Research in the digital age: Secondary data, social media and governance issues

Dr Muir Houston

Chair CoSS CREC

Muir.Houston@glasgow.ac.uk

16th International Conference on Higher Education Reforms.

The Digital Society and Higher Education

11-13th September 2019

INSPIRING PEOPLE

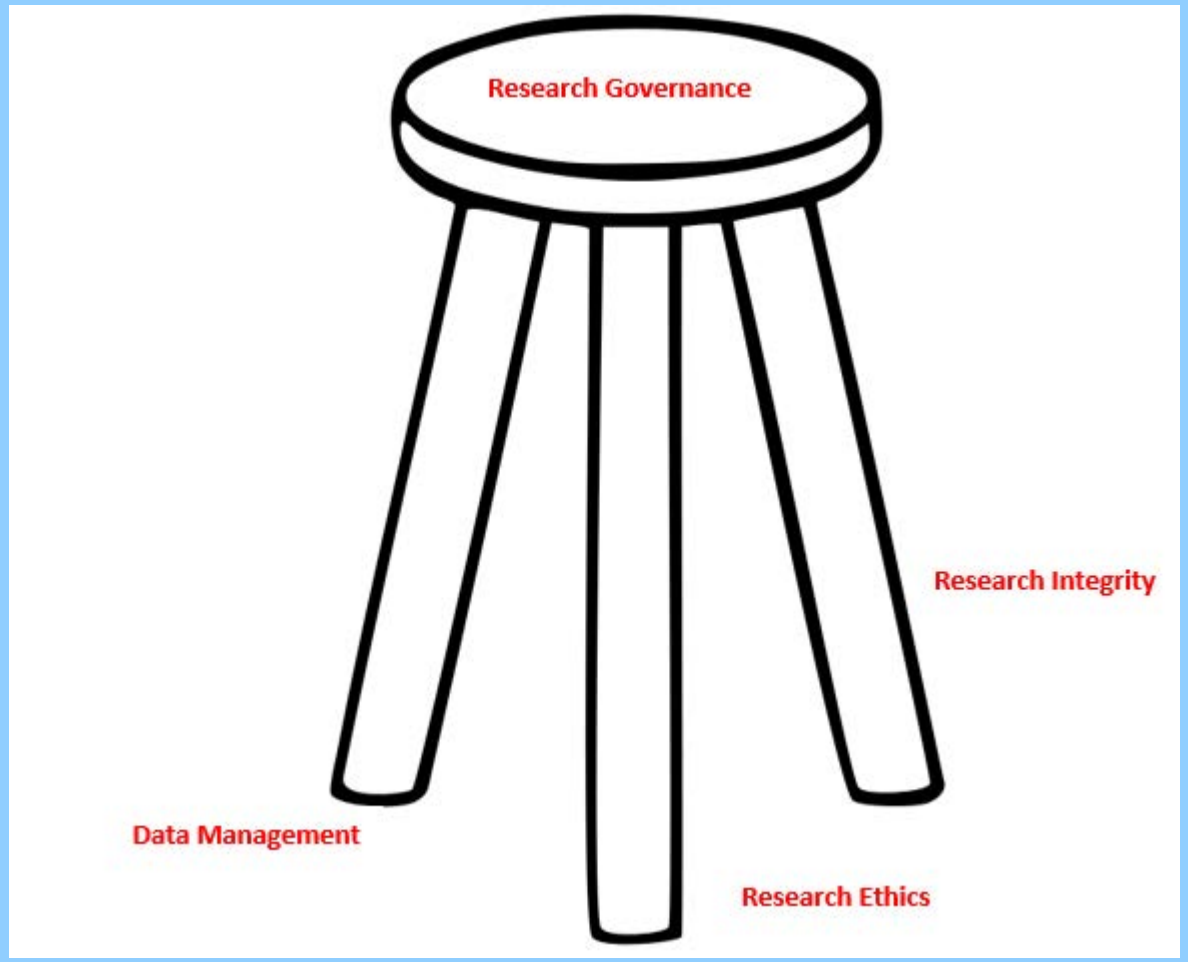
The issue

Big data in all its forms has vastly increased in terms of its scope and scale and covers almost all aspects of human life, our administrative, health, welfare, economic, retail and increasingly our social interactions.

However the introduction of GDPR and the negative publicity surrounding the activities of certain academic researchers and private sector organisations in the access, use, sale, algorithmic agglomeration and the ability to link and aggregate our data has resulted in renewed efforts to provide some form of guidance in relation to both the ethical and moral dimensions to the use of this type of data in academic research.

Research Governance

- Research integrity
- Ethics
- Data management



Research Ethics

- **Philosophically, can be linked to the works of Aristotle – Nichomachean Ethics (350 BCE) – and the pursuit of the good or virtuous life. (see also Confucius – Lunyu/Analects ca. 500 BCE)**
- **In the absence of agreed guidelines these aspects become more important in relation to big data and social media**

Developments in the Social Sciences

In order to maintain a climate of trust in the **practice and outcomes of social research**, it is inevitable that **systems of ethical assurance** will **be established** and will need to be **maintained** by all those engaged in work of this nature. The **trust** of the **public**, **professional colleagues**, those who **commission** and **fund research** and **those being studied** requires an **effective system of ethical review**, clear lines of responsibility and a **manageable degree of independent overview**. (Iphofen, 2011: p.5)

Iphofen, R. (2011) Ethical Decision Making in Social Research: A Practical Guide, New York: Palgrave MacMillan

ACADEMY *of* SOCIAL SCIENCES

The National Academy of Academics, Learned Societies and Practitioners in the Social Sciences

FIVE ETHICS PRINCIPLES FOR SOCIAL SCIENCE RESEARCH

1. Social science is fundamental to a democratic society and should be inclusive of different interests, values, funders, methods and perspectives.
2. All social science should respect the privacy, autonomy, diversity, values, and dignity of individuals, groups and communities.
3. All social science should be conducted with integrity throughout, employing the most appropriate methods for the research purpose.
4. All social scientists should act with regard to their social responsibilities in conducting and disseminating their research.
5. All social science should aim to maximise benefit and minimise harm.

Big Data and ethics

Libby Bishop (2017). Big data and data sharing: Ethical issues. UK Data Service, UK Data Archive. https://www.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethical-issues.pdf

Often differ from traditional research data in that they have not been generated specifically by researchers for research purposes. As a result, the usual ethical protections that are applied at several points in the research data life cycle have not taken place.

The speed of development in Big Data and associated phenomena, such as social media, has surpassed the capacity of the average consumer to understand his or her actions and their knock-on effects. We are moving towards changes in how ethics has to be perceived: away from individual decisions with specific and knowable outcomes, towards actions by many unaware that they may have taken actions with unintended consequences for anyone. (Zwitter, 2014)

The scale of big data

There is more data than ever in the history of data (Smolan and Erwitt 2012):

