# ICPSR

SHARING DATA TO ADVANCE SCIENCE



▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● のへで

# Updates on ICPSR's Social Media Archive (SOMAR)

Libby Hemphill

ICPSR

May 2019



### Talk Overview

#### SOMAR Research Activities

Research and Data Practices Survey Reviewing Social Media Research Methods Implications for SOMAR

< □ > < 同 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

SOMAR Development SOMAR Goals SOMAR Infrastructure

## Social Media Research and Data Practices Survey

- Summer 2018
- 73 responses
- 5 sections:
  - 1. general and demographic,
  - 2. data acquisition,
  - 3. data transformation,
  - 4. analysis and visualization, and

▲ロト ▲ □ ト ▲ □ ト ▲ □ ト ● ● の Q ()

5. data sharing and reuse

## Social Media Platforms Used

Platform	%	Ν
Twitter	39.7%	29
Facebook	28.8%	21
Instagram	11.0%	8
Reddit	11.0%	8
Wikipedia	6.8%	5
Tumblr	5.5%	4
Other	4.1%	3
Twitch	2.7%	2
YouTube	2.7%	2
Pinterest	1.4%	1

◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ● □ ● ● ● ●

## **Researcher Data Sharing Practices**

Mechanism	%	Ν
I don't make my data available.	31.5%	23
I make my data available.	46.6%	34
In a repository or archive	15.1%	11
Through a personal website	11.0%	8
Through journal or conference site	8.2%	6
Through a University affiliated website	6.8%	5
Through a third party data provider	5.4%	4
Other	15.1%	11

◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ◆ □ ◆ ○ へ ⊙

Reviewed 40 articles in 4 high-impact journals:

- 1. No one provided access to data
- 2. No one provided enough detail to replicate data collection

▲□▶ ▲□▶ ▲ □▶ ▲ □▶ ▲ □ ● ● ● ●

3. Data collected computationally (e.g., Python scripts) or through third-parties (e.g., Crimson Hexagon)

## How is social media like other social science data?

- 1. Researchers worry about getting scooped
- 2. Preparing data for reuse takes a lot of effort
- 3. Found data requires special manipulation and documentation

< □ > < 同 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

What makes social media data special?

- 1. Data properties: structure, scale, speed
- 2. Data practices: finding, curating, sharing, and storing

< □ > < 同 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

3. Ethics: private owners, PII

## Implications for SOMAR

- 1. richer documentation of provenance,
- 2. explicit documentation of the terms and conditions for acquiring the data, and
- 3. software and/or scripts used to acquire and manipulate the data
- 4. what constitutes a dataset
- 5. observation-level metadata enhancements and declarations

▲□▶▲□▶▲□▶▲□▶ □ のQで

## SOMAR Goals

- 1. Accept data from multiple platforms
- 2. Archive both data and code
- 3. Balance data terms, restrictions, and researcher needs

▲□▶ ▲□▶ ▲ □▶ ▲ □▶ ▲ □ ● ● ● ●

4. Scale responsibly

## SOMAR Technical Infrastructure



## SOMAR Technical Infrastructure



◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 - のへで

## **Ethical Dilemmas**



Figure 1. Visualization of the STEP framework for curating social media data.

#### Mannheimer and Hull 2018

#### Fiesler and Proferes 2018

▲ロト ▲ □ ト ▲ □ ト ▲ □ ト ● ● の Q ()

## Archiving Data + Code

LINKAGE LIBRARY	Sign In Upload Data Browse Data About
Privacy Preserving Record Linkage - Softwa vestorik: © Brian Thome, CSR0 errien: © VI and The © 9133 ablished: © January 11,2019	are System
Project Description Data and Documentation Publications Discussion L	@ Download this project
Infol You must be signed in to add linkages. Sign in	Usage Metrics 🚱 Overall Project Metrics
Clkhash creatings Unlage added by from Thome on 111/2019 4-65-69 PA Creatersty 6: On Insalytics Description: OC Locals hash pill for entity matching Tile/Software Types: O groups Alternative UKE: Or Insalytics and the control dedocs by	Cf Visit Repository     225     5     0       @ Read Me     Download Detailed Metrics   Published Versions
anonlink (EXXXVIIIII) Unique sóded la Virian Thome on UTI/2019 4-88-35 PA Creator(s): On Lanuly(c):	Cf Visit Repository Export Metadata B Read Me Distribution

