Internet Research: Ethical Guidelines 3.0
Association of Internet Researchers

Unanimously approved by the AoIR membership October 6, 2019

aline shakti franzke (University of Duisburg-Essen), Co-Chair
Anja Bechmann (Aarhus University), Co-Chair,
Michael Zimmer (Marquette University), Co-Chair,
Charles M. Ess (University of Oslo), Co-Chair and Editor

The AoIR IRE 3.0 Ethics Working Group, including: David J. Brake, Ane Kathrine Gammelby, Nele Heise, Anne Hove Henriksen, Soraj Hongladarom, Anna Jobin, Katharina Kinder-Kurlanda, Sun Sun Lim, Elisabetta Locatelli, Annette Markham, Paul J. Reilly, Katrin Tiidenberg and Carsten Wilhelm.[1]

Cite as: franzke, aline shakti, Bechmann, Anja, Zimmer, Michael, Ess, Charles and the Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0.* https://aoir.org/reports/ethics3.pdf

---

[1] A complete list of the EWG members is provided in **Appendix 7.2**. Contributions from AoIR members are acknowledged in footnotes. Additional acknowledgements follow **4. Concluding Comments.**

0. Preview: Suggested Approaches for Diverse Readers

The AoIR guidelines 2019 (Internet Research Ethics 3.0) are a collaborative document that builds on the previous guidelines (IRE 1.0, 2002; IRE 2.0, 2012) and should be read in conjunction with those. IRE 3.0 is written especially for researchers, students, IRB members or technical developers who face ethical concerns during their research or are generally interested in Internet Research Ethics.

As with the previous two AoIR documents, IRE 3.0 is premised on primary commitments to ethical pluralism and cross-cultural awareness, coupled with the experientially-grounded view that ethics starts by asking and answering critical questions rather than taking a more deductive, rule-oriented approach. In particular, long experience with both numerous real-world examples and critical reflection now shows us that each context and stage of research is different and provokes distinct questions.

This Preview seeks to guide you through the most relevant issues by asking for your research phase and context. In doing so, the Preview provides an initial starting point and points towards further resources. As with the IRE 3.0 guidelines and affiliated materials, we hope to build upon and improve this Preview in the future. If you find anything unclear or missing, don't hesitate to get in touch. Notice that each and every point is open for debate and ethics is an ongoing process.

If you are a **student,** you may find the 2002 guidelines (IRE 2.0) a good starting point of reflection. If you are looking for a draft to obtain informed consent, look into the appendix of the 2002 guidelines (IRE 1.0, https://aoir.org/reports/ethics.pdf; IRE 2.0, https://aoir.org/reports/ethics2.pdf).

If you are a **researcher**, you might want to begin with the 2019 guidelines to see if the provided resources are a helpful starting point. If you are looking for ways to solve the issue of informed consent you might want to have a look into the appendices of the 2002 document. A catalogue of important questions to start with can be found in the 2012 guidelines. For additional information, especially that focuses on recent technological developments, you might find the 2019 document a useful. Especially political and institutional pressure on researcher has gained importance.

If you are an **IRB, Ethical Review Board, or member of a similar research ethics oversight board**, we highly encourage you to take a look into the 3.0 guidelines. Notice that we emphasize deliberative processes of ethical reflection. At the same time, we believe that in times of Big Data, experimental research needs to be done that requires considerations beyond informed consent, but further includes careful reflection on research design, the context of research, and the basic requirement to minimize associated risks and harms. An ongoing ethical reflection might be more helpful and beneficial in the long term for society than now restricting research.

If you are a **developer**, you might find it helpful to have a closer look into the Companion Resources (6.0). These include “AI and Machine Learning: Internet Research Ethics Guidelines” (6.1) and an “Impact Model” (6.4) for ethical reflection which may be helpful.

1. Summary

This document introduces Internet Research Ethics (IRE) 3.0. We begin with a review of the AoIR ethical approaches and guidelines that we now designate as IRE 1.0 (Ess and the AoIR ethics working committee, 2002) and IRE 2.0 (Buchanan, 2011, p. 102; Markham & Buchanan, 2012; Ess, 2017). While driven by on-going changes and developments in the technological, legal, and ethical contexts that shape internet research, IRE 1.0 and 2.0 ground **a basic ethical approach** that continues as foundational for IRE 3.0.

IRE 3.0 is then illustrated by way of two elements – namely, (greater) attention to **stages of research** (a continuation of distinctions developed in 1.0 and 2.0) and what has become a standard problem of **informed consent** in particularly (but not exclusively) Big Data research approaches. We then list and briefly discuss the primary additional ethical challenges in IRE 3.0 as identified by the AoIR Ethics Working Group (EWG). We offer a **general structure for ethical analysis**, designed to help identify the ethically-relevant issues and questions, along with additional suggestions for how to begin to analyse and address these challenges in more detail. We offer this general structure as a guide for developing more extensive analyses of specific issues, both current and future. Initial examples of what such analyses can look like are offered here in **6. Companion Resources: Topical Guidelines and Ethical Frameworks.** We hope that additional analyses will be developed in response to emerging specific and ongoing socio-technical developments. In this way, we hope to produce a "living document," i.e., a set of guidelines that will continue to develop and unfold.[2]

2. Background and Introduction

The Association of Internet Researchers has foregrounded the development of Internet Research Ethics (IRE) since its inception. A first set of IRE guidelines was developed, approved by the AoIR membership, and published in 2002 (Ess and the AoIR ethics working committee). These guidelines – referred to here as IRE 1.0 – achieved broad acceptance and use internationally, in both diverse research communities as well as in U.S. Institutional Review Boards (IRBs), U.K. Ethical Review Boards, and their institutional counterparts and national data agencies in other countries (Buchanan & Ess, 2008).

In response to new technological developments, including the emergence of Social Networking Sites (from ca. 2005), the mobility revolution (i.e., dramatic expansion of internet access by way of mobile phones and devices, ca. 2008), and the initial emergence of Big Data, a second set of guidelines (IRE 2.0) were developed and published in 2012 (Markham

---

[2] This shift towards more open and dynamic documents and resources is inspired by a similar decision on the part of the Norwegian National Committee for Research Ethics in the Social Sciences and the Humanities (NESH), as the requisite response to a research and ethical landscape that continues to change and transform, often in dramatic ways, over a very short period of time (Enebakk, 2018). At the time of this writing, the implementation of the new EU General Data Protection Regulation (GDPR), coupled with dramatic new restrictions on research imposed by Facebook in response to the 2017 Cambridge Analytica scandal, are two prime examples of such potentially radical changes that will – yet again – require still further enhancement and revision of IRE (cf. Ess & Hård af Segerstad, 2019).

& Buchanan). While dramatically expanding and refining the first set of guidelines (IRE 1.0), the second set of guidelines, IRE 2.0, is explicitly rooted in and builds upon the first set.

2.1 Primary Ethical Norms

This continuity is apparent first of all in *the* ***primary ethical norms*** taken to be central for IRE, as initially rooted in *The Belmont Report* (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979): respect for persons, beneficence, and justice (for further discussion and elaboration, see: Buchanan, 2011, p. 84 f.; Markham & Buchanan, 2017).

2.2 A Basic Ethical Approach

Both documents likewise share a ***basic ethical approach,*** as initially developed in IRE 1.0 – what IRE 2.0 aptly identifies as "a process approach," one that aims to develop guidelines from the bottom up in a case-by-case based approach without making a priori judgements whether some research per se is unethical. This means focusing on the day-to-day practices of researchers in a wide range of disciplines, countries and contexts, in contrast to a more usual top-down approach that tries to provide a universal set of norms, principles, practices, and regulations (IRE 2.0, p. 5 f.). This process approach is first of all *reflective and dialogical* as it begins with reflection on own research practices and associated risks and is continuously discussed against the accumulated experience and ethical reflections of researchers in the field and existing studies carried out.[3] This further means an emphasis on the fine-grained contexts and distinctive details of each specific ethical challenges.

