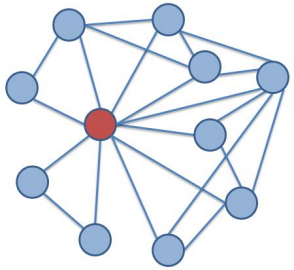


Sharing selves

Developing an ethical framework for curating social media data

Sara Mannheimer
Montana State University

Elizabeth Hull
Dryad Digital Repository

#IDCC2017 | Edinburgh | 21 February 2017


twitter
in
YouTube
digg

Emil OW Kirkegaard @KirkegaardEmil · May 8

The OKCupid paper has now been submitted. This means that the dataset is now public! Enjoy! :) openpsych.net/forum/showthre…

34 47

Ethan Jewett @esjewett · May 11

@KirkegaardEmil This data set is highly re-identifiable. Even includes usernames? Was any work at all done to anonymize it?

3 18

Emil OW Kirkegaard
@KirkegaardEmil

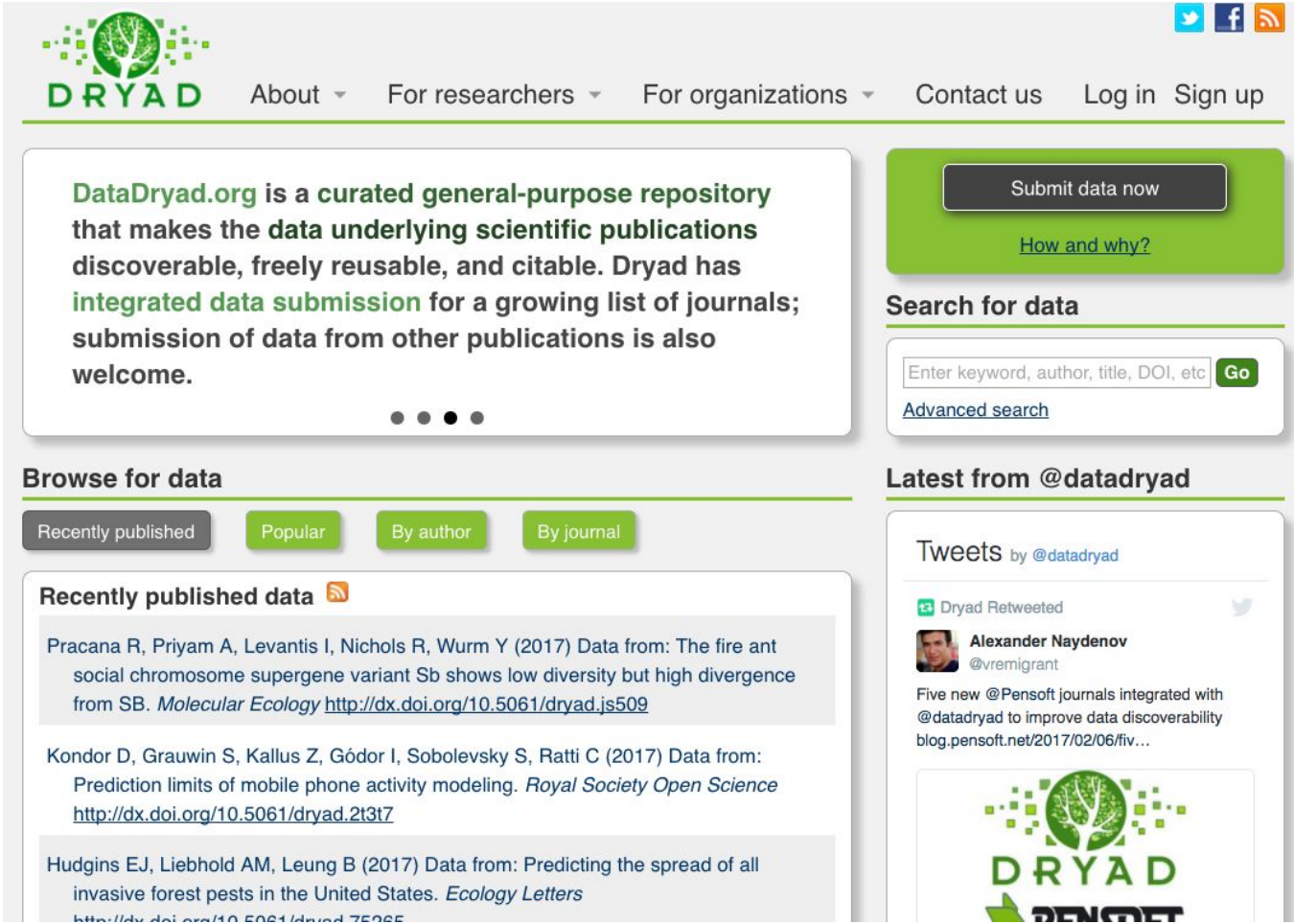
Follow

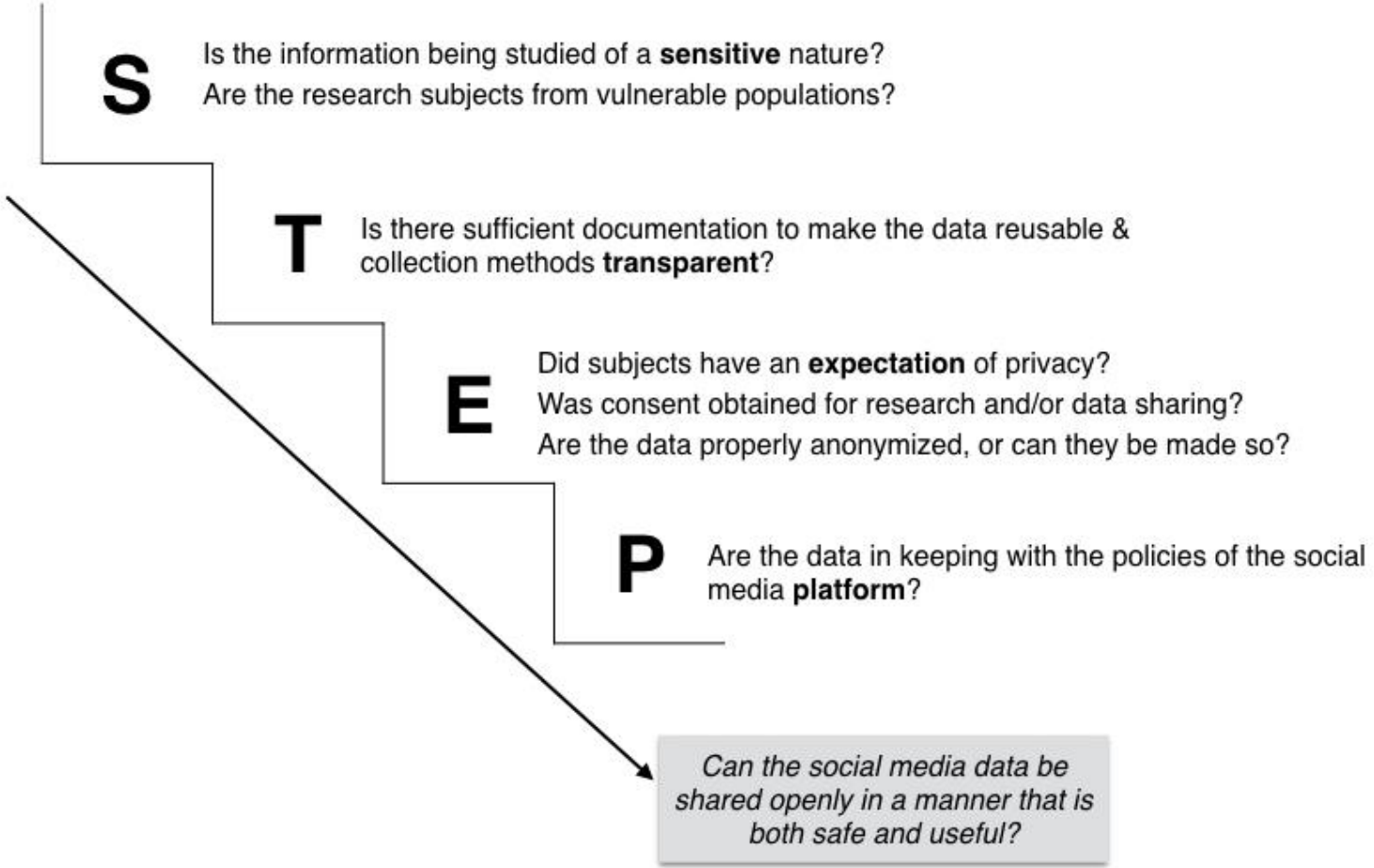
@esjewett No. Data is already public.