◆□▶ ◆□▶ ◆目▶ ◆目▶ ▲□▶ ◆□◆

## Archiving Data + Code

#### What does data look like?

reparticitions's nuit, consistent's nuit, consequent "The My B 444-13 4400 4407, "destings" (Consistent's (Data Large Constraint's Application (Lorent Constraint's Application), and and application (Lorent Constraint's Application (Lorent Constraint's Application), and and application (Lorent Constraint's Application), and and application (Lorent Constraint's Application), application (Lorent Constraint), application), application (Lorent Constraint), application), a "19479798", "indices": [15, 31], "newe": "tris Alfredsson", "screen\_nawe" "AlfredssonFris"))), "favorite\_cont": 7, "favorited": false, "gen": null, "if" 1135483884853504, "id\_stri: "1135483886855004, "in\_reply.to\_screen\_nawe": null, "in\_reply.to\_status\_if" null, "in\_reply.to\_screent\_in and "in\_reply.to\_screent\_in and": in\_reply.to\_screent\_in and: "in\_reply.to\_scre guage\_code": "en", "result\_type": "recent"), "place": null, "possibly\_sensitive": false, "retweet\_count": 2, "retweeted": false, "source": "<a href=\"http://twitter.com\" rei=\"mofollow\">hutter Web Client</a=" "text": "Lits Alfredson MalfredsonTris from SND is giving a preview of what's to come at the #IASSIST19 Poster Session. Colu2026 https://t.co/wWM9ZDxcq1", "truncated" true, "user": {"contributors\_enabled": false, "created\_at": "Wed Oct 11 13:26:18 +0000 2017", "default\_profile": true, "default\_profile": true, "default\_profile\_image": false, "description": Thirstitucture for research data valification trubbler forshologists. Teittar pladké svensa och engelika av svabbédet som vabbed viktigt fuadför att gludbför data tilgjubbegiga.", enittars förstrator: fvorsi [1], "un": fursts [Vispag.ur": "sud.gu.ek", "oppanden drt: http://ww.kod.gu.ek", "inders": [0, 2], "ur": "http://ww.kod.gu.ek", com": 20. "colouregents.tent": Side, "follower", "att "inder internation": folg 5), University of the second secon /' 's', 'liste.count': / 'lecalon': 'understeering wering . Them : '24', 'motification': 'acti, provide assignment of the series of the series and the serie "statuses.comt: 768, "time\_zond": null, "translator\_type": "none", "wrl": "https://t.co/wuKlogHOB", "wrc\_offset": null, "verified": false)), "source": "«a href=\"http://twitter.com\" rel=\"nofollow\"sTwitter Web Client</as", "text": "RT @sndSweden: Iris Alfredsson @Alfredsson.Fis from SMD is giving a preview of what's to nere at the #IASISTIP Poster Sesion, Come Learning 2007, "truncated": false, "user": footributors enabled": false, "created att": "Fri Ses 18 09-53:52 +0000 2010". construction of an announce of advance continuous continuous and a set of a set we have a project aimed at delivering a proposa "following": false, "friends\_count": 405, "geo\_enabled": false, "has\_extended\_profile": false, "id": 189078837, "id\_str": "189078837", "is\_translation\_enabled": false, "is translator": false, "lang": "en", "listed count": 4, "location": "Ljubljana", "name": "Janez \u000001001000; motifications: false, "profile\_background\_color" Control of the section of the sec 245915762, "id\_str": "245915762", "indices": [82, 99], "name": "Ashley Jester", "screen\_name": "ashleyrjester")]}, "extended\_entities": {"media": [{"display\_url": "pic.twitter.com/BulWkiquze", "expanded\_url": "https://twitter.com/stephlabou/status/1133955697998678177/photo/1", "id": 1133955492688957442, "id\_str": partexterrouwouwouwer, euweneum, moarrouwir owosynawowostawa uszawa wykania w siszawa w siszawa w siszawa w si 19355594208584, indicest i lini, 1941, real w 1" "http://bob.imig.com/endicest/Dyspecifik/Ip4K.jpg", "edia.ur/ http:// http://bob.imig.com/edia/ Dyspecifik/Ip4K.jpg", siszest: ("large": ("http://bob.imig.com/edia/bab/sectuari: ("http://bob.imig.com/edia/ w":edda), thumas' ("htt: 194, resize": "fit", "http://bab/sectuari: ("http://bab/sectuari: ("http://bab/se http://bab/sectuari: ("http://bab/sectuari: ("htt "id": 1133959499994978177, "id\_str: 1133959499998078177", "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id="tri", null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id="tri", null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id="tri", null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id="tri", null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id="tri", null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id="tri", null, "i place": null, "possibly\_sensitive": false, "retweet\_count": 0, "retweeted": false, "source": "<a href=\"http://twitter.com/download/iphone\" rel=\"nofollow\">Twitter for Vencoultant Toma Science Libertian at K. Sen Dopp Jorsen J Ayrben J waking data hun," Mettitant' (Vencoultant' Internet) (Sentonian at K. Sentonian J Artis, J Ayrben J Awaing data hun," Mettitant' (Vencoultant' Internet), Sentonian at K. Sentonian J Artis, J Artis and J Artis an ttrat", null, "verified": false)

