Gathering research data online: practical and ethical considerations

UKRIO Annual Conference 2021

Online, 20th May 2021

Claire Hewson,

Department of Psychology,
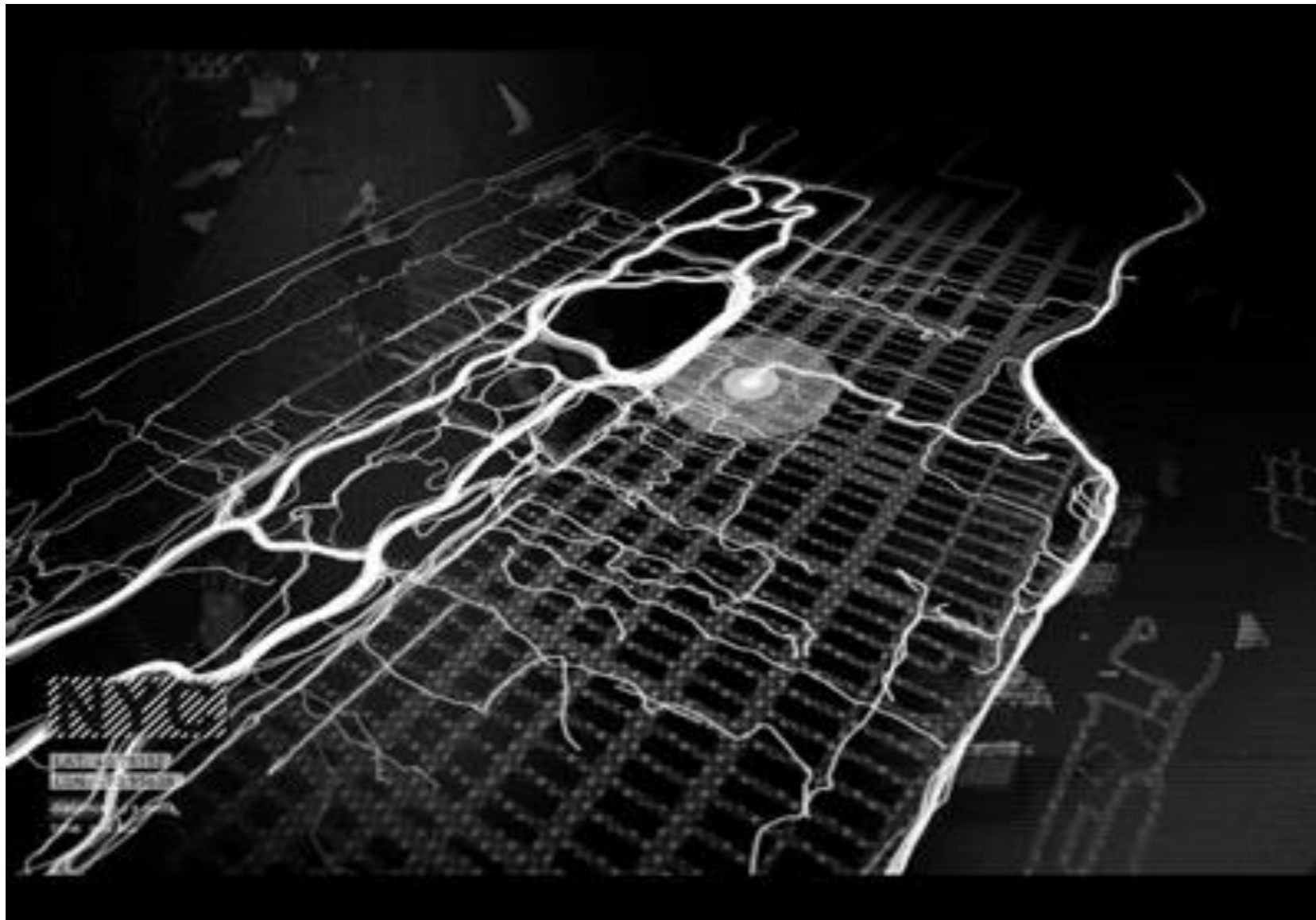
The Open University

claire.hewson@open.ac.uk

Gathering research data remotely (for primary research)

- **Obtrusive methods (surveys/questionnaires, experiments, interviews, focus groups, observation).**

- **Unobtrusive methods (observation, data mining, web crawling, 'big data').**

The Visualisation of Complex Networks

Nike+ involves the placement of a sensor underneath the footbed of your Nike running shoe in order to collect data about where you've run, how long it took and where you can improve over time.