This process- and context-oriented approach further requires the point developed by Annette Markham: ethics is method – method is ethics (Markham, 2006)[4]. This is to say, our choice of methods vis-à-vis given research questions and design evoke specific ethical issues – but these in turn (should) shape our methodological choices. Moreover, especially as we are forced through the course of the research project itself to revise original research design and methodological choices – we are likewise confronted with the need to revisit our initial ethical assumptions and designs. A key virtue of this point is that it helps counter a common presumption of "ethics" as something of a "one-off" tick-box exercise that is primarily an obstacle to research. On the contrary – as our subsequent experience has demonstrated – taking on board an ongoing attention to ethics as inextricably interwoven with method often

---

[3] We begin with the philosophical and anthropological insight that most of us *are* centrally *ethical* human beings. This is to say: as acculturated members of our societies, we have acquired and learned how to apply in *praxis* a primary set of ethical norms and values. Moreover, we are primordially *cybernetic* beings, in the original sense developed by Plato, namely, creatures equipped with a particular kind of reflective *ethical* judgment (*phronēsis*): we know from our experience what is possible and not possible – and when our judgments prove to be mistaken in light of subsequent experience, we are capable of learning from these mistakes (Ess, 2013, p. 239).
This further leads to a *casuistics* or case-based approach, i.e., one that seeks to discern close analogies between a current context and ethical dilemma and those of antecedent cases and examples for the sake of developing the best possible *resolution* (i.e., as one among perhaps several defensible responses, and one open to further revision and refinement in light of subsequent experience, rather than a "solution" as single and final): see McKee & Porter, 2009; Markham & Buchanan, 2012, p. 7.

[4] See also Buchanan, 2011, p. 92 f.; Markham, Tiidenberg & Herman, 2018.

leads to *better* research as this attention entails improvements on both research design and its ethical dimensions throughout the course of a project.[5]

2.3 Ethical Pluralism and Cross-Cultural Awareness

Moreover, we presume ***ethical pluralism*** and ***cross-cultural awareness***. *Cross-cultural awareness* is required when internet research projects involves either researchers and/or subjects/participants/informants from diverse national and cultural backgrounds.

2.3.1 A Range of Ethical Frameworks (Utilitarianism, Deontology, Feminist Ethics, etc.)

As both IRE 1.0 and 2.0 document, such research thereby implicates often strikingly different legal and ethical frameworks, norms, practices, and traditions. As a primary example: European and Scandinavian approaches to research ethics in general and privacy matters in particular are strongly *deontological* – i.e., they emphasize first of all the central imperative to protect the rights basic to human beings as autonomous citizens in democratic societies. So, the NESH (2019) guidelines, for example, emphasize "dignity, freedom, autonomy, solidarity, equality, democracy and trust" as foundational norms and values in IRE – as these, in turn, are rooted in for instance the GDPR 2018 (pp. 16-21; cited in NESH, 2019, p. 3).[6]

By contrast, approaches in the US and the UK (and, perhaps, in other Anglophone countries, given their shared historical and philosophical backgrounds) are shaded in more *utilitarian* directions, considering the greater good for the collective and society in general. This means that ethical frameworks and decision-making in the latter, by comparison, are more willing to risk protection of basic rights for a relatively few for the sake of a greater good or benefit for the many, as resulting from the research (IRE 1.0, p. 8; Buchanan, 2011, p. 84). For example, the U.S. Office of Human Research Protections (OHRP) documents start with a clearly *utilitarian* approach: "Risks to subjects" are allowed, if they "are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result" (OHRP 2018, 11; cf. Principle 4, "Maximising Benefits and Minimising Harm", British Psychological Society, 2017, p. 18 f.).

As this example further suggests, we can analyse and seek to resolve our ethical dilemmas and challenges through *a range of ethical frameworks*: specifically, in addition to deontology and utilitarianism, feminist ethics, ethics of care, and virtue ethics have become increasingly central within Information and Computing Ethics (ICE) more broadly, and IRE more specifically (e.g., Jackson, Aldrovandi & Hayes, 2015; Zevenbergen et al., 2015; Ess, 2018). For accessible introductions to these frameworks, see: *utilitarianism*, Sinnott-Armstrong, 2019; *deontology*, Alexander & Moore, 2016; *virtue ethics*, Hursthouse & Pettigrove, 2018; feminist ethics and ethics of care are introduced in section **6.3**.

---

[5] This emphasis can also help counter a further problem with such "tick-box" approaches (and "top-down" presumptions about ethics more broadly) - namely, as the latter can inspire forms of self-censorship that would have us reshape our research terrains and designs, including choice of subjects and informants, in hopes of more easily "getting by" an ethical review process (Thanks to Carsten Wilhelm for pointing this out).

[6] For more on the relationship between legal aspects and ethics, see the section "Legal aspects" below.

2.3.2 Conceptions of Selfhood

Still larger differences emerge more globally – beginning with (largely) Western *assumptions regarding human beings* as primarily *individual* persons and moral agents, vis-á-vis more *relational* conceptions of selfhood, in which our sense of identity is largely constituted by multiple relationships, spanning from the family through larger communities and, in some cases, natural and "supernatural" orders as well. These are defining elements in many non-Western and Indigenous cultures: at the same time, these conceptions are changing and include emerging middle grounds, such as recent conceptions of *relational autonomy* (Christman, 2004; Westlund, 2009; Veltman & Piper, 2014; see Ess, 2014).

2.3.3 Ethical Pluralism

To be sure, these contrasts can hence be striking and, in some ways, irreducible. But the IRE 1.0 endorsement of *ethical pluralism* has proven to work out well in practice in the great majority of cases in our experience. That is, in contrast with simple ethical or cultural relativism, ethical pluralism resolves such contrasts by way of showing, e.g., how different practices of privacy protection (as in the stark contrasts between Norway and the US) may be understood as diverse interpretations, applications, or understandings of a *shared* norm (namely, privacy: see IRE 1.0, footnote 6, p. 29 for further examples). An essential virtue of this approach is precisely that it acknowledges the legitimacy of specific local norms, practices, etc., while nonetheless conjoining these across significant differences with more encompassing and shared norms, values, and so on (see **3.2.3 Ethical pluralism**, below; the example of ethical pluralism presented in the Thai context by Soraj Hongladarom (2017) in **3.2.4 Cultural Dimensions**, below).

2.3.4 Judgment Calls: Guidelines, Not Recipes

But such pluralistic approaches again foreground the role of judgment and the possibility of multiple, ethically legitimate *judgment calls* – in contrast, that is, with more rule-bound, "one size fits all" ethical and legal requirements.

Taken together, all of this means that the best we can do is develop "guidelines, not recipes" (IRE 1.0, p. 3). More carefully, given the range of possible ethical decision-making procedures (utilitarianism, deontology, feminist ethics, etc.), the multiple interpretations and applications of these procedures to specific cases, and their refraction through culturally-diverse emphases and values across the globe – the issues raised by Internet research are *ethical* problems precisely because they evoke more than one ethically defensible response to a specific dilemma or problem. *Ambiguity, uncertainty, and disagreement are inevitable* (ibid., p. 3 f., emphasis in the original).

2.3.5 Asking the Right Questions

As a result, the emphasis in both IRE 1.0 and 2.0 is on ***asking the right / relevant questions*** (see IRE 1.0, pp. 4-8; IRE 2.0, pp. 8-11). That is, once basic frameworks and guidelines are

introduced and illustrated by way of examples – both documents then foreground the sorts of questions that researchers, as well as their oversight boards, should ask in order to (a) better discern the primary ethical issues in play within a given research approach and context, and (b) to inspire precisely the dialogical reflection usually needed for developing both individual and collective *judgments* as to the best resolutions of core ethical challenges.
As a start, **IRE 1.0** highlighted the following guiding questions:

A. Venue/environment - expectations -authors/subjects - informed consent

- Where does the inter/action, communication, etc. under study take place?
- What ethical expectations are established by the venue?
- Who are the subjects? Posters / authors / creators of the material and/or inter/actions under study?
- Informed consent: specific considerations (e.g., timing, medium, addressees, specific research purposes)

B. Initial ethical and legal considerations

- How far do extant legal requirements and ethical guidelines in your discipline "cover" the research?
- How far do extant legal requirements and ethical guidelines in the countries implicated in the research apply?
- What are the initial ethical expectations/assumptions of the authors/subjects being studied?
- What ethically significant risks does the research entail for the subject(s)?
- What benefits might be gained from the research?
- What are the ethical traditions of researchers and subjects' culture[7] and country? (AoIR, 2002, p. 1)

IRE 2.0 (Markham & Buchanan, 2012) dramatically expanded upon this earlier list. Again, these expansions were catalysed by the rise of Web 2.0, especially sites featuring user-generated content (e.g., YouTube) as well as Social Networking Sites (SNSs) more generally; the "mobility revolution" as more and more people began accessing the internet via mobile devices; and early Big Data approaches (see Markham & Buchanan, 2012, pp. 8-11). First of all, these developments markedly increased challenges to protecting privacy. One notorious problem was that especially younger people were sharing more and more information online in what amounted to public or quasi-public fora (the latter protected, e.g., by passwords, registered profile requirements, etc.). But they often nonetheless *expected* that these exchanges were somehow private – either individually private or in some form of group privacy. Even though these expectations were not warranted by the technical realities of a given forum or SNS, especially deontological ethics calls for respecting these expectations,

---

[7] Internationally, the nature and even existence of oversight boards differ substantially. Linked to local political and academic cultures, these aspects are the object of ongoing discussions both interdisciplinary and specific to national and disciplinary research associations and fields. See e.g., Domenget & Wilhelm (2017).

and thus protecting these exchanges as anonymous or pseudo-anonymous, and/or requiring informed consent for their use (IRE 2.0, p. 6 f., 8; footnote 12, p. 14).