#### Pretty printed

## Archiving Data + Code

What does code look like?

- twint command line
- open humans Jupyter notebook

< □ > < 同 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

- hydrator desktop app option
- purpletag command line

## Metadata

## COPEIN ICPSR Find Data Share Data openICPSR Repositories + End Data / Who Did US Congress Research in 2017

Who Did US Congress Retweet in 2017

Principal Investigator(s): O Libby Hemphill, University of Michigan; Angela Schopke, University of Michigan; Caroline Hodge, University of Michigan; Chris Bredernitz, University of Michigan

Version: 😡 V1

Version Title: 😡 Initial deposit

Name 🖸	File Type 🖸	Size 🖸	Last Modified 🖸		OOWNLOAD THIS PROJECT		
retweets 2017 complete.csv	text/csv	427.7 MB	01/28/2019 06:27:AM				
Project Citation: Hemphill, Libby, Schooke, Angela, Hodge, Care	line, and Bredernitz, Chris, Wh	o Did US Congress Retweet in	2017. Ann Arbar. Mi: Inter-university Consortium for	Usage Metrics 🚱 Overall Project Me	trics		
Political and Social Research [distributor], 2019 Persistent URL: http://doi.org/10.3886/E1085	9-01-28. https://doi.org/10.3886 803V1	5/E108303V1		217 Views	14 Downloads	0 Publications	
Project Description				Download Detailed	Metrics		
Summary: Of this idease includes the retweets posted on Twitter by accounts associated with members of the US Congress in 2017. The list of accounts associated with members in the second strategies and the second strategies and the second strategies and the second strategies and the second strategies and the second strategies and the second strategies and the second strategies the two levels associated strategies and the second strategies and the second strategies and the second strategies and the second strategies and the secon			Published Versions V1.(2019-01-28)				
Tweets were collected using Twitter's Search Al	PI through the twitter_user_coll	ector Python script ( <u>https://g</u>	thub.com/casmlab/twitter_user_collector).				
Scope of Project				Export Metadata Dublin Core			
Subject Terms: 9 Congress; twitter				DDI 2.5			
Geographic Coverage: 😡 United States				DDI 3.1			

#### Congress's retweets example

libbyh@umich.edu+

▲□▶ ▲圖▶ ▲臣▶ ★臣▶ = 臣 = のへで

## Integrating Search + Analysis



◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 三臣・のへで

## **SOMAR** Questions

- 1. How should SOMAR prioritize its goals? (e.g., replication, preservation, active collection)
- 2. How would users most likely search for data in SOMAR?
- 3. What tools exist already on your campus for facilitating social media data access? (e.g., Crimson Hexagon, Sysomos, Python/R consulting)
- 4. Which ethical considerations should guide our decisions, and how do they lead to different technical outcomes?

< □ > < 同 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <