DMDS: Social Media Research Data Ethics and Management

Andrea Zeffiro: zeffiroa@mcmaster.ca
Jay Brodeur: brodeujj@mcmaster.ca
Sherman Centre for Digital Scholarship
05-April, 2018

Outline

- Case studies & discussion [30 mins]
- Ethical considerations [15 mins]

**********Break********** [10 mins]

- Managing & sharing SM materials [20 mins]
- Evaluating frameworks & wrap-up [25 mins]

Case studies

Some considerations

Is the data private?

Can the subject matter be considered sensitive?

Are any of the subjects vulnerable?

Is consent necessary? Is it given? How to obtain it?

How (if at all) should source information be presented in publications?

How (if at all) should the data be shared?

Should the researcher identify themselves?

Is the research exploitative?

Is the data representative?

Is there a need to account for bots, trolls, and spam?

Is ethics approval necessary?

Are there other ethical and methodological considerations?

“

A researcher wishes to conduct a content analysis of tweets related to the 2016 US Presidential Election, to explore how Trump supporters argued for their candidate on Twitter.

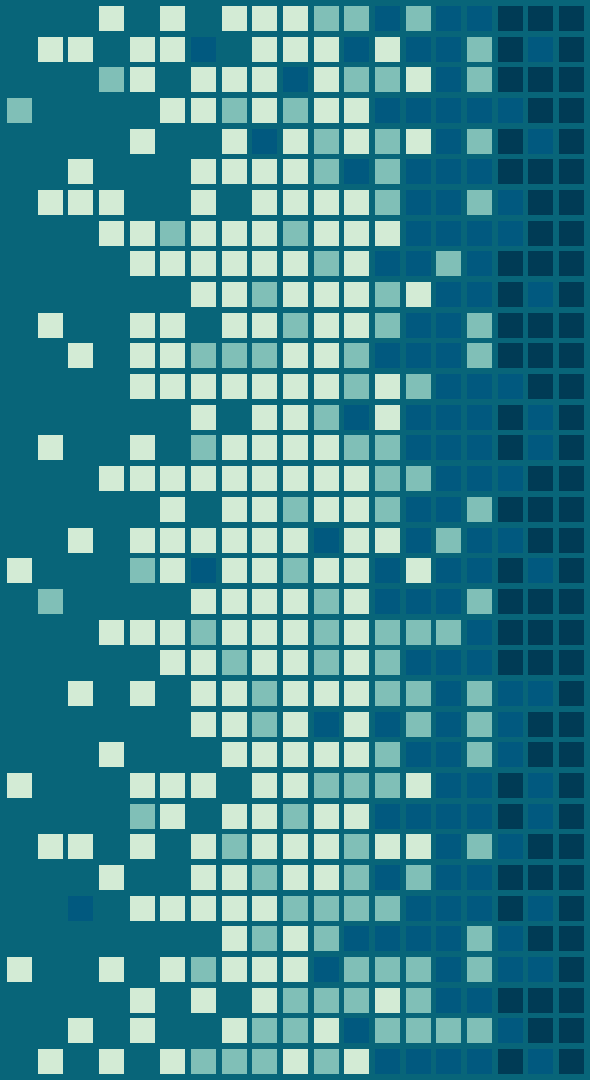
They have paid a third-party service to provide data related to tweets using the hashtags #DonaldTrump, #TrumpTrain, #VoteTrump2016, #AlwaysTrump, #MakeAmericaGreatAgain, and #Trump2016 that span the period leading up to and shortly after the election.

Scenario 1

“

A researcher wishes to study support mechanisms and discourse amongst members of a discussion forum which deals with mental health issues such as depression and feelings of suicide.

The forum is closed and password protected, and registration must be approved by a gatekeeper (a site admin).



Scenario 2

“

Researchers studying how Facebook is used by people in Puerto Rico in the aftermath of Hurricane Maria… planning to conduct an in-depth qualitative analysis of public/private Facebook pages used by local people to communicate and organise in the aftermath.

There are a wide range of topics being discussed on the boards including people searching for lost family and friends…

The researchers want to join the private groups, and then observe how different types of public and private Facebook pages are being used by people as they respond to the disaster.

Scenario 3

“

A researcher wishes to use Tinder to study public interactions on social dating platforms. Although the posts being studied are public (rather than through private messaging), she needs to sign up to Tinder to view them.

By signing up, she has to fill in a registration form including questions such as “I am a woman looking for a man/woman” etc. It is therefore reasonable to think that users of the platform expect that other people viewing their profile might be doing so for similar (dating) reasons. The users of the platform are aware that there is a very large number of people using the platform and potentially able to access their profile.

Scenario 4

“

A researcher wishes to perform a discourse analysis of interactions between environmental activists, organizations (such as greenpeace), and corporations on Twitter through close reading of tweets for selected (less than 10) individuals and a number of prominent public groups.

They wish to share excerpts of the interactions in an upcoming publication.

Scenario 5

“

Working in the days after the Charlie Hebdo attacks in Paris, researchers aim to conduct a network and sentiment analysis of Twitter users using the hashtag ***#jesuischarlie***. They plan to use an online commercial tool to collect tweets (legal + aligned with T&C) that will be fully identifiable.

...plan to create network visualisations showing how tweets became popular through retweeting practices. They also want to visualise how sentiment about the events emerged over time amongst different networks of Twitter users.

They want to make an interactive online visualisation in which users will be able to zoom in on particular areas of the network to view specific tweets and their submitting users.

Scenario 6

Ethical & methodological considerations

Social Media

Websites and applications that enable users to create and share content or to participate in social networking.

Sharing information, ideas, personal messages and other content such as images and videos.

Types of Platforms

- Networking, information sharing, content curation
  - (i.e. Facebook, Twitter, Youtube, LinkedIn, Reddit)
- Online forums for specific communities
  - (i.e. PatientsLikeMe, Mumsnet, BaristaExchange)
- Private collaborative tools
  - (i.e. Trello, Yammer, Slack)
- Crowdsourcing platforms
  - (i.e. GoFundMe, Kickstarter, etc.)

(Taylor and Pagliari 2017)

Enable the conduct of research

Informal and formal modes of scholarly exploration.

- Gathering opinions
- Recruiting Participants
- Fostering stakeholder involvement

(Taylor and Pagliari 2017)

As a source of data for research

‘Secondary uses’ include studies seeking to profile or understand users’ behaviours, demographics, interactions and networks, or to assess their responses or sentiments towards particular topics, products or policies.

(Taylor and Pagliari 2017)

Benefits of social media research

- Reach larger numbers of participants
- Reduce cost
- Analyse trends and associations within large corpuses of data
- Interaction across extended time periods
- Less prone to bias than approaches involving direct contact between researchers and participants
- Involvement of citizens in research process
- Creating new channels for research dissemination

(Taylor and Pagliari 2017)

Methodological Considerations

- Representativeness
- Inequalities in access
- Heterogeneous data
- Non-traditional sampling approaches
- Social media service provider

Ethical Considerations

The complexity of interactions between individuals, groups, and technical systems present a number of challenges for scholars seeking to use social media data in research.

Recommendation: Ethical considerations guide the research design and methodological considerations.

Contextual

It’s impossible to adopt a ‘one size fits model’:

- Every social media context is unique
- Ethical considerations are grounded in the specifics of the social media community, the methodology and research questions
- Ethical decision making is a deliberative and iterative process

Common (Ethical) Challenges

I. Public vs Private

II. Informed Consent

III. Anonymity

IV. Managing and Sharing Data

i. Public vs Private

Terms and Conditions are written in legal discourse and contain clauses on how one's data is managed and used by a platform and accessed by third parties, including researchers.

Ethical considerations about access to and use of data cannot be ignored simply because a service provider deems data as 'public'.