A further set of questions were gathered under the umbrella query "How are we recognizing the autonomy of others and acknowledging that they are of equal worth to ourselves and should be treated so?" (IRE 2.0, p. 11). Questions here focused on the complexities of *informed consent* as a primary way of protecting the (deontological) norms of *autonomy* and *equality*. The basic question of how to "ensure that participants are truly informed?" is raised here, along with recognition that online fora and our engagements within them are intrinsically *relational*: informed consent may need to be sought not only from individuals, but also "from communities and online system administrators?" (ibid.).

A last set of questions illustrate the growing importance of Big Data research, beginning with umbrella question "How are data being managed, stored, and represented?" (IRE 2.0, p. 9). The questions presented here help researchers take on the ethical matters of properly securing, storing, and managing "potentially sensitive data" (ibid.). Utilitarian concerns of both benefits and risks of attempting to de-identify data are raised vis-à-vis the central requirements to protect anonymity, privacy, and confidentiality. The last question here – presciently – asks us to consider how future technological developments – specifically, "automated textual analysis or facial recognition software" – might compromise such protections (ibid*.,* p. 9 f.).

The utility of asking the right questions – specifically for the sake of catalysing researchers' own judgments and reflections based on their often extensive experience – is suggested by the subsequent emergence of additional question-oriented guidelines, such as the NESH guidelines (2019), and DEDA (Data Ethics Decision Aid for Researchers), developed by aline franzke for the University of Utrecht - consisting of over 230 questions: (<https://dataschool.nl/deda/deda-for-research/?lang=en>). In this document, we build on and extend this question-oriented approach both in our general considerations and specifically in the **3.2 A General Structure for Analysis**, below.

3. Internet Research Ethics 3.0

As with the transition from 1.0 to 2.0, ongoing technological developments – most especially under the umbrella of Big Data and associated technologies of mining and collecting data – have evoked both new versions of familiar research ethics issues (e.g., informed consent) as well as relatively novel issues. In seeking to address these issues, the Ethics Working Group 3.0 (EWG) continues this process of revising and expanding as necessitated by still more contemporary developments.

As a first step, we refer those seeking more detailed and helpful introductions to the primary features of IRE to IRE 1.0 and 2.0. Additional guidelines can also be helpfully consulted: see **3.2.1 Related Guidelines,** below.

As we have seen, these documents outline the basic ethical requirements for IRE, beginning with respect for persons, beneficence, and justice. Respect for persons, for example, grounds primary Human Subjects Protections such as protecting identity by way of anonymity, confidentiality, and informed consent. IRE 1.0 and 2.0 then offer more detailed exploration of additional ethical norms in both social sciences- and humanities-based research in conjunction with primary ethical frameworks. Reading into these documents can then be usefully supplemented by more recent overviews and case studies.[8]

We now turn to some initial topics IRE 3.0, namely attention to ***stages of research*** and ***protecting the researchers.*** These are followed by a list of ***primary topics*** to be further explored in IRE 3.0, coupled with a ***general structure for analysis***, as defined by primary considerations and questions.

3.1 Initial Considerations

3.1.1 Stages of Research

One of the key contributions of IRE 2.0 was a taxonomy of the ***stages of research***, beginning with a primary distinction between the initial phases and then the dissemination phases of a research project (Markham & Buchanan, 2012, p. 5). IRE 3.0 extends this approach by way of a more extensive taxonomy of stages, especially as suited to Big Data as well as other large-scale research projects, as defined by national and/or international funding processes. These stages or phases include:

Initial research design, including initial considerations of potential ethical issues, in seeking grant funding.

Initial research processes, including acquiring data: these stages typically entail specific requirements for de-identifying data, securely storing data, and so on.

Analyses, including assessment of how use of particular techniques, formulas, or instruments may re-identify data through aggregation of multiple data sets. This includes considering downstream ethical effects arising from the unpredictability of now-common analytical processes, often algorithmically driven.

---

[8] A bibliographic database of these resources is available: <https://www.zotero.org/groups/2235867/aoir-ethics/items/>

Dissemination – i.e., various ways of publicizing research findings and data: this typically includes conference presentations (including injunctions not to tweet or otherwise share sensitive information presented within relatively closed contexts) and publications. An increasingly pressing set of issues are further generated by requirements by national and international funding bodies to make research data openly available.

Close of the project – including the destruction of research data and related materials.[9]

Note that research cannot (always) be clearly structured in different stages, but frequently reflection on ethics is interwoven. This staged approach thus helps structure and parse research ethics concerns in more defined ways that thereby allow for a more precise analysis and fine-grained resolution.

3.1.2 Informed Consent

These distinctions are particularly helpful in what has emerged as a standard problem in more contemporary Big Data projects – namely, **informed consent**.[10] Such projects use a range of data collection techniques, including automated scraping for semi-public data and the use of API (Application Programming Interface) processes for accessing private data. Especially as personally identifiable information (PID) and/or sensitive information is collected, strong steps are required to protect the identity of individual subjects and, in many cases (where possible), to obtain their informed consent to the research being carried out upon them and/or their data. Such consent is manifestly impracticable in the case of Big Data projects, however, resulting in a serious ethical dilemma. Researchers have taken different steps to mitigate risk against research subjects in such cases (Bechmann & Kim, 2020). Some researchers are trying to obtain first-degree informed consent, others are focusing on deleting names and other highly identifiable information from the dataset when storing and processing the data. Most commonly, researchers pseudonymize their data separating keys from the actual dataset and also make sure to justify both any questions using/processing sensitive data and/or how risk in this process has been dealt with (e.g. storage, aggregation of data, publication of aggregates).

Another way of trying to mitigate risk and resolve the dilemma is by reserving the acquisition of informed consent to the *dissemination stage* of a project, i.e., by asking for informed consent from specific subjects for the publication of a quote or other data that might make them and their personal information identifiable (e.g., through a string search or more sophisticated data-matching techniques). Especially as such quotes or references may be

---

[9] These suggestions are partly based on the ethical research guidelines developed by aline franzke for the University of Utrecht. See aline franzke & Mirko S. Schaefer: Data Ethics Decision Aid for Researchers (DEDA researchers), unpublished questionnaire and report, University Utrecht. On dissemination ethics, see: Ess & Hård af Segerstad, 2019; Rensfeldt, Hillman, Lantz-Andersson, Lundin & Peterson, 2019. See also "agile ethics," Neuhaus & Webmoor, 2012.

[10] Example forms for informed consent will be included in subsequent versions of this document.

necessary for only 10-20 research subjects, informed consent is thereby easily tractable (e.g., Van Schie, Westra & Schaefer, 2017; Ess & Hård af Segerstad, 2019).[11]

3.1.3 Protecting the Researcher(s)

IRE 3.0 further emphasizes attention to the growing need for ***protecting the researchers***, as well as our subjects and informants. Phenomena such as "Gamergate" and similar events highlight comparatively new risks and levels of risk posed to researchers whose work – and/or simply their public identity (e.g., ethnicity, minority identity, sexual identity, political activism, etc.) – triggers strong ideological reaction: these include death threats, "doxing" (publishing private information about the researchers, thereby fanning the flames of further hate speech, threat, etc.) and so on (e.g., Massanari, 2016). Similarly, research on violent online and offline political extremists, including jihadists, risks direct threats and retaliation should researchers' identities become known. As well, simply reviewing and curating, e.g., videos of beheadings and other forms of extreme violence can have serious consequences for researchers' psychological health and well-being that in turn require – ethically, if not always legally – one or more therapeutic counter-measures as part of the research project and process (VOX-Pol, 2018). Accordingly, collecting resources offering diverse ways of protecting and enhancing researcher safety is a primary component of IRE 3.0.[12]

In addition, another essential measure is that institutions develop policy detailing support procedures for researchers experiencing online threats or harassment related to their work.[13] Beyond these concerns with researcher safety, data-intensive research methods implicate a wide spectrum of issues surrounding data security, see **Appendix 7.1 Keith Douglas, "Operational Security: Central Considerations".**

---

[11] It is also an *ethical-methodological question*: why would publication of an exact quote be necessary? Such publication is typically required, e.g., by methods of Critical Discourse Analysis, i.e., as documental critical examples necessary to a larger analysis or argument. But of course, other methods may not require such publication in the first place. In some cases, transforming the quotations (Markham, 2012) and images (Tiidenberg, 2018; Warfield, Hoholuk, Vincent & Camargo, 2019), or creating samples of discourse that are composites from various participants may mitigate risk. Jeremy Hunsinger also points out: Informed consent can also be necessary to modify, erase, or misrepresent people's identities. In some contexts, people have a right to have their words and identities preserved, even against the judgement of the researcher who may be trying to protect them. However, the researcher cannot assume they should be protected but must get their consent to protect them. Do not assume that modifying, misrepresenting or erasing someone's participation or language to protect them is the ethical action, it is never ethical without consent (Personal communication, 2019).
For additional considerations of the contemporary complications of informed consent, see Obar, 2015; Halavais, 2019; *Nature,* 2019. Our thanks to Jeremy Hunsinger for opening this conversation.