Beginning of recorded history till 2003—5 billion gigabytes

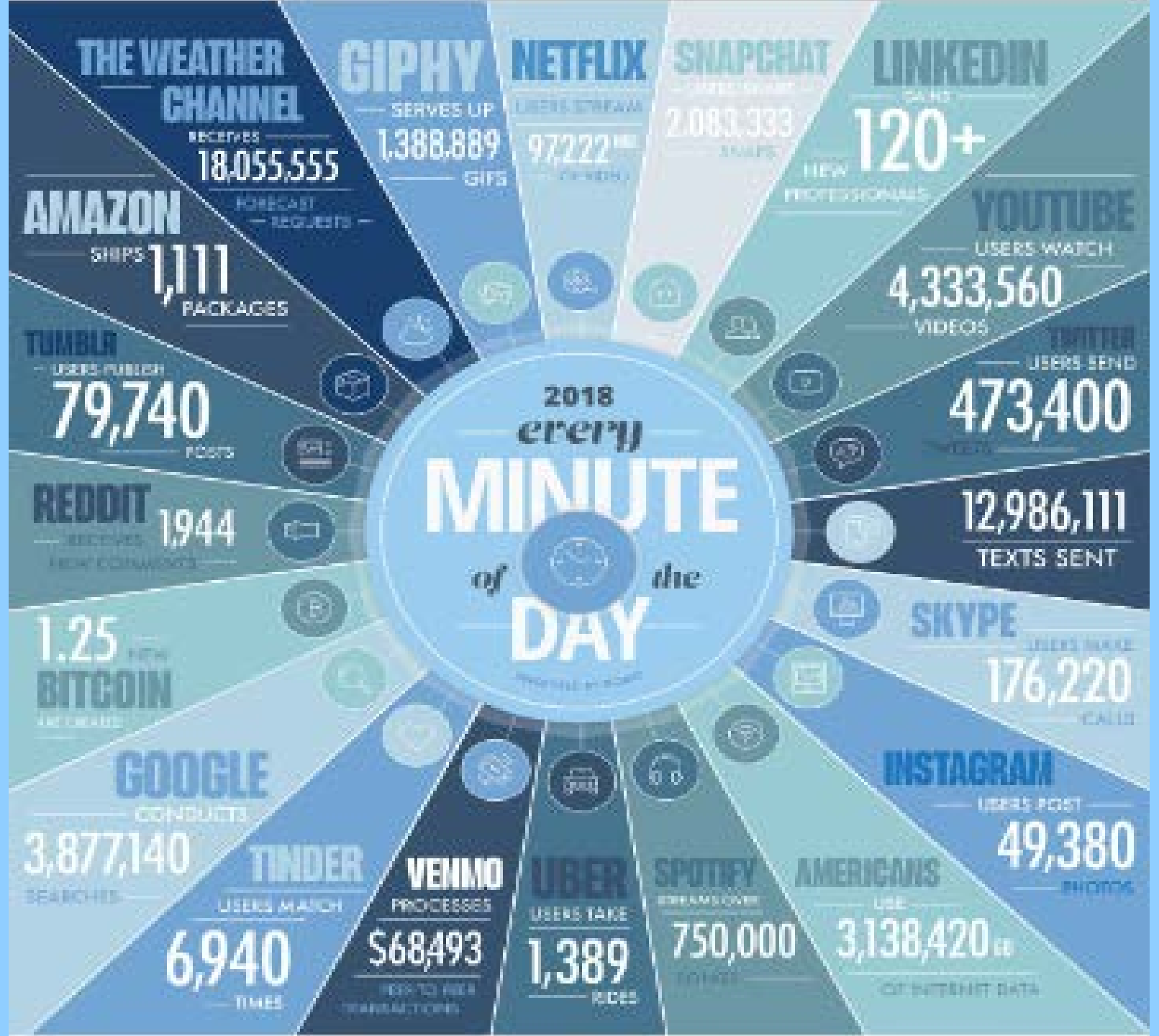
2011—5 billion gigabytes every two days

2013—5 billion gigabytes every 10 min

2015—5 billion gigabytes every 10 s

By 2025 463 exabytes every day

(1 GB= 1,000,000,000 bytes; 1 EB=1,000,000,000,000,000,000 bytes)


2018
every
MINUTE
of the
DAY
THE WEATHER CHANNEL
RECEIVES
18,055,555
FORECAST REQUESTS
AMAZON
SHIPS
1,111
PACKAGES
TUMBLR
USERS PUBLISH
79,740
POSTS
REDDIT
1,944
1.25
BITCOIN
GOOGLE
CONDUCTS
3,877,140
SEARCHES
TINDER
USERS MATCH
6,940
TIMES
VENMO
PROCESSES
$68,493
UBER
USERS TAKE
1,389
RIDES
SPOTIFY
750,000
AMERICANS
USE
3,138,420
INSTAGRAM
USERS POST
49,380
SKYPE
176,220
12,986,111
TEXTS SENT
TWITTER
USERS SEND
473,400
YOUTUBE
USERS WATCH
4,333,560
VIDEOS
LINKEDIN
120+
NEW
GIPHY
SERVES UP
1,388,889
GIFS
NETFLIX
97,222
SNAPCHAT
2,083,333

Big Data Sources

Corporate:

Facebook
Google
Twitter
Amazon
EBay
YouTube
Loyalty Cards
The Internet of ‘things’ …

Public:

Census
Longitudinal cohort studies
Administrative data:
- **Health records**
- **Social security**
- **Employment and income**
- **Crime records …**