Reasonable Expectations of Privacy

The perception of privacy very much depends on a particular platform's or group's protocols and privacy boundaries, audience and aims, which vary greatly from platform to platform, individual to individual, group to group, hashtag to hashtag.

Does the social media participant reasonably expect to be observed by strangers? What about researchers? Does the participant consent to be part of a research study/project? Will publicizing units of data identify a social media participant?

Informed Consent

- To what extent are we ethically bound to seek informed consent?
- Social media participants are not always aware of their participation as research subjects/subjects of research
- It’s difficult to acquire or expect informed consent with large data sets
- Necessary when reproducing individual units of data for publication and/or public presentations

Anonymity

- Anonymising social media data is still a complex process
- Researchers need to consider the data
- Different issues arise for different data
  - Text-based units of data
  - Interoperability of datasets

Risk of Harm

Republishing quotes verbatim and/or using screen grabs can expose the identity and profile of the social media participant.

- Paraphrase
- Seek informed consent for research output
- Consider more traditional approaches

Managing & sharing social media research materials

Forms of dissemination & sharing

- Disseminating materials through publications, presentations, blog posts, visualizations
  - Text, images, video, audio, etc.
  - Aggregated results
  - Small units (excerpts)

- Sharing research datasets with collaborators / reviewers / research community / public

If, how, and where to share

The ‘if’, ‘how’, and ‘where’ to share depend on:

- The data (privacy, sensitivity, specificity/granularity)
- The subjects (vulnerability, expectation of privacy)
- The SM platform’s terms of use and conditions
- The format of dissemination (text vs. image vs, video)
- Institutional, disciplinary, and funding body guidelines

Considerations for dissemination

- Read thoroughly (and revisit!) the terms and conditions for both ***users*** and ***data users***
  - Who maintains (copy)rights to the information?
  - Can direct excerpts be published?
- Seek consent where required, appropriate, and possible
- Protect participants' identity
  - Anonymize by removing/treating direct (handles, usernames, emails) & indirect (gender, location) identifiers
  - Fictionalize aspects of the research
  - Paraphrase materials

Considerations for sharing datasets

Why share social media research datasets[2]?

1. To support research transparency
   - i.e. reproducibility and verification
2. To enable broad access to data
3. To benefit research efficiency through reuse
4. To satisfy publisher / funder requirements

[2]Weller and Kinder-Kurlanda p. 166

Considerations for sharing datasets

- Read the terms and conditions!
- Anonymize datasets by removing handles, usernames, and other direct identifiers
- Consider (and minimize) potential for re-identification through indirect identifiers
- Control access to datasets
  - Restrict access by accounts, groups, domain
  - Require potential reusers to request data or notify author

Draft Tri-Agency RDM Policy

- Tri-Agency draft data management policy expected April, 2018
- 6-month consultation period; feedback will inform policy

Proposed policy — 3 possible requirements:

- Institutions: Institutional Strategy
- Researchers: Data Management Plans
- Researchers: Data Deposit ***where appropriate***