[12] Leonie Tanczer (UCL, PETRAS IoT Hub), for example, recommends the following guides for teaching yourself tools and skills:

- https://ssd.eff.org/en/playlist/academic-researcher
- https://tacticaltech.org/projects/security-in-a-box-key-project/
- https://www.accessnow.org

See also: Lindsay Blackwell, Katherine Lo, Alice Marwick, http://datasociety.net/output/best-practices-for-conducting-risky-research/

[13] Our thanks to Fiona Martin and colleagues from "The academy and gendered harassment: Individual, peer and institutional support and coping in harsh online environments" 2019 preconference workshop (Jaigris Hodson, chair) for pointing out the absence of relevant university policies.

3.1.4 Additional Topics

The work by the co-chairs and EWG, including two rounds of IRE panels at the annual AoIR conferences (2017, 2018), have helped identify an extensive range of additional topics and issues requiring attention in IRE 3.0. As but two examples:

- **Quality of research questions**

This emphasizes the importance of developing and articulating research questions so as to take on board how method and ethics always interweave with one another – and will very likely do so throughout the research project. A further consideration here is to hone the research questions so as to ensure that no more data, especially sensitive and personal information, is collected than is strictly necessary (the principle of "data-minimization", Ess & Hård af Segerstad, 2019).

- **Power and Power Imbalances**

This point foregrounds ethical issues connected with the fact that Big Data research, as well as all research on people's networked practices, social media behaviours, and so on, takes place within and depends upon information architectures and infrastructures, including social media venues as well as platforms in a broader sense (e.g., Google, Microsoft, Apple, and so on). While these platforms can offer extraordinary research possibilities through Big Data techniques their design and day-to-day functioning can impose constraints largely outside the control of researchers. Special care should be taken when collecting data from social media sites in order to ensure the privacy and dignity of the subjects. A current example is Facebook's recent decision to severely restrict research by way of using APIs (Bruns, 2018). A very large array of additional topics and issues have been suggested that likewise require reflection and discussion: these are summarized in **3.2.9 Other Concerns**.

3.2 A General Structure for Ethical Analysis

Based especially on more recent experience with these sorts of issues, we propose that each of these topics (as well as similar or emerging ones) adopt the following structure – something of a conceptual taxonomy or list of essential components – as part of a relatively complete analysis of and possible resolutions to the ethical challenges in play.

3.2.1 Related Guidelines.

Beyond the guidance available in IRE 1.0 and 2.0, researchers are encouraged to explore other guidelines, including: NESH (The [Norwegian] National Committee for Research Ethics in the Social Sciences and the Humanities) (2019) *A Guide to Internet Research Ethics*. Oslo: NESH.

The NESH Guidelines are especially helpful as:

(a) They take on board the strict new privacy protection requirements articulated in the GDPR (2018).

(b) They explicitly "combine individualist and relational perspectives on human life, which is especially relevant for distinguishing between private and public matters on the internet" (2019, p. 6) - and thereby foreground Nissenbaum's account of privacy as *contextual integrity* (2010), i.e., as dependent upon the relational contexts within which information is shared, in contrast with prevailing individualist approaches (e.g., in the EU and US) that focus on more static concepts such as "Personally Identifiable Information" (PII), etc.

(c) They foreground the centrality of (reflective) *judgment.* Specifically, NESH states that a central function of the guidelines is "…to aid in the development of *sound judgement* and reflection on issues pertaining to research ethics, resolutions of ethical dilemmas, and promotion of good research practices" (2019, p. 4, emphasis added, CME).

Additional national guidelines include The Canadian Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (2014) (http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/). The ethical guidelines from organizations such as the ACM, IEEE, ICA, etc. and those of other civil society organizations like the UN may also sometimes be helpful. This will be more likely true in the future as new guidelines are being developed cooperatively between these and related organizations.

In addition, as noted in IRE 2.0, funding agencies such as the U.S. National Science Foundation and the Office for Human Research Protection (specifically: "Subpart A of 45 CFR Part 46: Basic HHS Policy for the Protection of Human Subjects", https://www.hhs.gov/ohrp/sites/default/files/revised-common-rule-reg-text-unofficial-2018-requirements.pdf) are also good sources for more specific and nuanced understandings, e.g., of consent.[14]

Recent changes to the **U.S. Common Rule** (OHRP 2018) are particularly relevant to IRE. To begin with, there are new activities deemed not to be research, including scholarly and journalistic activities such as oral history, journalism, biography, literary criticism, and legal research, and so are exempt from its requirements for informed consent and related protections, as well as how Institutional Review Boards (IRBs) are to be constituted and function. This is the US only, of course, but given the historical and contemporary influence of these US regulations and practices, these developments will have significant consequences for a great deal of internet research. [15]

A particular concern here is that "many universities seem loathe to loosen the existing requirements and are hewing to a higher standard - needlessly. IRBs in the US continue to be hyperconservative" (Steve Jones, personal communication 2019). Similar observations have been offered from the Australian context (Naomi Barnes, personal communication, 2019): indeed, the tendency of ethical review boards to be overly cautious is a widespread complaint internationally. (This points to one of the founding justifications for and central subsequent uses of the AoIR IRE guidelines – namely, the importance of being able to help inform ethical review boards of the distinctive characteristics of internet research, as well as to provide researchers with resources, beginning with the guidelines themselves, to help them in the processes of negotiating the process of seeking approval for their research with such boards.

---

[14] Anne Henriksen has pointed to Zook et al. (2017) and to *Data, Ethics & Society* (https://bdes.datasociety.net) as particularly helpful here as well.

[15] Our thanks to Steve Jones for these observations.

We hope the current document, including the Companion Resources collected in **6.0**, will continue to be useful in these ways).

Finally, comparative work on national ethics initiatives might also yield interesting insights as to the cultural and social representations of ethical research[16] (e.g., Domenget & Wilhelm, 2018; see also **3.2.4 Cultural Dimensions,** below).

3.2.2 Legal Aspects

Typically, law falls behind both ethical reflection and technological development: nonetheless, both **national** and **international laws** are often primary frameworks for defining specific requirements within IRE, beginning with privacy protection.

In addition to interrogating the relevant national laws relevant for specific researchers, IRE 3.0 typically intersects with:

- **GDPR**, including questions of copyright derivatives, database derivatives, and so on (see, e.g. Kotsios et al., 2019).
  (NB: the implications of and research literature on the GDPR for IRE are only slowly emerging, following its implementation in 2018. For a first overview, see the "Data Management Expert Guide" developed by the Consortium of European Social Science Data Archives.[17] A future iteration of this document will include a companion resource that will develop a fuller overview).
- **Terms and Conditions**
  The increased reliance on internet platforms and applications to gather research data and/or recruit research subjects means researchers will increasingly confront terms of use that dictate the conditions under which such research activities can take place, if at all. For example, Twitter's terms of service for third-party developers requires that they "respect users' control and privacy" by deleting any "content that Twitter reports as deleted or expired," as well as any content that has been changed from public to private.[18] These terms and conditions suggest that researchers relying on Twitter-based datasets must continually check if any tweets have been deleted from the platform, and then remove them for their research dataset accordingly (see, for example, Tromble & Stockmann, 2017).

  Various research activities might violate the terms and conditions of online platforms. For example, Facebook and LinkedIn prohibit the automated scraping of data from their platform, limiting researchers' ability to collect needed data through such means (e.g., Lomborg & Bechmann, 2014). Other forms of research might require violating a platform's prohibition of creating multiple fake user accounts or contributing prohibited content. Recent efforts on the part of platforms to increase privacy and security, e.g., through verifying one's identity by requiring a working

---

[16] The French research program GENIC, "Group on Ethics and the Digital in Information-Communication," in the context of the absence of guidelines in communication and media studies in France and inspired by the AOIR guidelines 1.0 and 2.0, wishes to enrich the current discussion on recent current practice of research ethics in communication and information studies with an ambition towards cross-cultural comparison.