LIKES
3

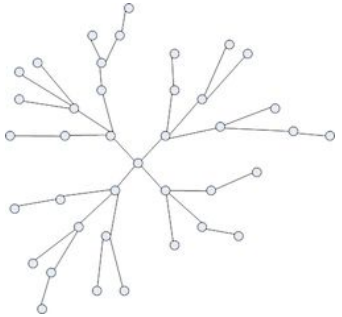
11:30 AM - 11 May 2016

3

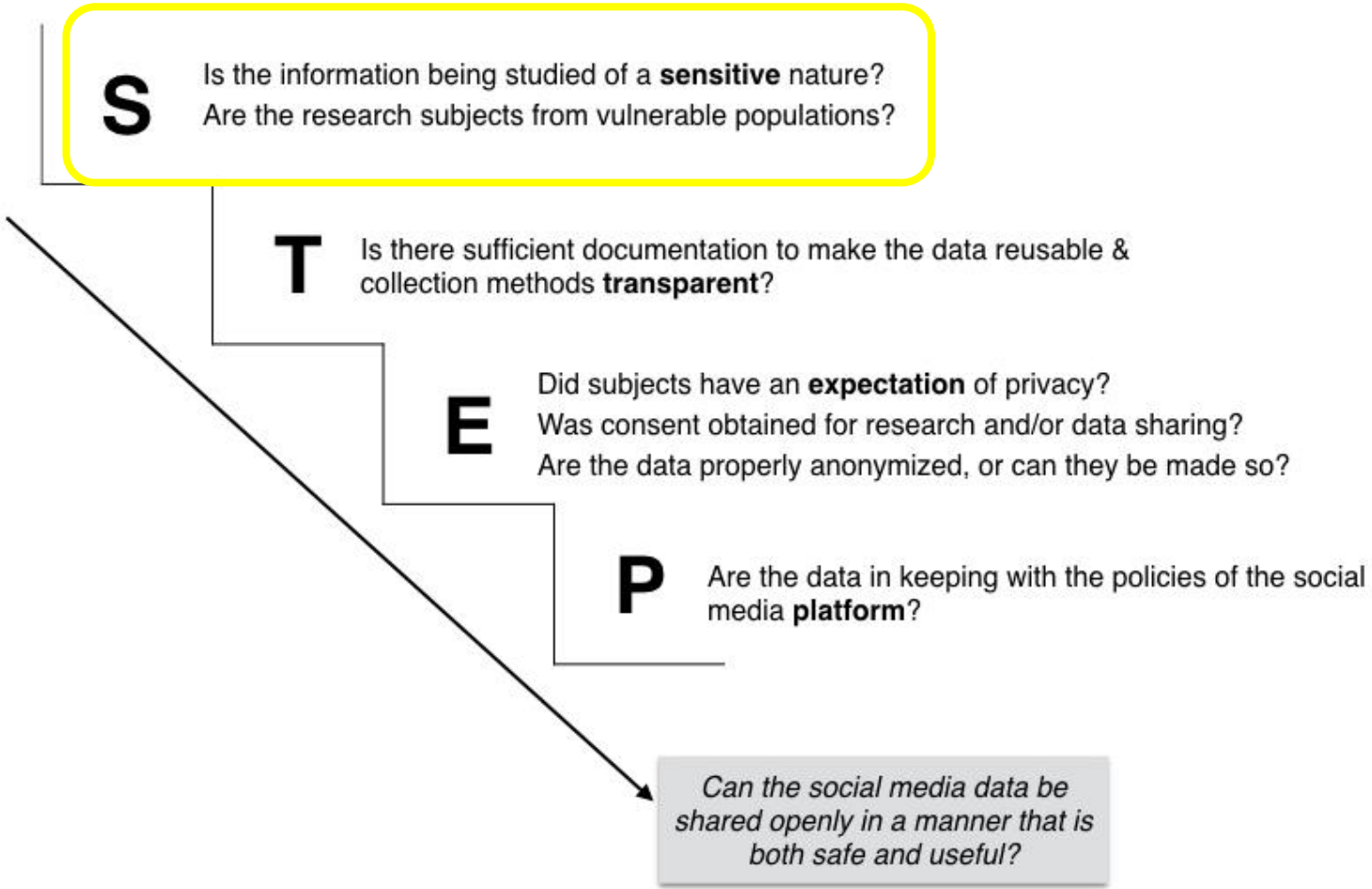

DRYAD
About
For researchers
For organizations
Contact us
Log in
Sign up
DataDryad.org is a curated general-purpose repository that makes the data underlying scientific publications discoverable, freely reusable, and citable. Dryad has integrated data submission for a growing list of journals; submission of data from other publications is also welcome.
Submit data now
How and why?
Search for data
Enter keyword, author, title, DOI, etc
Go
Advanced search
Browse for data
Recently published
Popular
By author
By journal
Recently published data
Pracana R, Priyam A, Levantis I, Nichols R, Wurm Y (2017) Data from: The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. Molecular Ecology http://dx.doi.org/10.5061/dryad.js509
Kondor D, Grauwin S, Kallus Z, Gódor I, Sobolevsky S, Ratti C (2017) Data from: Prediction limits of mobile phone activity modeling. Royal Society Open Science http://dx.doi.org/10.5061/dryad.2t3t7
Hudgins EJ, Liebhold AM, Leung B (2017) Data from: Predicting the spread of all invasive forest pests in the United States. Ecology Letters
Latest from @datadryad
Tweets by @datadryad
Dryad Retweeted
Alexander Naydenov
@vremigrant
Five new @Pensoft journals integrated with @datadryad to improve data discoverability blog.pensoft.net/2017/02/06/fiv…
DRYAD