Visualisation of a year's worth of runs uploaded to the Nike+ website. It plays back runs throughout three cities: New York, London and Tokyo. The runs showed tens of thousands of peoples' runs animating the city and bringing it to life. The software visualizes and follows individual runs, as well as showing the collective energy of all the runners, defining the city by the constantly changing paths of the people running in it.

The Visualisation of Complex Networks

http://www.visualcomplexity.com/vc/

Running routes in major cities

http://www.citylab.com/tech/2014/02/mapping-where-people-run/8313/

Thorstad & Wolff (2019)

'Predicting future mental illness from social media: A big data approach' (BRM)

- words posted on clinical 'subreddits' (Reddit) could be used to distinguish several types of mental illness;

- similarly, words posted on non-clinical subreddits could also be used to distinguish several categories of mental illness;

- words derived from non-clinical subreddits predicted posts to clinical subreddits (implying everyday language contains signal about the likelihood of future mental illness).

Practicalities of IMR

Key feature of IMR: involves acquisition of data from or about individuals in absence of face-to-face co-presence

COUPLED WITH greater scope for carrying out quite complex interactive procedures.

Leads to issues of RELIABILITY, VALIDITY and ETHICS due to reduced levels of researcher control over, and knowledge of, participants, participant behaviours and study procedures.

Novel feature of IMR

Opportunities for accessing traces of peoples' everyday interactions and behaviours on an unprecedented scale, in ways not previously possible with traditional offline methods

Leads to issues of PRIVACY, CONSENT and CONFIDENTIALITY (ethical challenges).

Issues of validity/reliability in IMR

- Historically a major concern (mid-90s).

- Reliability issues – largely resolved for many methods (software/technology developments; robust, well-tested procedures).

- Validity concerns – largely alleviated (various validation studies now exist across a range of methods).

- Sampling bias concerns – largely resolved (ubiquity of IUP, some limitations but many opportunities).

IMR design guidelines and useful resources

Selected texts

- Callegaro, Manfreda & Vehovar (2016). Web Survey Methodology. Sage.

- Fielding, Lee & Blank (Eds.) (2017). The Sage Handbook of Online Research Methods (2nd Ed.) Sage.

- Hewson, Vogel & Laurent (2016). Internet Research Methods (2nd Ed.). Sage.

- Roberts, Snee, Hine, Morey & Watson (Eds.) (2017). Digital Methods for Social Science: An Interdisciplinary Guide to Research Innovation. Palgrave.

- Woo, Tay & Proctor (2020). Big Data in Psychological Research. American Psychological Association.

AoIR: http://aoir.org/

WebSM: http://www.websm.org/

Selected Software Resources

- SurveyMonkey: surveymonkey.com
- Qualtrics: qualtrics.com
- JISC Online surveys: onlinesurveys.ac.uk
- WebSM.org provides a useful, searchable database of survey software packages

- WEXTOR: wextor.org/wextor/en
- Gorilla: gorilla.sc

- LogAnalyzer: www.sclog.eu
- Google Analytics: www.google.co.uk/analytics
- Mozdeh: http://mozdeh.wlv.ac.uk/
- Ncapture (part of N Vivo)

- Zoom, Skype, Messenger, WhatsApp, etc.

Ethics in Online Research

Some ethics issues require special consideration in an Internet Mediated Research (IMR) context, the complexities of which are not always obvious to researchers or participants.

These issues require extra care, and consultation of relevant ethics guidelines, to ensure minimising potential risks.