[17] https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide

[18] https://dev.twitter.com/overview/terms/policy

phone number, make it more difficult to take this approach. Still, it remains debatable as to whether following a website's terms and conditions is a legal requirement for academic researchers whose work benefit the knowledge level of society at large. In the United States, failure to follow such terms may violate the Computer Fraud and Abuse Act (CFAA), and interpretation that is being fought in court by the ACLU on behalf of academic researchers who feel such a stance chills research (e.g., Zetter, 2016). There are tools for doing research for scraping and analysing Twitter that report back to the researcher if a tweet has been deleted.

3.2.3 Venues and Platforms

As the focus on terms and conditions indicates, much depends on the specific venues within which research will take place, beginning with user requirements under the Terms of Service of online platforms such as Facebook and Instagram, Diaspora, Pinterest, Snapchat, etc. – but also Google, Microsoft, and others as platforms whose access and use is also contingent upon ToS agreements.[19]

3.2.4 Cultural Dimensions

As noted from the inception of IRE 1.0, internet-facilitated communication almost always crosses multiple cultural boundaries, thereby implicating multiple local cultural norms, practices, beliefs, etc. – both in terms of local uses, approaches, etc. and local research ethics. These challenges begin with differences in ethical values, schools, and traditions (including frameworks such as utilitarianism, deontology, feminist ethics, ethics of care, virtue ethics, and so on). They often further entail foundational differences such as greater emphasis on human beings as *individual* persons/agents vis-á-vis more *relational* conceptions (Ess, 2014). As we have also seen, at least some of these differences can be approached by way of *ethical pluralism* as a way of conjoining shared values with diverse interpretations / applications, etc.

In addition, there is a growing literature on IRE within more specific cultural contexts that may be relevant and useful to consult, e.g. For example, Soraj Hongladarom contrasts the ethical assumptions and frameworks of Confucians and Buddhists. Confucians – like most "Westerners" – believe in a self as *real*; Buddhists consider the self to be "ego delusion," an illusion to be overcome. Nonetheless, both agree on a foundational ethical norm - that "…an individual person is respected and protected when she enters the online environment" (Hongladarom, 2017, p. 155). This is a *pluralistic* solution as a basic norm is shared across radically diverse ethical frameworks.

Hongladarom likewise argues for such pluralism as conjoining the sometimes even greater differences between Western and Eastern traditions - specifically regarding *informed*

---

[19] See: Halavais, 2019; Puschmann, 2019; Bruns, 2019.
As noted above, there is some indication that ToS agreements may not be as legally binding as the platforms would like us to believe. On the other hand, developing fake profiles is a form of deceptive research, which opens up additional ethical issues – e.g., whether or not researchers should ask/require their students to use such a research design?

consent in the context of Thai research on mothers' blogs. Thai culture is more collective (in part, as relying on a more *relational* sense of selfhood as stressed in both Confucian and Buddhist traditions). The ethical force of an originally Western requirement for *individual* informed consent is not immediately recognized by Thai researchers; nor is the importance of ensuring *individual* anonymity when needed. Hongladarom observes, however, that legitimate versions of these requirements are understood and applied in the Thai context nonetheless.

Hongladarom points towards an emerging global IRE, constituted in part by shared norms understood in such a pluralistic fashion – i.e., allowing for differing interpretations and applications as these are diffracted through specific local traditions. This requires continued emphasis on the pluralism enunciated in AoIR 2002, beginning with recognizing and respecting that a given national or cultural IRE is "… grown from the local source, meaning that the [ethical] vocabulary comes from the traditional and intellectual source of the culture in which a particular researcher is working" (Hongladarom, 2017, p. 161).

Additional discussions and resources regarding "culture":

- Denison & Stillman (2012) discuss a case study from South Africa regarding the academic and ethical challenges in participatory models of community research (The article includes sections on data management, data ownership and access rights).
- Digital Ethnography in the Asian context – see Mukherjee (2017).
- Within Europe, especially France (Domenget & Wilhelm, 2018), Germany (Heise & Schmidt, 2014; Pentzold, 2015), and Scandinavia (Ess & Hård af Segerstad, 2019)
- Internet Research Ethics in Canada (Seko & Lewis, 2017).
- Internet Research Ethics in a cross-cultural big data social media case study in Denmark and South Korea (Bechmann & Kim, 2020).
- Andrew Whelan (2018) develops a close critique of ethics forms from 10 Australian universities, first of all substantiating "the standard critique of prospective ethics review from social media researchers: that the opportunity for a proper conversation about research ethics in the community of researchers is supplanted by an administrative exercise in 'box ticking', while more broadly foregrounding "the ethical consequences of the stance [these forms] require the applicant to take with respect to prospective research participants, and the implications of their formulation of research as a process of data extraction" (1).
- Eileen Honan (2014) argues that Western concepts of protecting subjects' privacy via informed consent can sometimes become unethical in the context of Papua New Guinea (p. 8 f.).[20] At the same time, Bassett & O'Rierdan pointed out quite early on within Western contexts (2002) that default assumptions concerning the need to protect privacy sometimes run counter to other ethical considerations, as when well-meaning anonymization of participants in LGBQ groups were criticized by participants as reinforcing their marginalization and silencing in the larger society. For further discussion, see IRE 1.0, footnote 12, p. 31.

---

[20] Thanks to Naomi Barnes for this point and reference.

Finally, different platforms have different use cultures that lead to different ethical implications. Depending on users' personal perception of privacy, the role of social norms, and the government – people perceive platforms practices as intrusive in diverging degrees (e.g. Hudson & Bruckman, 2004; Beninger et al., 2014; Patterson, 2018).

3.2.5 Involved Subjects

A primary ethical imperative is to avoid harm - to subjects as well to researchers. But the primary question is, who are the subjects? This question then interacts with a classical ethical principle: the greater the vulnerability of our subjects, the greater our responsibility and obligation to protect them from likely harms (cf. IRE 1.0, p. 5; Tiildenberg, 2019). Some more specific considerations include:

- Downstream harms or harms after the fact (Harmon, 2010; Sterling, 2011; Markham, 2018).
- Minors (e.g. Robards, 2013; Hudson et al., 2004).
- Politically sensitive research (Kaufmann, 2019; Reilly & Trevisan, 2016).
- Women (Luka, Millette & Wallace, 2017).
- Groups (e.g. persons with disabilities, Trevisan & Reilly, 2014; Hård af Segerstad, Kasperowski, Kullenberg & Howes, 2017).
- Special emotional states such as grieving and/or trauma, illnesses; e.g.: suicide prevention (Eskisabel-Azpiazu, Cerezo-Menedez & Gayo-Avello, 2017); Griefsquatting (Klastrup, 2017); Digital Death (Gotved, 2014; Lagervist, 2013)

Additional considerations have been suggested, including specific attention to minorities, LGBT individuals and/or communities.

A particular set of concerns here are evoked when researchers encounter information suggesting that their subjects may be engaged in behaviour threatening to their own well-being, e.g., a researcher studying bloggers describing self-cutting (Stern, 2004) or manifesting ever greater focus on suicide (Seko, 2006; see discussion, McKee & Porter, 2008, p. 15 f., p. 72 ff., pp. 88-90, pp. 95-96). Relatedly, researchers have long been confronted with the possibility of discovering information suggesting the potential for committing crimes. Very broadly, there is often a professional and/or legal obligation for researchers to report such information to the relevant authorities: but this varies widely, first of all, depending on local and national legislation, specific professional ethical guidelines, and/or specific policies that may apply within a given research context. In this direction, what are researchers' obligations – if any, and if so, under what circumstances – to report such potential threats to the platforms in which they appear? [21]

As well, a cluster of issues are certain to emerge in conjunction with the rise of AI as deployed, e.g., in chatbots and social robots, along with the burgeoning development of innumerable "smart" devices incorporated into the emerging Internet of Things (IoT). There is already an established and growing literature on the moral status and potential rights of robots and AIs as they become increasingly autonomous (as an early and now standard

---

[21] Our thanks to William Wolff for raising these issues.

reference, see Coeckelbergh, 2010). And human responses to inflicting (apparent) harm on virtual agents (e.g., in efforts to replicate the [in]famous Milgram experiments [Slater et al., 2006]) make clear that many humans *feel* that to do so is ethically problematic (this is not surprising, given the now well-documented phenomenon of human beings attributing feelings to and thereby obligations to *care for* robotic devices, beginning with the very simple Tamagotchi (Turkle, 2011). As these devices become more and more implicated in research, whether as quasi-subjects (e.g., as surrogate caregivers (e.g., Kaiko, Shuichi, & Shinichi, 2016)) and/or as co-researchers and/or as the objects of research, attempting to establish their moral status (if any – and if so, what?) and concomitant ethical implications and obligations towards them will become more and more important.[22] In these directions, the emerging field of Human-Machine Communication (HMC) will develop critical insights from the multiple disciplines involved, including robotics, media and communication, computer science, psychology, applied ethics, and so on (Jones, 2014; Guzman, 2018; Ess, 2018).