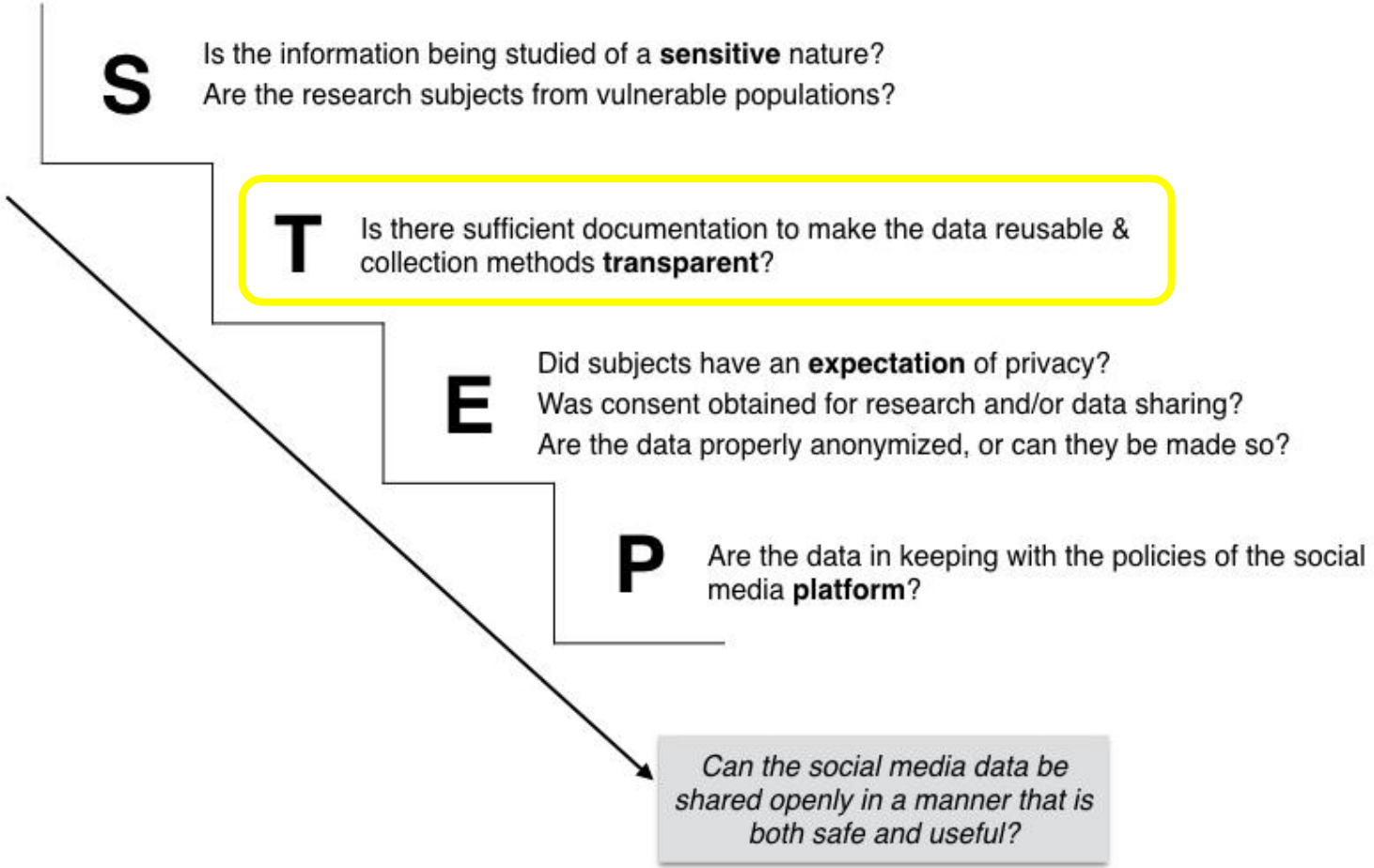

S
Is the information being studied of a **sensitive** nature?
Are the research subjects from vulnerable populations?
T
Is there sufficient documentation to make the data reusable & collection methods **transparent**?
E
Did subjects have an **expectation** of privacy?
Was consent obtained for research and/or data sharing?
Are the data properly anonymized, or can they be made so?
P
Are the data in keeping with the policies of the social media **platform**?
Can the social media data be shared openly in a manner that is both safe and useful?