- Phased and incremental implementation

Resources for ethical sharing of social media data

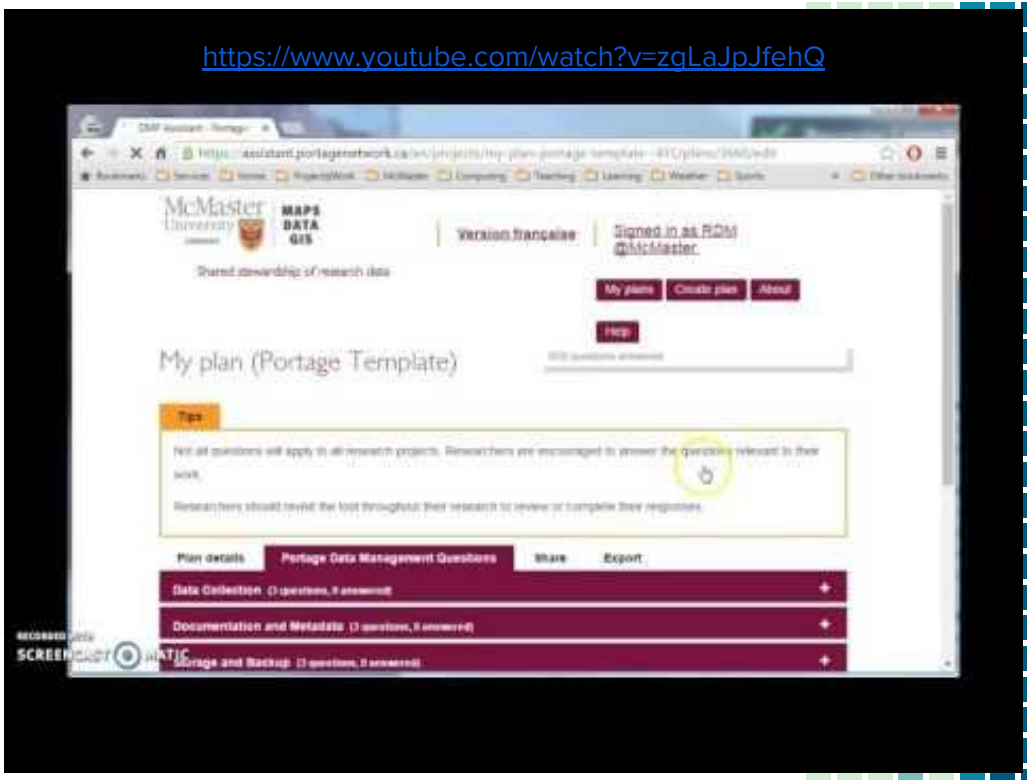
- Planning
- Storing
- Sharing

Portage DMP Assistant

DMP ASSISTANT
DATA MANAGEMENT PLANS
PLANS DE GESTION DES DONNÉES
PGD

- A web-based, bilingual data management planning tool.
- Available to all researchers in Canada.
- A guide for best practices in data stewardship.
- Exportable data management plans.

https://www.youtube.com/watch?v=zgLaJpJfehQ



https://assistant.portagenetwork.ca/

Considerations for managing data

What types of data (and how much) will you collect?

How will you organize, secure, and backup your data?

Are there ethical or commercial conditions?

- Should your data be encrypted?

How will you describe your data so that others understand it?

How will you control access to your data?

How will you manage data versions?

	RHPCS - Backup Services	RHPCS - Hosted Server Packages	MacDrive	Microsoft OneDrive / Teams
Storage Quota	1 TB; more available for fee	1 TB; more available for fee	300 GB per account	1 TB per account; up to 5 TB by request
Rates / cost	$500 / yr + one time set up fee ($125 / machine) Additional space: $300 / TB Restore services: $125 / hour	$500 - $4000 / yr Setup fee: $500 - $1000 Additional space: $450 / TB	No cost to users	No cost to users
Backups / versioning	Nightly, 14-day rotating cycle; Restore services through RHPCS	Nightly, 14-day rotating cycle; Restore services through RHPCS Nextcloud sync service available.	Ongoing real-time sync 4-month version history Full Library restore through UTS	Ongoing real-time sync Unlimited version history (?)
Who can use this service?	Any subscribing users or research group	Any subscribing users or research group	McMaster Faculty and Staff Graduate students can obtain zero-quota accounts	All McMaster faculty, staff and students
Server location	A.B. Bourns building	A.B. Bourns building	Replicated clusters in Gilmour Hall and JHE	**OneDrive**: Canadian servers **Teams**: Soon in Canadian servers only
Other notes			Supports encrypted libraries, file and directory sharing, Desktop client, web interface	Supports file and directory sharing, Desktop client, web interface
More info	rhpcs.mcmaster.ca/current-rates	rhpcs.mcmaster.ca/current-rates	macdrive.mcmaster.ca/ Documentation: https://goo.gl/AvRGwx	portal.office.com/ Documentation: mcmaster.ca/uts/licensing

Access the matrix: goo.gl/45iy38

Considerations for sharing datasets

How will your data products be stored in the long-term?