3.2.6 Conceptual Issues: Ethical Frameworks and Concepts

As with IRE 1.0 and 2.0, a wide range of diverse ethical frameworks and concepts can be fruitfully brought to bear upon issues emerging within IRE 3.0. i.e., as helpful conceptual tools for analysing and, ideally, resolving ethical conflicts. These begin with *utilitarianism* and *deontology*, alongside *virtue ethics*, *feminist ethics*, and (feminist) *ethics of care*, which have become much more prominent in recent decades.[23] These large frameworks are frequently tuned specifically to internet research issues: so our Companion Resources include: **6.3 aline shakti franzke, Feminist Research Ethics** (explicitly taking up feminist ethics and ethics of care) and **6.4 Annette Markham, An “Impact Model” for Ethical Assessment** (which takes a primarily utilitarian approach).

Broadly, these frameworks address such foundational ethical norms and commitments as autonomy and freedom (especially in deontological ethics), along with the basic elements of IRE, such as informed consent and confidentiality. AoIR members have also pointed to specific ethical topics such as *accountability*, *trust*, and *transparency* as also critical to good research practice and ethics – though not always easy to define or apply in *praxis*. Similarly, AoIR researchers often highlight *responsibility* – towards oneself, one’s institution, and the larger society, including returning some sort of benefit back to the communities under study. How researchers will understand and practice such responsibility will depend, to begin with, on basic assumptions of selfhood (more individual vis-à-vis more relational) as well the specific ethical frameworks they take up: as noted early on, feminist ethics and ethics of care – as applied, for example, within a participant-observation methodology – characteristically leads to a greater sense of ethical responsibility to protect informants’ privacy, confidentiality, and so on (Hall, Frederick & Johns, 2003).

Numerous specific ethical components of IRE have also been carefully explored, specifically with a view towards their application in *praxis*. These include: *anonymity* (Robson, 2017); *privacy* (whether social privacy, horizontal privacy, and/or group privacy:

---

[22] Our thanks to Sarah Quinton for pointing this out.

[23] Cf. the initial discussion of these frameworks, example applications in IRE, and suggested resources for further reading above, **2.3.1 A Range of Ethical Frameworks**.

Matzner & Ochs, 2017); *justice* (Hoffmann & Jonas, 2017); *accuracy* (Puschmann, 2017); and *bias* (Tromble & Stockmann, 2017).

3.2.7 Assumptions, Questions, Issues, Procedures for Data Acquisition, Analysis, Storage, Dissemination [24]

IRE 2.0 included the following questions:

- How are data being managed, stored, and represented?
- What method is being used to secure and manage potentially sensitive data?
- What unanticipated breaches might occur during or after the collection and storage of data or the production of reports? (For example, if an audience member recorded and posted sensitive material presented during an in-house research presentation, what harms might result?)
- If the researcher is required to deposit research data into a repository for future use by other researchers (or wishes to do so), what potential risks might arise? What steps should be taken to ensure adequate anonymity of data or to unlink this data from individuals?
- What are the potential ethical consequences of stripping data of personally identifiable information?
- How might the removal of selected information from a dataset distort it such that it no longer represents what it was intended to represent?
- If future technologies (such as automated textual analysis or facial recognition software) make it impossible to strip personally identifiable information from data sets in repositories, what potential risks might arise for individuals? Can this be addressed by the original researcher? If so, how? How will this impact subsequent researchers and their data management? (IRE 2.0, p. 9 f.)

These questions emphasize the hard fact that the best current practices and techniques can only "de-identify" data, i.e., not perfectly anonymize data. Therefore, emphasis on data security is even more important and **IRE 3.0** stresses the importance of continuing this line of questioning.

Questions of definitions, understandings of what data is/represents:

Hacked/stolen Data (Poor, 2017): when is it allowable (if ever) to use data that would otherwise be prohibited ethically and/or legally because of privacy protections, etc. – but has been made public because of an accidental breach and/or intentional hack (e.g., Ashley Madison)? (For example, those using a deontological perspective, the danger that use of such data might be harmful to persons involved would almost invariably preclude its use. However, those from a more utilitarian perspective might seek to weigh any possible harm against the possible benefits that could result from analysing data that would otherwise be unavailable.)[25]

---

[24] For further exploration of ethical research practices regarding automated data extraction see e.g., Alim (2014).

[25] Our thanks to Mark Johns for suggesting this elaboration.

Issues about the accuracy of data research – how far can we rely on data provided from a commercial provider? What are the possibilities of in-built biases, etc. in algorithms used for collection and analysis? The need to be aware of the problematic representativeness of even large datasets of social media users as compared to overall populations etc.

Differences between different types of data (aggregated sets, derived sets, purchased sets: see Kitchin & McArdle, 2016; **6.2 Elisabetta Locatelli, Academy/Industry Partnership and Corporate Data: Ethical Considerations**)

Downloaded data (Barbosa & Milan, 2019): Does the type and amount of (meta)data together potentially disclose identity, and if so, what measures have been taken in order to project human subjects? **Issues related to Metadata** (Humphreys, 2017)

Issues and procedures in collecting, managing, storing, and destroying data:

Data minimization: When is data enough for research purpose? The rule of data minimization in, for instance, the GDPR (2018), is to some extent conflicting with the very purpose of big data, namely, to query the data for answers within the scope of an overall research interest or research question with the aim to learn inductively. This is a completely different approach than within 'normal' quantitative approaches where hypotheses are driving the research. In order to make the full benefit of big data methods the researcher often find themselves in a dilemma of wanting to 'ask' the data across time, different data points or a large volume of/diverse set of IDs or data subjects, yet wanting to fulfill the need for data minimization (see also: Ess & Hård af Segerstad, 2019; Bechmann & Kim, 2020).

Data storage vs. the (EU) right to be forgotten (Tromble & Stockmann in Zimmer & Kinder-Kurlanda, 2017; Geiger et al., 2018)

Deleting data – e.g., when in the life of a research project, and how?

Databases, storing, archiving: A long history of web and internet archiving has raised many questions relevant for IRE 3.0, including questions such as: What data are stored as relevant and what are discarded as irrelevant? What happens to archives when data storage formats become obsolete? What factors determine, or later alter, the searchability of databases? What decoding information/formula is (or should be) stored alongside data to ensure future readability? (cf. Kinder-Kurlanda, Jessica Odgeden, 2017; Agostinho, 2016; Thylstrup, 2019; Markham & Pereira, forthcoming)

Data sharing: Even if not legally prohibited, what are the levels of impacts possible from sharing large datasets? Can one be sure these are adequately anonymized (e.g., the OKCupid data release case where supposedly anonymized profiles were easily re-linked to persons)? (cf. Responses from Zimmer, 2016; Markham, 2016 and other overviews from Weller and Kinder-Kurlanda in Zimmer & Kinder-Kurlanda, 2017). However, social media big data analysis relies on platforms to share data to some extent with the research community and/or to allow researchers to collect data through APIs. This has made the topic of data sharing and the balance between transparency and privacy a top political priority in Europe, for instance (Moller & Bechmann, 2019). The strategy in current data exchange solutions is to secure privacy on the level of both who gets access (data grants) and how many data points are provided access to (Dwork, 2006). Yet, both solutions also entail profound ethical problems of scientific freedom and favour large countries over smaller countries, famous researchers and universities over less famous ones, and thus hamper the creation of a community around such research. Working on different safe space approaches is therefore a high priority in the internet research community now (Moller & Bechmann, 2019; Bruns, 2019).

