

# **Social Media Data in the Academic Environment: Crimson Hexagon in the Library**

Tim Dennis @ University of California, San Diego

Stephanie Tulley @ University of California, Santa Barbara

# Overview

1. Description of the tool
2. Background @ UCSB & UCSD
3. Content provided/Querying
4. Access models & usage policies
5. Challenges & Interesting uses

# What it is: Crimson Hexagon

1. Social media data library
2. Content analysis software program, via Foresight platform
3. Primarily offered as a business intelligence tool
4. Access is purchased by number of users and number of monitors
  - a. We'll get into that more later...

# Content:



Crimson Hexagon

Data Library provides access to billions of posts from providers including:



Twitter (Full firehose July 2010 – )



Tumblr (Full firehose January 2015 – )



Google+

... Blogs, Forums, FB, Instagram Hashtags, VK Keywords, Review sites, News sites, YouTube...

Content sources are dynamic, new ones get added and removed (i.e. Weibo)



Crimson Hexagon @UCSB

## 1. Background:

- a. Many requests for social media data
- b. Faculty & Grad requests since 2013 for CH
- c. UCSB had pockets of access for faculty
  - i. Poli Sci/CS
- d. **New service + Big dreams + \$\$ =  
Perfect time for access**

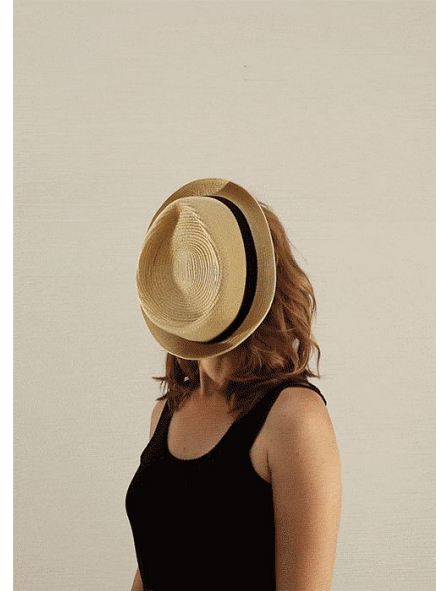


Image: <https://blog.lonewolfmag.com/moving-pictures-romain-laurent/>



Crimson Hexagon @UCSD

1. Many requests for social media data
2. Library explored using Social Feed Manager
3. New PoliSci faculty with IQSS background
4. Strong campus support for 'big data' efforts
5. Library innovation funding piloted CH for 2 years

# Getting data:



Crimson Hexagon

1. Data is selected through the creation of queries, referred to as “monitors” in CH
2. 3 types of monitors (least to most sophisticated):
  - a.  Social
  - b. Buzz
  - c. Opinion

**\*You purchase access by “monitors” and “users”**

# Results:



Crimson Hexagon

1. View results on the monitor page, using dashboard and CH visualization tools
2. Download data for running own analysis/creating visualizations
3.  Up to 10,000 posts allowed for bulk export - per user, per day
  - a. Twitter content is 50,000 user/day
4. Can result in massive data set over time (previous UCSB researcher amassed 1TB of Twitter data using this method)



**Access @UCSB**



# Access @UCSB

## Subscription level:

- a. **50 Buzz/Opinion Monitors**
- b. **20 Social Account Monitors**
- c. Unlimited users

# Access @UCSB

## **Who gets access:**

UCSB students, faculty, and staff conducting academic research can request access to Crimson Hexagon by submitting a form describing the project scope and timeline

# Access @UCSB

## How access is granted:

- Following a meeting with CH admins in the Collaboratory, a specified number of monitors are allocated for a timeframe appropriate to the project.
- CH admins create new user accounts and assign dummy monitors
- Prior to access, users must:
  - 1) View pre-recorded training webinars
  - 2) Sign a T&C agreement

# Access @UCSB

## When & Where:

- Access is (mostly) limited to Collaboratory workstations
  - **Artificially tie a web-based portal to physical space by having staff** log users into their accounts
- Collaboratory is open 7 days a week, 60 hours/wk for walk-in, and by a appointment
- “Power users” are given their own logins and passwords, these are faculty with long-term projects with the Collaboratory

# Access @ UCSD





Crimson Hexagon

@ **UCSD**

**UCSD researchers** conducting academic research can request access to Crimson Hexagon by sending a short description of the project to CH admins in the Library.

The CH admins will then contact them to discuss specific needs and set up the accounts.