- How to ensure that it remains *integral* and *secure*?
- Who will assume long-term *responsibility* for your data?

How will others access your data products?

- What data (if any) can/should be shared? Who should have access?
- How will you manage legal, commercial & ethical constraints?

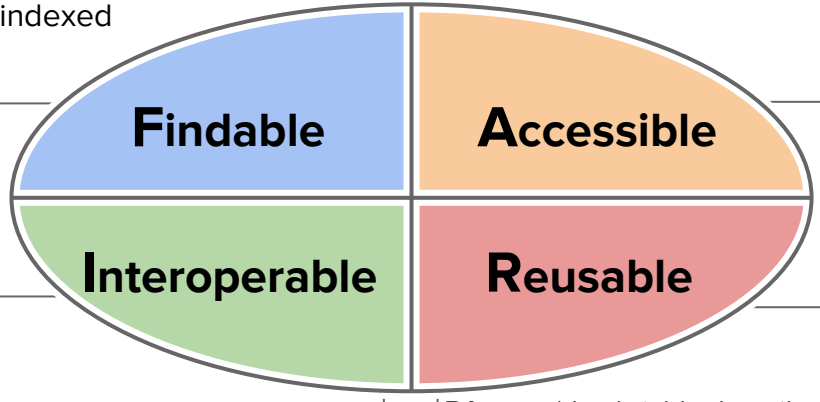
How to maximize credit for sharing your data?

- In which repository should you deposit your data?
- How to ensure that your data is FAIR (*findable*, *accessible*, *interoperable* and *reusable*)?

The FAIR Guiding Principles

F1: (meta)data have a globally unique and eternally persistent identifier
F2: data are described with rich metadata
F3: metadata clearly and explicitly includes the ID of the data it defines
F4: (meta)data are registered and indexed in a searchable resource

A1: (meta)data retrievable by their ID using a standardized protocol
A1.1: protocol is open, free and universally implementable
A1.2: protocol allows for AuthT/ AuthZ where needed
A2: metadata is always accessible



I1: (meta)data use a formal,accessible, shared, broadly applicable language for knowledge rep.
I2: (meta)data use vocabularies that follow FAIR principles
I3: (meta)data include qualified references to other (meta)data

R1: meta(data) richly described with accurate and relevant attributes
R2: (meta)data released with a clear and accessible data usage license
R3: (meta)data associated with detailed provenance
R4: (meta)data meet domain-relevant community standards