Additional considerations include: what are the primary purposes of the collected data that you are using? Is it possible that the original context influences/interferes with your own results? In which sense do you feel that all aspects of the original context can be or to ought to be considered when reusing already stored datasets? (cf. **7.1 Keith Douglas, Operational Security: Central Considerations**)

3.2.8 Algorithms and Learning Models

How can we make sure that our models and subsequent results are adequately documented in order to replicate our results and in order to represent our data subjects? Two fundamental issues arise when dealing with big data and subsequent learning methods. First, how do we make sure that our data subjects are adequately represented? Here, questions of privacy are of course an issue, but also how we choose to classify our data subjects in, for instance, different genders? Or: How we clean our data and leave out some information over others thereby increasing the likelihood of concluding something over others (Bechmann & Bowker, 2019). Second, how do we make sure that it is possible for the community to actually review the study made? This is especially important due to the increase of misinformation in circulation in the intersection between popular dissemination and research results.

Learning models are often divided into supervised and unsupervised models. Supervised models are controlled (or semi-controlled) models where the researcher feed the models with, for instance, classifiers or labelled material that the model can learn from: by contrast, unsupervised models try to find clusters in the data from a more inductive approach, even though one could claim that classification takes place in other forms (for a more extensive discussion, see Bechmann & Bowker, 2019). The many layers and iterations of cleaning/pre-processing, classification, training/retraining in order to predict a trend, tendency or pattern require new standards for documentation in the publication outlets that has yet to catch up with the actual research. This is also the case with sensitive data or other personal data made available only for the research team in question e.g. from Social Science One (Facebook data).

Other strands of research within internet research use reverse engineering in order to understand closed algorithms e.g. how algorithms react on privacy sensitive settings (e.g. data disclosure), on potential systemic discrimination (e.g. advertising interface) or public sensitive matters (e.g. election campaigns). Here, only part of the data is available and scraping might be an issue (Karpf, 2012; Sandvig, Hamilton, Karahalios & Langbort, 2016). Also, the algorithm is protected by intellectual property rights even though especially European regulation has recently tried to create mandatory insights into the logics of the algorithm. Therefore, researchers may find themselves in an ethical dilemma in wanting to disclose societal problems on the one side, and needing to protect themselves against lawsuits and platform identity 'attacks'. This may require them to unite in larger international teams, for instance, that can collaborate with (and place pressure on) internet industry players in a different way.

3.2.9 Other Concerns

To be sure, the list of ethically-relevant issues, challenges, and resources continues – and certainly will continue – to expand. For example, there is the category of *ethically relevant institutional and commercial issues* that researchers often must also face. These include broad issues of conflicting interests between researchers (who are supported by their institution and/or specific funding bodies) and involved third parties (including participants who supply data and/or commercial data providers, who may set restrictions on who is allowed to access purchased data). These questions point specifically to concerns over *data ownership*, for example, in projects involving collaboration between two or more institutions: if a team of researchers from more than one institution undertake data-collection processes – where are these data to be stored (securely), and thereby who can claim ownership of the data, especially if it is preserved for subsequent use? Similar – and additional – concerns are raised as public and/or national research committees increasingly press for Open Access publication and data sharing.

Other specific issues related to research funding and finances have received some attention. For example, the prevalent use of crowdsourcing platforms such as Amazon Mechanical Turk for both online data collection (i.e., distributing surveys for completion) as well as for data processing (i.e., having Amazon Turk users assist in processing raw research data) spur concerns over whether such crowdsourcing workers are being paid a living wage (Pittman & Sheehan in Zimmer & Kinder-Kurlanda, 2017, p. 213 ff.). Furthermore, as Amazon Turk and similar crowdsourcing commercial initiatives are paid per unit, and thus create incentives to label as many units per hour as possible, there are also ethical concerns regarding the labelling quality and as to what extent documentation is needed in order to account for not only representation, sampling (of Turks) but also what incentives might generate of issues in the research findings and how researchers have tried to account/adjust for these.

Last but certainly not least: we need to elaborate an ethics addressing the distinctive issues clustering around the production, sharing, and thereby research on visual images. This is in part as visual images may evoke stronger emotional senses of investment and impact both among those who develop and share especially more personal or dramatic images (e.g., political protest, police violence, terrorist attacks, events in war, etc.) as well as among their audience. Broadly, greater impact entails greater care and reflection for the sake of respecting (possible) expectations of privacy, avoiding harm, and so on. As well, there is a growing series of anecdotes of researchers in diverse disciplines who include examples of visual data to illustrate and undergird their research in conferences – e.g., social media profile photos, specific images shared on sensitive topics such as domestic abuse and other forms of violence, racism, and so on: as occurring within a restricted professional audience, researchers may assume that the ethical considerations that would apply for publication do not hold. But as it has become increasingly common to photograph or livestream such presentations for the sake of sharing on social media – the conference venue increasingly resembles a form of more public publication of these materials, thereby increasing the possibilities of various harms.

These sorts of issues are certain to become increasingly common; but to our knowledge, there is little established literature or sense of best practices that can be referred to. Similar comments hold, for example, regarding how we can best resolve ethical conflicts and dilemmas through deliberative processes. Hence these – and still further issues as they emerge – should be on the agenda for further discussion and reflection in an IRE 3.1, 3.2, and so on.

4. Concluding Comments

Internet Research Ethics (IRE) 3.0 was driven by on-going changes and developments in the technological, legal, and ethical contexts that shape internet research. While IRE 1.0 and 2.0 provided grounded ethical approaches to assist internet researchers with sets of then-emerging problems, IRE 3.0's focus is centered more on the ethical challenges that emerge across different stages of research, and the growing challenge of informed consent in data-intensive research protocols. Like its predecessors, IRE 3.0 aims to provide a general structure for ethical analysis, designed to help identify the ethically relevant issues and questions, supplemented with additional suggestions for how to begin to analyse and address these challenges related to ongoing socio-technical developments in the domain of internet research.

As Zimmer and Kinder-Kurlanda (2017) note, "ethically-informed research practices come out of processes of deliberation and decision making under great uncertainty, which often may go wrong or seemingly force us towards less-ideal options" (p. 303). While solutions to the ethical challenges faced by internet researchers might not always be obvious, easy, or even ideal, IRE 3.0 provides researchers with an enhanced framework to engage in the necessary "processes of deliberation" to empower ethically informed research practices.

Lastly, as noted from the outset of IRE 1.0 and throughout the development and unfolding of IRE 2.0 and 3.0 – a central component in these *reflective, dialogical* process-oriented approaches is precisely the catalysation of our capacities for reflective ethical *judgment* (*phronēsis*: see **2.2.**, footnote. 3; **3.2.1.,** and the NESH guidelines' emphasis on developing "sound judgement" [2018, p. 4]). Such judgment is core to our ethical decision-making in the face of the difficult challenges we face as researchers, participants, oversight authorities and/or larger stakeholders. It is honed through experience (what turn out to be either effective judgments and/or misjudgments that can sometimes be corrected or at least learned from), but is also a capacity that can be enhanced and cultivated. Perhaps most importantly: such judgment is deeply *relational* and *intersubjective*. As such, it requires continuous cross-checking, mutual critique, and/or corroboration and expansion – precisely through ongoing dialogue and discussion with peers and trusted colleagues. In this direction, one of the most important ethical techniques to be recommended is one of the simplest: talk things over with colleagues and friends. Any number of researchers – as well as anyone else struggling to resolve a difficult ethical challenge – will witness to how essential and helpful such discussions have been in their own ethical reflection, choices, and development.

Finally, as ethical challenges will continue to multiply much more rapidly than our capacities to develop considered guidelines, it seems clear that researchers and oversight boards are going to increasingly be thrown back on their own judgments and thus their own responsibility for those judgments. Hence it would seem prudent (another term for *phronēsis*) to cultivate our ethical judgments precisely as researchers, rather than presume (positivistically) that ethics and judgment are just for professional philosophers to worry about.

We hope that the resources gathered in IRE 3.0 – including frameworks and their applications in practice, as well as pointers to further references in the relevant literatures – will be helpful tools for researchers and oversight boards who, most likely, will be increasingly responsible for both the ethical judgments they draw and communicating them to other researchers, research participants, and the broader community.

Acknowledgements

In addition to the numerous contributors to AoIR ethics panels and this document over the past three years, we are especially indebted to:

> The DATALAB, Center for Digital Social Research, Aarhus University (Aarhus, Denmark), for financial and administrative support of workshops focusing on the further development of the IRE 3.0 processes and documents, including Anne Hove Henriksen for research and administrative assistance, and Stine Riis Ebbesen for final reference checking, formatting, and pagination.
>
> The Department of Media and Communication, University of Oslo, for financial and administrative support of IRE 3.0 workshops, as well as for a visiting research fellowship for aline shakti franzke in 2017**.** In particular, Niamh Ní Bhroin was central in the planning and implementation of two workshops on IRE that directly enriched the content and development of this document.
>
> Anna Jobin (ETH, Zürich) has made extraordinary contributions, especially with the development and maintenance of our digital support resources, including our Slack workspace and the Zotero bibliographic database.
>
> Our many colleagues affiliated with NESH (The Norwegian National Committee for Research Ethics in the Social Sciences and the Humanities) and its continued revision and expansion of the NESH internet research guidelines – most especially Vidar Enebakk and Elisabeth Staksrud – are due particular thanks for their innumerable contributions to and ongoing support of the AoIR IRE 3.0 project.