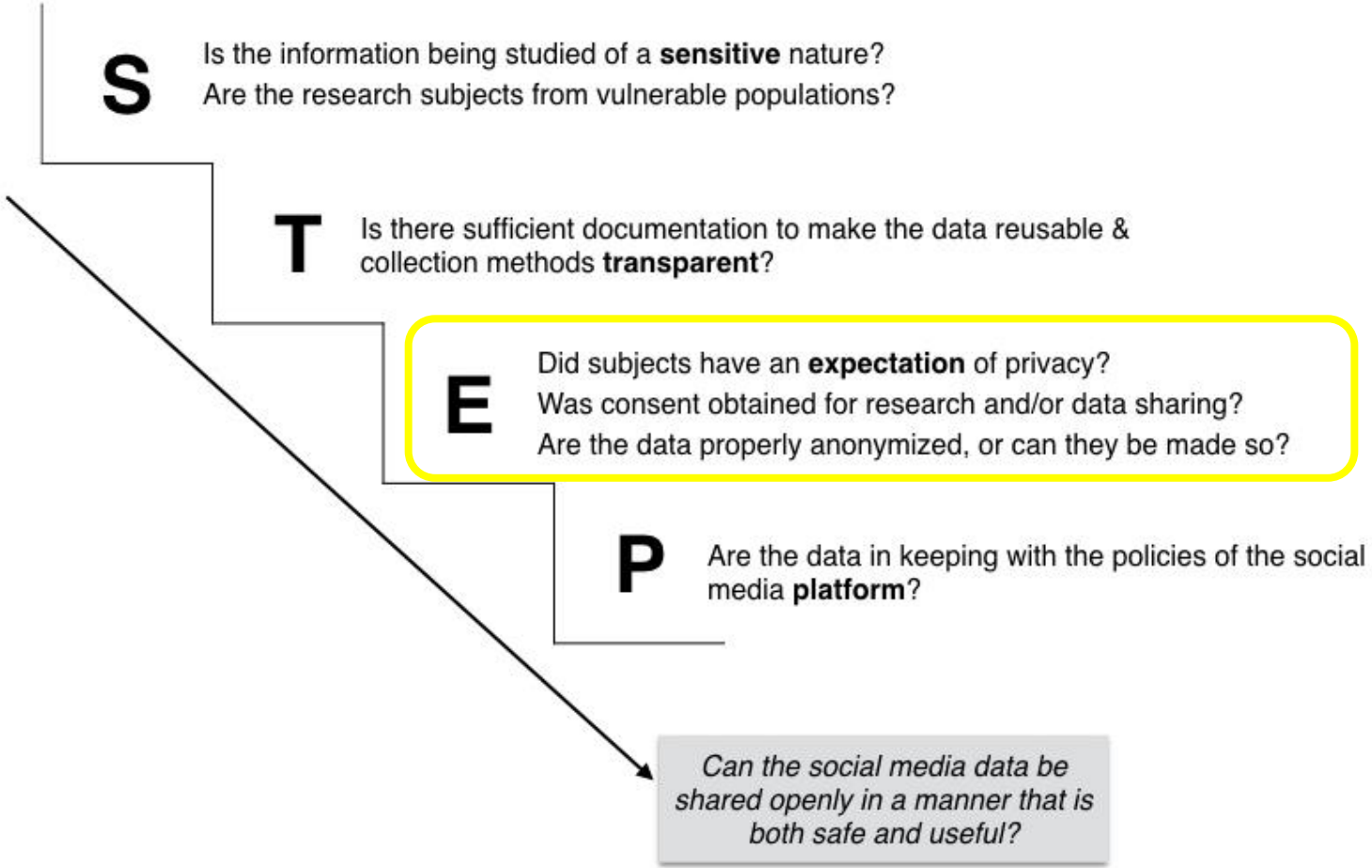
STEP Framework guiding principles

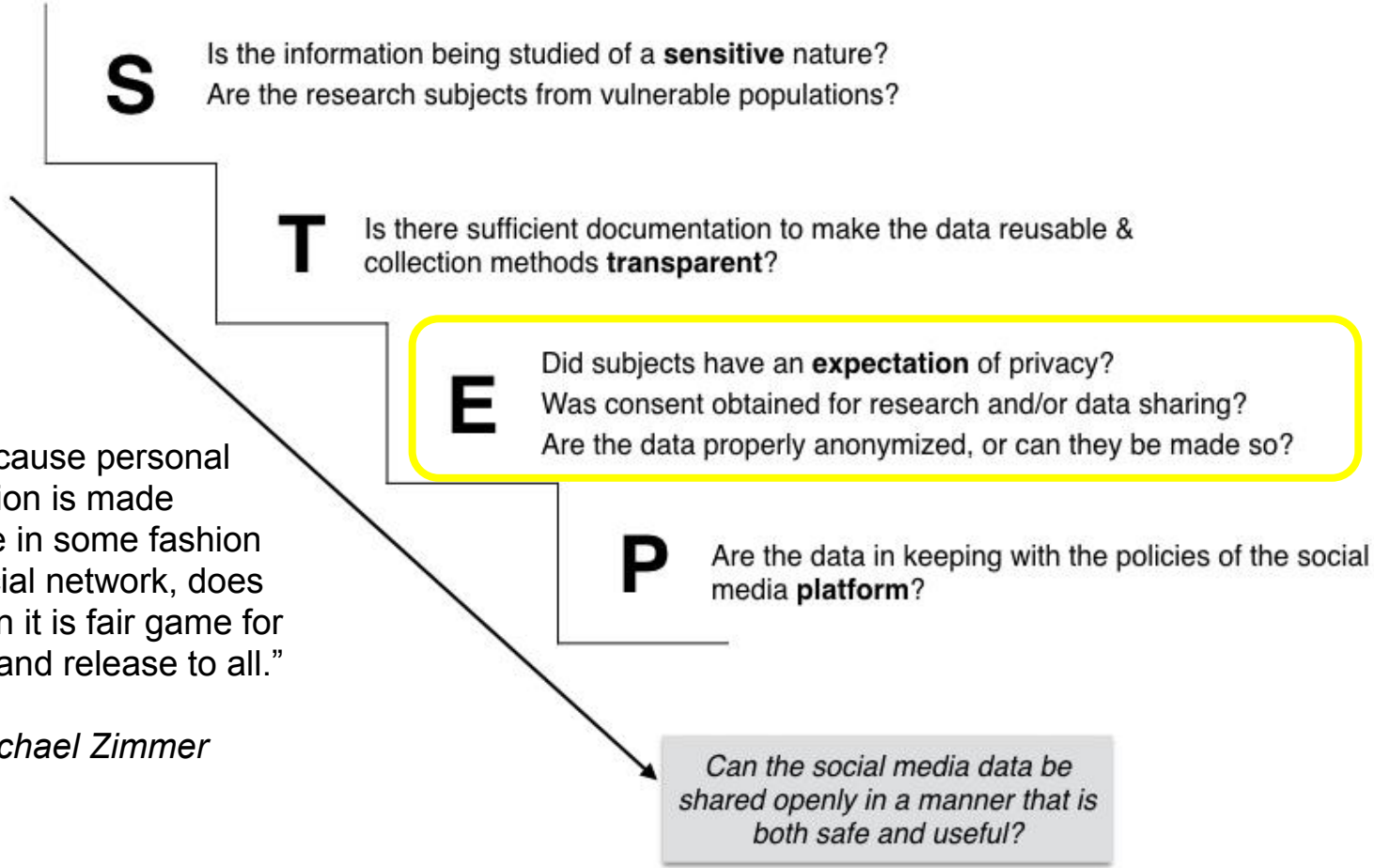


1. **Value analysis**. When sharing social media data, researchers and data curators must measure the benefits of sharing data against the potential risks to human subjects.
2. **Responsibility**. Data curators can help educate researchers about ethical data sharing, but researchers themselves are ultimately responsible for the data they share.
3. **Continual inquiry**. Ethical practice requires ongoing dialogue and examination.


S
Is the information being studied of a **sensitive** nature?
Are the research subjects from vulnerable populations?
T
Is there sufficient documentation to make the data reusable & collection methods **transparent**?
E
Did subjects have an **expectation** of privacy?
Was consent obtained for research and/or data sharing?
Are the data properly anonymized, or can they be made so?
P
Are the data in keeping with the policies of the social media **platform**?
Can the social media data be shared openly in a manner that is both safe and useful?


S
Is the information being studied of a **sensitive** nature?
Are the research subjects from vulnerable populations?
T
Is there sufficient documentation to make the data reusable & collection methods **transparent**?
E
Did subjects have an **expectation** of privacy?
Was consent obtained for research and/or data sharing?
Are the data properly anonymized, or can they be made so?
P
Are the data in keeping with the policies of the social media **platform**?
Can the social media data be shared openly in a manner that is both safe and useful?