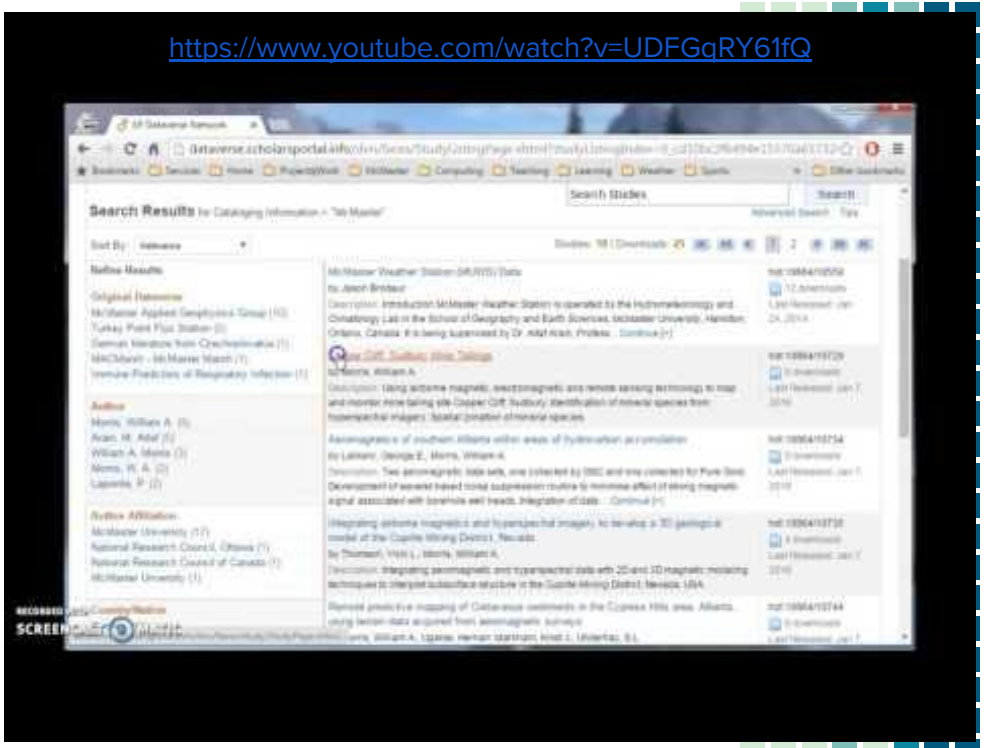
Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3. *doi:10.1038/sdata.2016.18*

Scholars Portal Dataverse

Scholars Portal
A Service of the Ontario Council of University Libraries

- A data repository for researchers at Ontario's universities -- **free** and **open** for all researchers in Canada
- An online platform to share, preserve, cite, explore and analyze research data.
- Allows researchers to control how they share their data.
- Supports data DOI registration through Datacite Canada.

https://www.youtube.com/watch?v=UDFGqRY61fQ



http://dataverse.scholarsportal.info

Evaluating frameworks for ethical use of social media data

Frameworks:

Townsend & Wallace (2016)

Williams et al. (2017)

Case studies: Revisited

- Revisit your case studies & re-evaluate
- Use the provided frameworks, where helpful

Follow-up discussion

- What has become clearer? What has not?
- Are the frameworks helpful?
  - Where are they lacking?
- Lingering questions?

Sources cited

Conway, Mike. "Ethical issues in using Twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature." Journal of medical Internet research 16.12 (2014).

Moreno, Megan A., et al. "Ethics of social media research: common concerns and practical considerations." Cyberpsychology, Behavior, and Social Networking 16.9 (2013): 708-713.

Nissenbaum, Helen. "A contextual approach to privacy online." Daedalus 140.4 (2011): 32-48.

Office of the Information and Privacy Commissioner of Ontario. "Big Data Guidelines". Online: https://www.ipc.on.ca/wp-content/uploads/2017/05/bigdata-guidelines.pdf

Shilton, Katie. "Emerging Ethics Norms in Social Media Research." Online: https://bigdata.fpf.org/wp-content/uploads/2015/12/Shilton-Emerging-Ethics-Norms-in-Social-Media-Research1.pdf

Taylor, Joanna, and Claudia Pagliari. "Mining social media data: How are research sponsors and researchers addressing the ethical challenges?." Research Ethics (2017): 1747016117738559.

Townsend, Leanne, and Claire Wallace. "Social media research: A guide to ethics." University of Aberdeen (2016). Online: https://www.gla.ac.uk/media/media_487729_en.pdf

Williams, Matthew L., Pete Burnap, and Luke Sloan. "Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and algorithmic estimation." Sociology 51.6 (2017): 1149-1168.

Unwin, Lindsay, and Kenny, Anita. "The Ethics of Internet-based and Social Media Research: Report of a Research Ethics Workshop held on Thursday 14 July 2016". Online: https://www.sheffield.ac.uk/polopoly_fs/1.644904!/file/Report_Ethics_of_Social_Media_Research_Jul16.pdf

Thank you

Andrea Zeffiro: zeffiroa@mcmaster.ca

Jay Brodeur: brodeujj@mcmaster.ca