "850006245121695744",
  "text": "1\ Today we\u2019re sharing our vision for the future of the Twitter API platform!\nhttps://t.co/XweGngmxlP",
  "user": {
    "id": 2244994945,
    "name": "Twitter Dev",
    "screen_name": "TwitterDev",
    "location": "Internet",
    "url": "https://dev.twitter.com/",
    "description": "Your official source for Twitter Platform news, updates & events. Need to know what's new? Visit https://dev.twitter.com/ for the latest news.",
    "entities": {
      "hashtags": [
        {
          "text": "TapIntoTwitter",
          "indices": [
            118, 138
          ]
        }
      ],
      "urls": [
        {
          "url": "https://t.co/XweGngmxlP",
          "expanded_url": "https://t.co/XweGngmxlP",
          "display_url": "https://t.co/XweGngmxlP",
          "indices": [
            58, 88
          ]
        }
      ],
      "user_mentions": [ ]
    },
    "place": {
      "name": "Building the Future of the Twitter API Platform",
      "full_name": "Building the Future of the Twitter API Platform",
      "coordinates": {
        "type": "Point",
        "coordinates": [
          37.6213, -122.3319
        ]
      },
      "country": "US",
      "geo": {
        "type": "Point",
        "coordinates": [
          37.6213, -122.3319
        ]
      }
    },
    "user_mentions": [ ]
  },
  "retweet_count": 0,
  "reply_count": 0,
  "retweeted_status": null,
  "favorited": false,
  "retweeted": false,
  "lang": "en"
}
```

Collection Documentation

Documenting a Collection of Twitter JSON Data

The role of documentation is central to the archival process of data. Without a detailed description of when and how the data was collected, the dataset is effectively useless to the outside world. This tutorial will walk you through the process of documenting Twitter datasets with the ultimate goal of making the data available to researchers both in and outside of your institution. It is recommended that you complete the second and [third](#) tutorials in this series before completing this tutorial.

Difficulty level: Beginner

Prerequisite(s)

- [Tweet JSON](#)
- [Command Line](#)
- [Collection Design](#)
- [Collection with Twarc](#)

Material(s)

- [Open the Collection Documentation Checklist in a new tab](#)

Lesson objectives

- Complete the Collection Documentation Checklist
- Create a README file for a Twitter data collection

Key Terms

- Twarc
 - A command line tool and python library for collecting Twitter JSON data
- README
 - A plain text (.txt) file that contains information for users about a software program, utility, or other application.

Table of Contents

Lesson objectives	1
Key Terms	1
Table of Contents	2
Introduction	2
Step 1: Collection Title	3
Step 2: Collection Description	3
Step 3: Twarc Command	3
Step 4: Collection Time & Date	4
Step 5: Preparing a README	4
Step 6: Creating A README	5

Introduction

Through the thorough documentation of your Twitter data, you can help other researchers better understand your methods. While this level of documentation may seem intimidating, the easiest way to avoid a lengthy process is to record information as you receive it. This tutorial will help you structure future collections in a way that prioritizes documenting each step. By the time you have finished your collection, you won't have to scramble to find collection dates and times. Staying organized will allow you to focus on your research with the added benefit of helping anyone interested in your work to better understand your unique approach to data collection.

Important: If you haven't already, [open the Collection Documentation Checklist](#) in a new tab. Once you have finished the tutorial, make a copy of the spreadsheet and use it to prepare for future collections.

[illegible]

Collection Documentation Checklist

Step 1: Collection Title

Create a collection title that will give an outside viewer an idea of what the collection contains. This should be concise and capture the original purpose for the collection you outlined in the collection design tutorial.

Step 2: Collection Description

The collection description should be brief and contain important topline of the dataset. Consider including the following information in your collection description:

- Collection purpose
- Relevant events and/or dates
- Total number of tweets in the collection
- Brief description of how data were gathered
- Include search terms if applicable
- Collection dates
- Collection times

Step 3: Twarc Command

Record each Twarc query on a separate line in the collection documentation checklist. This will help you prepare your readme file later on. The command can be copied directly from the command line to avoid errors. Properly documenting your search terms will help other

researchers have a better understanding of your method, allowing them to replicate it if necessary.

Step 4: Collection Time & Date

One of the most important steps in the collection documentation process is recording the time and date your collection started and finished. While it may be more difficult to determine exactly when your collection finished, including the date will allow other researchers to gain access to the collection process behind your dataset.

Step 5: Preparing a README

Think of the README file as your way of communicating your work to the world. It should be brief—containing the topline of your dataset and details of collection that would be valuable to researchers looking to use your data in their research. It should also be descriptive enough so that anyone reviewing your research has a clear idea of how your collection came to be. By following the Collection Documentation Checklist, you can better prepare yourself to create a Readme file that accurately reflects the contents of your collection.

In creating the 1 October Twitter Collection README, the following sections below were included. Read through them and start to think about how you want to describe and share your collection:

- 1. Description**

The collection description describes the purpose of the collection, provides the total number of tweets within the dataset, and the collection timeframe.

- 2. Queries**

Lists the six queries (commands in Twarc) that were made to Twitter's status/lookup API to create the collection. These queries all contain the same search term 'Vegas'. The files have been separated into 24-hour intervals.

- 3. Files**

Shows the file sizes for both the public and private version of the collection.

- 4. File Description**

Describes the type of data within each file (Twitter JSON), the search parameters (7-day search window imposed by Twitter's API), and any additional modifications to the data.

5. Use

The use section incorporates the Twitter Terms of Service to let anyone accessing the public or private dataset know that, by downloading this data, they are agreeing to Twitter's Terms of Service. The purpose of this is to ensure that everyone who accesses the dataset knows the legal boundaries in which they can share and distribute these data, with [a link](#) to Twitter's Terms of Service for further clarification.

6. Prepared by

This section names the creators of the collection. Additional questions about the contents of the collection can be answered by contacting UNLV Libraries' Digital Collections Department.

Step 6: Creating A README

Now that you have an idea of what a README file describing a collection of Twitter data looks like, go ahead and create your own.

Use [this template](#) as a reference to create your own readme file.

Once you're done, change the file name to 'readme' and download the file as a plain text (.txt) file.

🎉 Congratulations! You have created a readme.txt file for your collection. 🎉