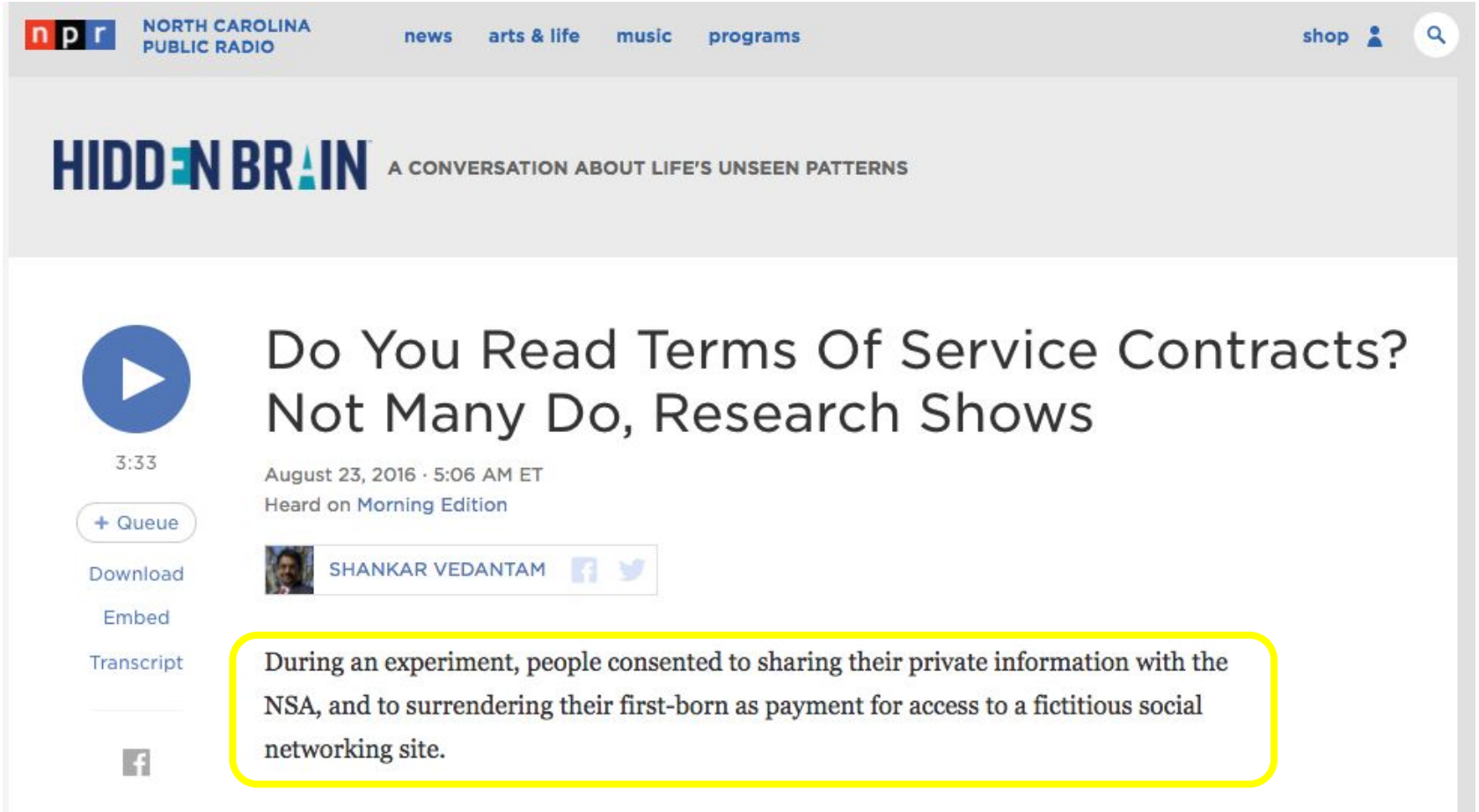

S
Is the information being studied of a **sensitive** nature?
Are the research subjects from vulnerable populations?
T
Is there sufficient documentation to make the data reusable & collection methods **transparent**?
E
Did subjects have an **expectation** of privacy?
Was consent obtained for research and/or data sharing?
Are the data properly anonymized, or can they be made so?
P
Are the data in keeping with the policies of the social media **platform**?
Can the social media data be shared openly in a manner that is both safe and useful?



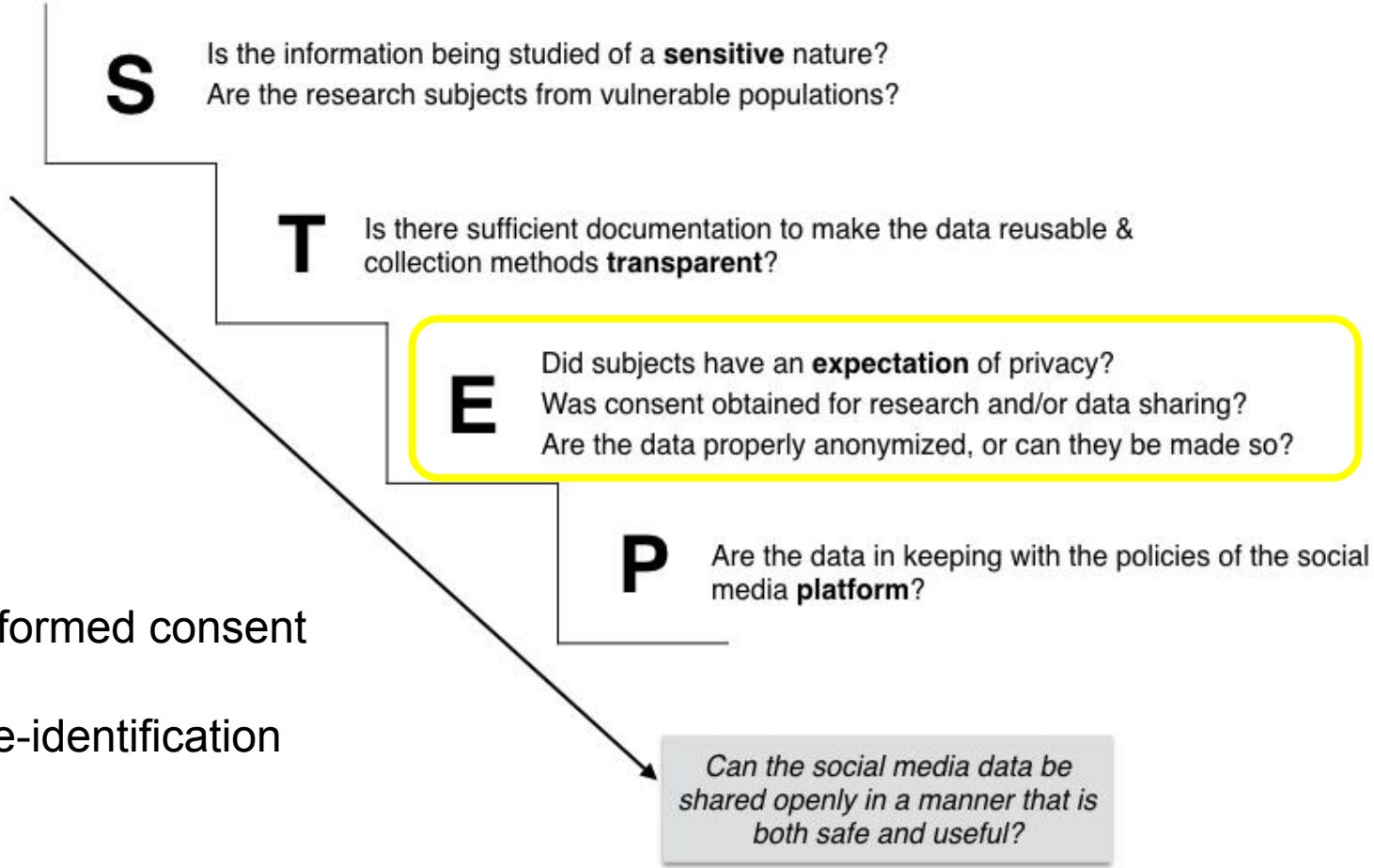
"Just because personal information is made available in some fashion on a social network, does not mean it is fair game for capture and release to all."