Tensions in big data analytics

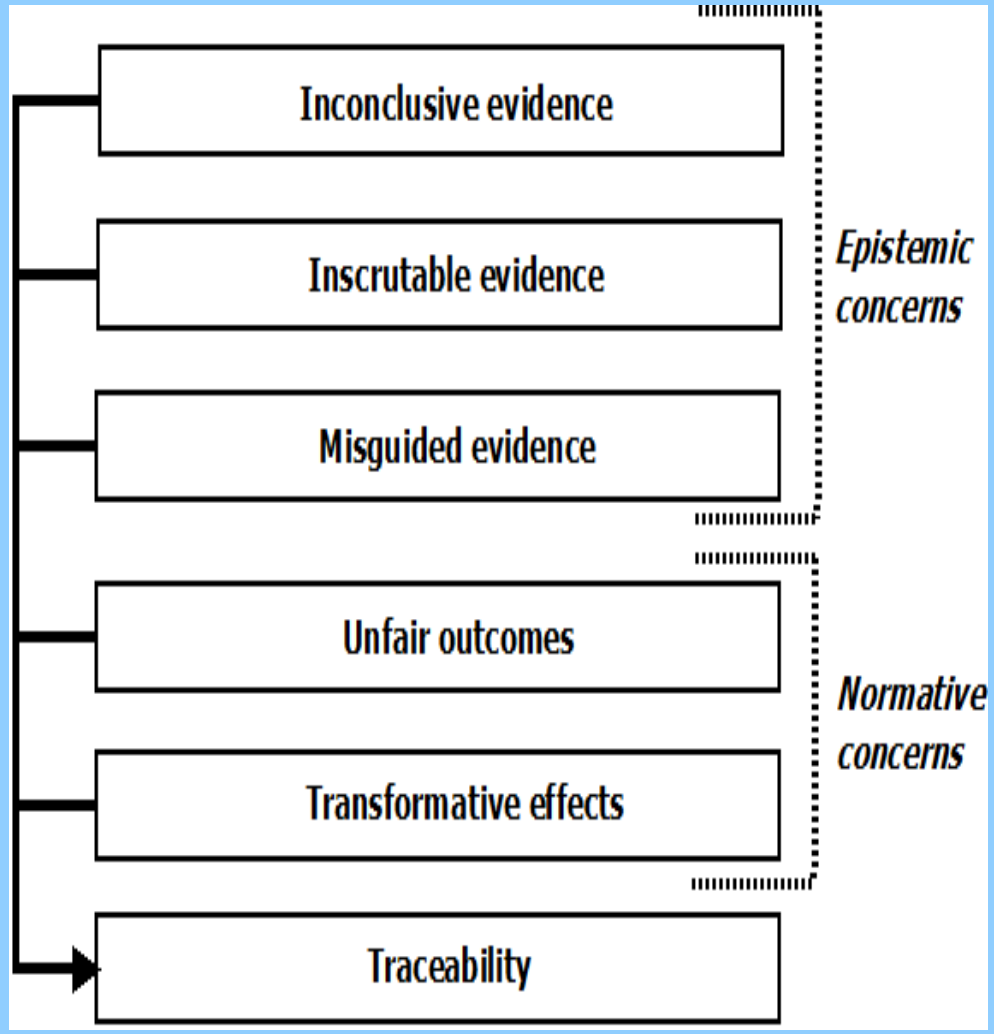
Drivers are Corporate market share, customer profiling and targeting, while in academia the public good and benefit should be key. However, academics are increasingly operating in a context of performativity and so publications, funding, citation counts and impact drive promotion.

Example when these collide – Facebook and Cambridge Analytica

Ethical issues in big data analytics

Machine learning and algorithms –

Analytics identifies relationships and small patterns across vast and distributed datasets (Floridi, 2012). New types of enquiry are enabled, including behavioural research on ‘scraped’ data (e.g. Lomborg and Bechmann, 2014: 256); tracking of fine-grained behaviours and preferences (e.g. sexual orientation or political opinions; (Mahajan et al., 2012); and prediction of future behaviour (as used in predictive policing or credit, insurance and employment screening; Zarsky, 2016) (source Mittelstadt et al 2016)



Social media and online ethics

What constitutes 'privacy' in an online environment?

How easy is it to get informed consent from the participants in the community being researched?

What does informed consent entail in that context?

How certain is the researcher that they can establish the ‘real’ identity of the participants?

When is deception or covert observation justifiable?

How are issues of identifiability addressed?

How do country-specific legal requirements (eg for data protection) apply in internet-mediated research that crosses national boundaries?

ESRC – internet mediated research

Swatman (2012) on privacy:

Is the space being researched seen as private by its users?