CH admins:

1. Create an account and assign users to a team.
2. Make users aware of the [guidelines](#).
3. Encourage users to take advantage of CH help and webinars.
4. Ask users to
5. We do not require users to come to the Library.



Crimson Hexagon

@ **UCSD**

## **Access model:**

- a. 40 Buzz/Opinion Monitors
- b. **16 Social Account Monitors**
- c. Unlimited users





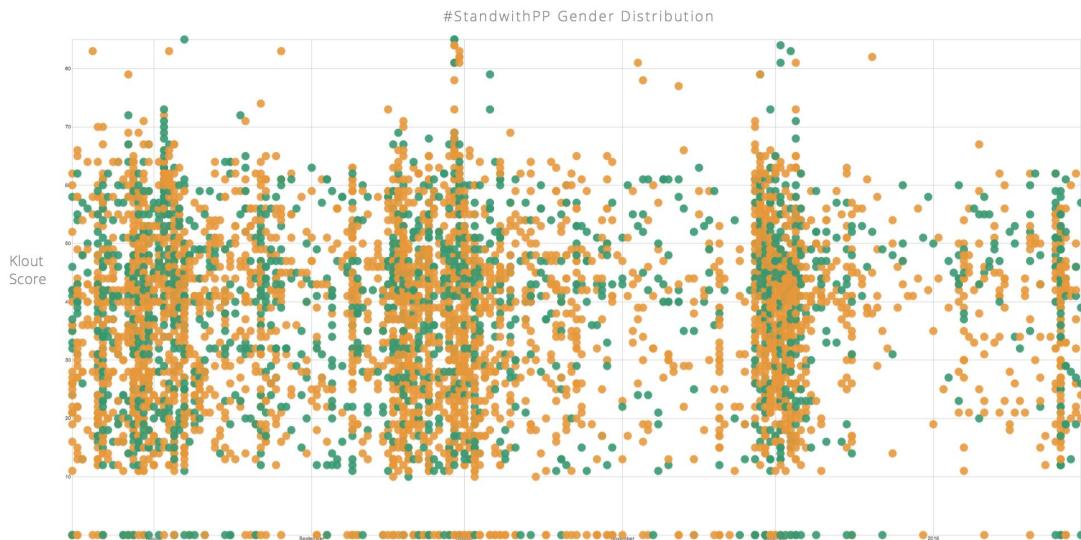
# Ongoing Challenges + Opportunities

# Challenges @UCSB

- Reproducibility is not the primary focus of this tool, and it can be a black box in some instances (like the algorithm for auto-sentiment)
- Demand > Supply
- Time consuming for our staff
  - Users need a lot of support\*
  - Paperwork!!

# Interesting Usages @UCSB

Data Visualization class used CH data to explore topics related to their dissertation research



# VISUAL PROVOCATIONS

**VISUAL PROVOCATIONS: Data, Culture, and Praxis Exhibition**

Created by an interdisciplinary group of 6 UCSB students

On display at UCSB Library, Interdisciplinary Research Collaboratory (2<sup>nd</sup> Floor, Mountain Side) April 27–June 12, 2016

Reception and demonstration: 5 pm Wed., April 27

Exhibition hosted by UC Santa Barbara Library and co-sponsored by Department of Film and Media Studies. Projects developed in the class FAMST 248: Data Visualization as Media Praxis. For more information, see [famst248.wordpress.com](http://famst248.wordpress.com).

SANTA BARBARA LIBRARY  
film & media

# Challenges @UCSD

- Lack of mediation
- Opacity on use, difficult to follow research lifecycle
- Funded with temporary innovation funding

# Interesting Usages @UCSD

As part of a Big Data class, students in the Global Policy school used CH to develop a Fitbit-focused domain dependent dictionary built as the classification criteria for positive or negative attitudes.

*Table 6 Revised Words in Domain-Dictionary for Sentiment Analysis*

Positive						Negative
achieve	burn	intensively	motivation	solved	weightloss	disconnect
achieved	burning	intensity	motivating	solving	workout	disconnected
achieving	burned	loss	obsessive	spoil	:)	disconnecting
addict	exercise	lose	obsessed	spoiled	lol	fall
addicted	exercising	lost	obsessing	spoilng	LOL	falling
addicting	fit	losing	rock	validate	Lol	fell
buy	fitting	friendship	rocks	validated		inaccurate
buying	intense	motivated	rocking	validating		wtf