- *Michael Zimmer*

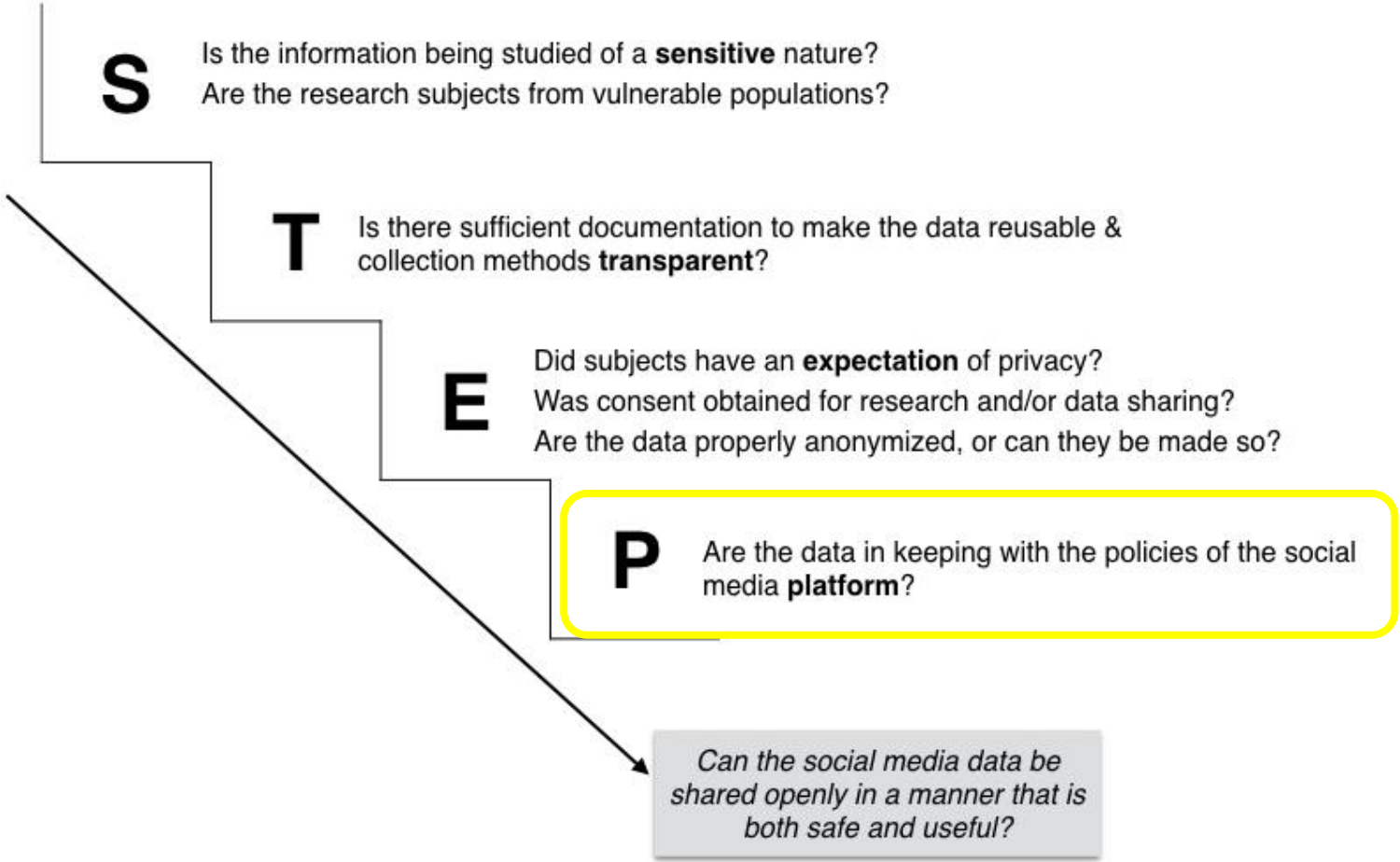
Zimmer M (2010) "But the data is already public": On the ethics of research in Facebook. http://doi.org/10.1007/s10676-010-9227-5