Existing Ethics in IMR Guidelines

- AAAS (Frankel & Siang, 1999)
- AoIR (IRE 3.0, 2019)
- APA (Kraut et al, 2002)
- BPS (2017, forthcoming 2021)

- Internet Research Ethics (IRE) entry in the Stanford Encyclopedia of Philosophy, 2021 (plato.stanford.edu)

- International Journal of Internet Research Ethics

Key ethics issues to pay attention to in IMR

- Data security
- Data traceability
- Anonymity/confidentiality
- Re-identifiability/de-anonymisation
- Blurred public/private distinctions online
- Ensuring robust consent, withdrawal and debrief procedures comparable to offline methods

Data security and traceability

Researchers should carefully consider where and how any research data will be collected and stored online. Considerations to pay particular attention to include:

- Use of third party services/servers – are these within or outside the UK/EU?

- Are adequate encryption and/or password protection methods in place?

- Levels of risk of tracing published data back to original sources/context, e.g. verbatim quotes from online discussion groups

Anonymity and re-identifiability

Considerations to pay particular attention to include:

- Are any potentially identifiable data being collected, e.g. IP addresses?

- Could anonymous/anonymised data potentially become identifiable, e.g. if combined with other data sources

- Use of machine learning algorithms, e.g. that infer additional characteristics/traits of individuals 'baked into the data'. Validity, consent and dissemination issues

Public/private domain distinctions online

Considerations to pay particular attention to include:

- Blurred boundary between public and private spaces online

- People's expectations about how their online data traces may be used; gaining consent.

- Copyright/ownership of online date, permissions required

- Ensuring rigorous safeguards are in place to protect personal identities (esp when no consent gained).

Consent, withdrawal, debrief

Ensuring rigorous, effective, reliable procedures are in place, comparable with those commonly used offline, particularly:

- Implementing procedures to protect vulnerable groups and/or those not able to give consent

- Ensuring participants engage properly with study information sheets / informed consent processes

- Effectively tracking/monitoring participants to detect and address any upset/distress/harm caused, esp when these risks are higher (e.g. sensitive topics)

Ethics issues: summary / advice

- Some risks are enhanced when collecting data online; researchers should carefully consider and implement safeguards to minimise risks, referring to the various guidelines available.

- Navigating the public-private domain distinction online is not straightforward; researchers should consider people's expectations and any associated potential harms before using traces without consent.

- Using data traces without consent requires a clear justification, risk assessment, and proper safeguards to be put in place.

- Research involving sensitive/personal topics/data needs extra careful consideration and safeguards in place to minimise the enhanced risks involved: this is the **principle of proportionality**

Research in a pandemic

ILLUSTRATIVE EXAMPLES

EXAMPLES to reflect on…

- (1) Collecting lockdown stories from marginalised, vulnerable groups; pics/stories/text sent via social media apps, email, etc.; existing shared online posts and new material; families and children involved; disseminating stories, images, etc. in outputs, exhibitions; opt-out if dissemination of personal materials not agreed.

- (2) Collecting children's expressive texts, stories, images during COVID-19; public pavement drawings, items displayed in home windows, garden, parks, etc.; consent not gained, opt-out approach; dissemination /sharing of items.

EXAMPLES to reflect on…

- (3) Collecting posts from Twitter, Instagram, Facebook (public groups), etc.; not password protected and in public domain; informed consent not gained; dissemination practices/approaches; traceability of outputs?.

- (4) Using corporate online panel survey data provider; survey questions distributed to and managed by third party; no consent / PI procedures possible; fully anonymised data set delivered to researcher with no way to trace individuals; sensitive questions on health included (have you tested positive for COVID-19?); no debrief; paid participation.

Research in a pandemic: Lessons to be learned?

- Many opportunities for adapting traditional offline methods to online approaches, e.g. interviews, focus groups, ethnographies, experiments.

- Advantages and disadvantages of remote approaches become salient: some benefits, e.g. enhanced recruitment, lower costs, accessibility, empowerment.

- Possible disadvantages: reduced data quality, sample restrictions, lack of control, ethics dilemmas.

- Researcher skills: exploring new approaches, creative solutions, opportunities in a digitally connected world.