Are they aware they are being observed?

Is everything what it seems?

Are Fred X and Mary Y really who they claim to be? Age? Gender, Bots?

How do researchers ensure their participants really are anonymous?

IP addresses are (usually) traceable; Tweets may contain identifiers...

See also Internet Specific Ethical Questions section in AOIR (2012)

Ethics in social media research: where are we now?

Early on in the research we quickly realised that many of the learned society ethical resources were of little guidance, given their focus on non-digital data. Where addendums on using Internet data were written, they had little to say about social media. Papers were being published in reputable journals with tweets quoted verbatim, with unacceptable and ineffective methods of anonymisation, and without informed consent from users[1]. ……. Research on users' views of the repurposing of their social media data consistently shows that the majority wish to be asked for informed consent if their content is to be published outside of the platform which it was intended for[2].

Some guidance

Williams, M.L., Burnap, P. and Sloan, L. (2017) Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation, Sociology 51(11) tweet publish or not.pdf

Townsend, L. and Wallace, C. (2017), "The Ethics of Using Social Media Data in Research: A New Framework", Woodfield, K. (Ed.) *The Ethics of Online Research* (*Advances in Research Ethics and Integrity, Vol. 2*), SMEF.pdf

Association of Internet Researchers (AOIR) - 2012: Ethical decision-making and Internet research 2.0: recommendations from the AoIR ethics working committee

2012: This chart provides a useful starting point for internet researchers to consider ethics. AOIR

2002: Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee

Our response

Traditional ethics processes and forms not ‘fit’ for Big Data and Social Media so we created an alternative route specifically for this kind of data – we heard anecdotally people were working in these areas and not applying for approval – often citing erroneously public domain defence. Some sections are cited below:

1 Please provide details of the data you wish to collect or access – please include details of the platform, app, data archive, API, etc.

2 You should have consulted the specific Terms and conditions of the specific platform or data source; please answer the following:

2.1 What do the terms and conditions say about retention of datasets?

2.2 What are the rules regarding publishing or re-sharing collected data?

2.3 Are there specific provisions within the terms and conditions that permit research usage of data collected?

2.4 What are the explicit limits on usage that may be relevant for planned research work?

3 Have you consulted the relevant legal guidelines disciplinary, funder or institutional guidelines in relation to the specific ethical concerns research of this nature can raise? E.g. copyright/Intellectual Property Rights/contracts/licensing/ privacy/GDPR.

4.1a Does the data contain personal data?

4.1b Does the data contain Special Category Data?

5. Is the data open to being freely used for any purpose?
6. Can the data reasonably be considered to be in the public domain? (AOIR ethics guidelines)
7. Can the individuals reasonably expect their data to be observed/accessed by strangers? Or reasonably expect their data to be used in the way that you propose in your research?
8. Please provide details of the data access and collection methods and or techniques and also provide details of the method(s) of analysis. A Data Management Plan should also be submitted
9. Are any individuals vulnerable (see guidelines on vulnerable groups)?
10. Is the subject matter of sensitive or controversial nature?
11. Will individuals be identifiable from published outputs?
12. Please provide a justification for not seeking consent or explain why this would not be possible and again provide a justification for why the research should proceed.

"Social scientists do not have an unalienable right to conduct research involving other people (Oakes, 2002). That we continue to have the freedom to conduct such work depends on us acting in ways that are not harmful and are just. Ethical behaviour may help assure the climate of trust in which we continue our socially useful labours (AAAS, 1995; Jorgensen, 1971; Mitchell and Draper, 1982; PRE, 2002; Walsh, 1992). If we act honestly and honourably, people may rely on us to recognize their needs and sensitivities and consequently may be more willing to contribute openly and fully to the work we undertake."
(Israel and Hay, 2006: p3)

Israel, M. and Hay, I. (2006) *Research Ethics for Social Scientists*, London: Sage.

An afterthought

Complex social, ethical and legal issues to be overcome –

ICO (2017) Big data, artificial intelligence, machine learning and data protection

Or

-SKYNET-
NEURAL NET BASED ARTIFICIAL INTELLIGENCE
CYBERDYNE SYSTEMS

University of Glasgow

Thank you

Muir.Houston@glasgow.ac.uk

WORLD CHANGERS WELCOME