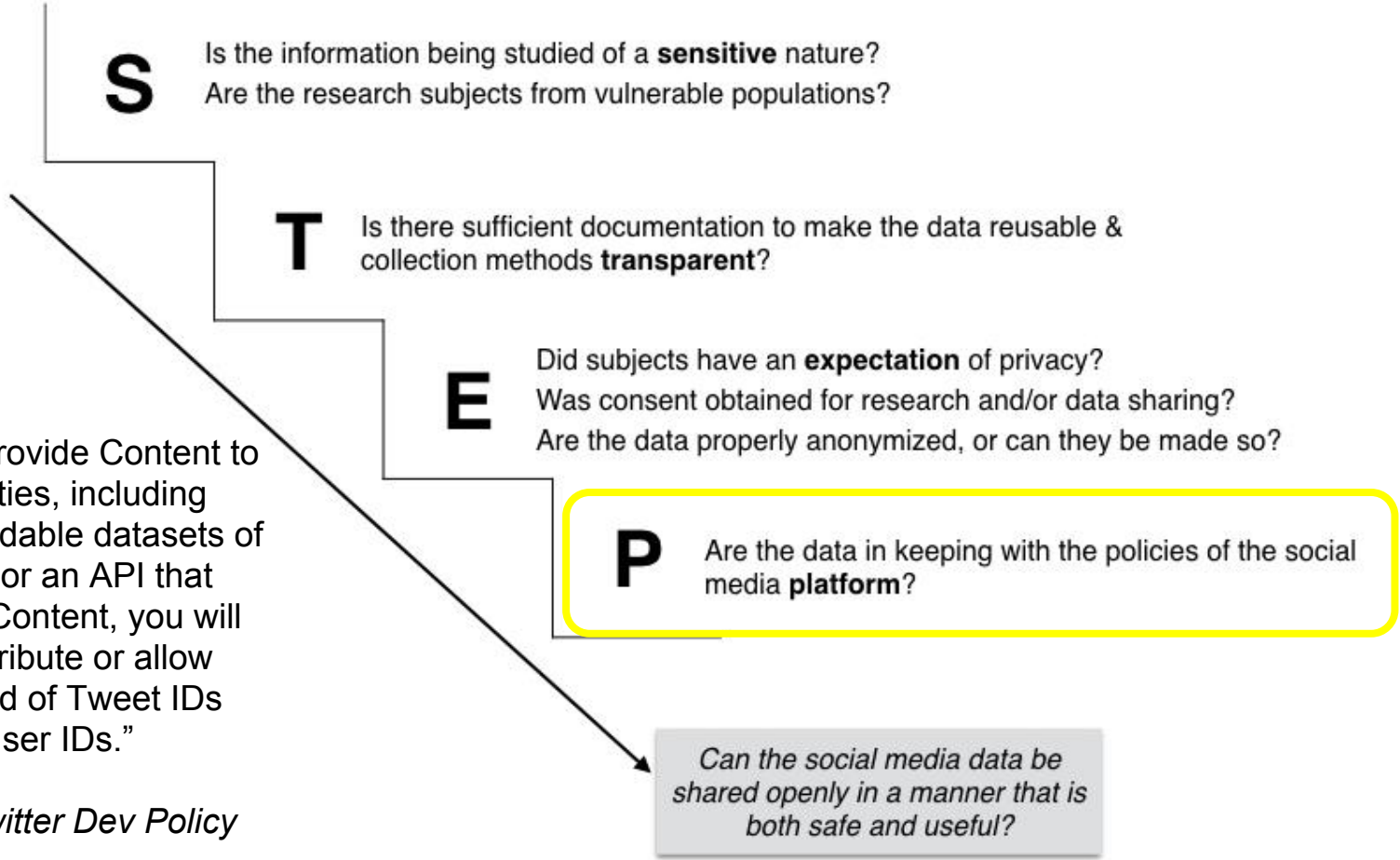


http://www.npr.org/2016/08/23/491024846/do-you-read-terms-of-service-contracts-not-many-do-research-shows



- Informed consent
- De-identification

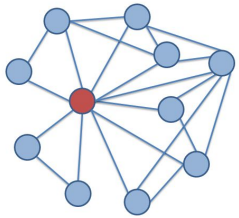

S
Is the information being studied of a **sensitive** nature?
Are the research subjects from vulnerable populations?
T
Is there sufficient documentation to make the data reusable & collection methods **transparent**?
E
Did subjects have an **expectation** of privacy?
Was consent obtained for research and/or data sharing?
Are the data properly anonymized, or can they be made so?
P
Are the data in keeping with the policies of the social media **platform**?
Can the social media data be shared openly in a manner that is both safe and useful?



"If you provide Content to third parties, including downloadable datasets of Content or an API that returns Content, you will only distribute or allow download of Tweet IDs and/or User IDs."

- *Twitter Dev Policy*

https://dev.twitter.com/overview/terms/agreement-and-policy



Case study 1: the #occupy case

- Authors used Twitter hashtag data to study the evolution of political discussion during and after the Occupy Wall Street movement.
- The associated Dryad data package includes one .csv file containing three variables: “user,” “hashtag,” and “time.”

Gargiulo F, Bindi J, Apolloni A (2015b) The topology of a discussion: the #occupy case. PLOS ONE 10(9): e0137191. http://dx.doi.org/10.1371/journal.pone.0137191

Gargiulo F, Bindi J, Apolloni A (2015a) Data from: The topology of a discussion: the #occupy case. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.q1h04

Case study 2: in the mood

- Study of the relationship between UK Twitter users' "sentiment levels" and the network structure created by @-mentions.
- Researchers selected 18 "communities" to monitor and used these to formulate a model for "reproducing measures of emotive response."
- The data package contains several dynamic mention networks split over 6 tables; variables include an anonymised tweet ID, anonymised user IDs, and timestamps of tweets.

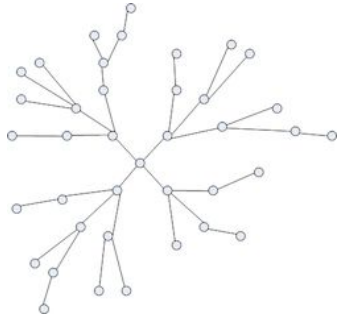
Charlton N, Singleton C, Greetham DV (2016b) In the mood: the dynamics of collective sentiments on Twitter. Royal Society Open Science 3(6): 160162. http://dx.doi.org/10.1098/rsos.160162

Charlton N, Singleton C, Greetham DV (2016a) Data from: In the mood: the dynamics of collective sentiments on Twitter. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.5302r

AREA


berghaus

Taking STEP further



- Much like social media research, the STEP framework should evolve over time
- More case studies and testing with a variety of repositories and platforms
- Expanded application to big data research, social science data journalism

Thank you

Sara Mannheimer

Data Management Librarian
Montana State University
@saramannheimer

Elizabeth Hull

Operations Manager
Dryad Digital Repository
@datadryad