5. References

Agostinho, D. (2016). ‘Big data, time, and the archive’, *Symploke*,, Volume 24,(Numbers 1-2), 2016, pp. 435-445.

Alexander, L. & Moore, M. (2016). Deontological Ethics. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.). https://plato.stanford.edu/archives/win2016/entries/ethics-deontological/

Alim, S. (2014). An initial exploration of ethical research practices regarding automated data extraction from online social media user profiles. *First Monday, 19*(7). Doi: http://dx.doi.org/10.5210/fm.v19i7.5382.

Barbosa, S. & Milan, S. (2019). Do Not Harm in Private Chat Apps: Ethical Issues for Research on and with WhatsApp. Westminster Papers in Communication and Culture, 14(1), 49-65.

Barnes, N. (2019). Personal communication. Email to C. M. Ess, October 10, 2019.

Bassett, E. H. & O’Riordan, K. (2002). Ethics of Internet Research: Contesting the Human Subjects Research Model. *Ethics and Information Technology, 4* (3), 233-249.DOI: https://doi.org/10.16997/wpcc.313

Bechmann, A. & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society 6*(1 January–June), 1–11). https://doi.org/10.1177/2053951718819569

Bechmann, A. & Kim, J.Y. (in press). Big Data: A Focus on Social Media Research Dilemmas. In R. Iphofen (ed.), *Handbook of Research Ethics and Scientific Integrity.* Berlin: Springer.

Beninger, K., Fry, A., Jago, N., Lepps, H., Nass, L., and Silvester, H. (2014). Research using Social Media; Users’ Views. London: National Centre for Research Methods. Available online: http://www.natcen.ac.uk/our-research/research/research-using-social-media-users-views/ [June 14, 2019].

British Psychological Society (2017). Ethics Guidelines for Internet-mediated Research. INF206/04.2017. Leicester. www.bps.org.uk/publications/policy-and-guidelines/ research-guidelines-policy-documents/research- guidelines-poli

Bruns, A. (2018). Facebook shuts the gate after the horse has bolted, and hurts real research in the process. *Internet Policy Review*, April 25, 2018. https://policyreview.info/articles/news/facebook-shuts-gate-after-horse-has-bolted-and-hurts-real-research-process/786

Bruns, A. (2019). After the ‘APIcalypse’: social media platforms and their fight against critical scholarly research, *Information, Communication & Society*, 1544-1566. DOI: https://doi.org/10.1080/1369118X.2019.163744710.1080/1369118X.2019.1637447

Buchanan, E. (2011). Internet Research Ethics: Past, Present, Future. In M. Consalvo and C. Ess (eds.), *The Blackwell Handbook of Internet Studies*, (C. Ess and M. Consalvo, (eds.), pp. 83-108). Oxford: Wiley-Blackwell.

Buchanan, E. & Ess, C. (2008). Internet Research Ethics, in K. Himma and H. Tavani (eds.), *The Handbook of Information and Computer Ethics*, (pp. 273-292). Hoboken, NJ: John Wiley & Sons, 2008.

Christman, J. (2004). Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 117* (1/2), : 143-164.

Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology 12* (3),: 209-221.

Denison, T., & Stillman, L. (2012). Academic and ethical challenges in participatory models of community research. *Information, Communication & Society, 15*(7), pp. 1037-1054. Doi: https://doi.org/10.1080/1369118X.2012.65613810.1080/1369118X.2012.656138.

Domenget, J.-C., & Wilhelm, C. (2018). « L'éthique des pratiques de recherche liées au numérique en sic: le rôle de la société française pour les sciences de l'information et de la communication ». In G. Chartron et al. (eds.) *L'éthique en 26ontext info-communicationnel numérique,* (pp. 101-112). Deboeck 2017, 101-112.

Dwork, C. (2006). Differential Privacy. *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II* (pp. 1–12). https://doi.org/10.1007/11787006_1

Enebakk, V. (2018). The new NESH guidelines: highlights, examples and important lessons. Presentation, "Digital Research: Contemporary Methods and Ethics." Oslo, Norway, Dec. 1, 2018.

Eskisabel-Azpiazu, A., Cerezo-Menéndez, R. & Gayo-Avello, D. (2017). An Ethical Inquiry into Youth Suicide Prevention Using Social Media Mining. In M Zimmer & K- Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts* (pp. (227-234)). Berlin: Peter Lang.

Ess, C. (2013). *Digital Media Ethics, 2nd Edition*. Cambridge: Polity Press.

Ess, C. (2014). Selfhood, moral agency, and the good life in mediatized worlds? Perspectives from Medium Theory and philosophy. In Knut Lundby (ed.), *Mediatization of Communication* (vol. 21, *Handbook of Communication Science*), (pp. 617-640). Berlin: De Gruyter Mouton, 2014.

Ess, C. (2017). Grounding Internet Research Ethics 3.0: A view from (the) AoIR (Foreword). In M. Zimmer & K- Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts* ((pp. ix-xv)). Berlin: Peter Lang.

Ess, C. (2018). Ethics in HMC: Recent Developments and Case Studies. In A. L. Guzman (ed.), *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves* (pp. 237-257). Berlin: Peter Lang.

Ess, C. and the AoIR ethics working committee. (2002). Ethical decision-making and Internet research: Recommendations from the aoir ethics working committee. http://www.aoir.org/reports/ethics.pdf

Ess, C., & Hård af Segerstad, Y. (2019). Everything Old is New Again: the Ethics of Digital Inquiry and its Design. In Å. Mäkitalo, T. E. Nicewonger & M. Elam (eds.), *Designs for Experimentation and Inquiry: Approaching Learning and Knowing in Digital Transformation* (pp. 179-196). London: Routledge.

GDPR. (2018). General Data Protection Regulation Regulation EU 2016/679. Approved 27 April 2016, implemented May 25 2018. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679.

Geiger, R. S., Sholler, D., Culich, A., Martinez, C., Hoces de la Guardia, F., Lanusse, F., Ottoboni, K, Stuart, M., Vareth, M., Varoquaux, N., Stoudt, S., & van der Walt, S. (2018). Challenges of Doing Data-Intensive Research in Teams, Labs, and Groups." BIDS Best Practices in Data Science Series. Berkeley Institute for Data Science: Berkeley, California. Doi: 10.31235/osf.io/a7b3m.

Gotved, S. (2014). Death online – alive and kicking! *Thanatos 3*(1), 112–26.

Guzman, A. L. (ed.) (2018). *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves*. Berlin: Peter Lang

Halavais, A. (2019). Overcoming terms of service: a proposal for ethical distributed research. *Information, Communication & Society*. https://doi.org/10.1080/1369118X.2019.1627386

Hall, G. J., Frederick, D. & Johns, M.D. (2003). "NEED HELP ASAP!!!" A Feminist Communitarian Approach to Online Research Ethics, in M. Johns, S.L. Chen & J. Hall (eds.), *Online Social Research: Methods, Issues, and Ethics* (pp. 239-252). New York: Peter Lang

Hård af Segerstad, Y., Kasperowski, D., Kullenberg, C. & Howes, C. (2017). Studying Closed Communities On-line: Digital Methods and Ethical Considerations beyond Informed Consent and Anonymity. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 213-226). Berlin: Peter Lang.

Heise, N. & Schmidt, J.-H. (2014). Ethik der Onlineforschung. In M. Welker, M. Taddicken, J.-H. Schmidt & N. Jackob (eds.), *Handbuch Online-Forschung* (pp. 519–539). Köln: Herbert von Halem Verlag.

Hoffmann, A. L. & Jonas, A. (2017). Recasting Justice for Internet and Online Industry Research Ethics. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 3-18). Berlin: Peter Lang.

Honan, E. (2014). Disrupting the habit of interviewing. *Reconceptualizing Educational Research Methodology, 5*(1), 1-9. http://journals.hioa.no/index.php/rerm

Hudson, J. M., and Bruckman, A. (2004). "Go Away": Participant Objections to Being Studied and the Ethics of Chatroom Research. *The Information Society*, *20*, 127–139.

Humphreys, L. (2017). Locating Locational Data in Mobile and Social Media. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 245-254). Berlin: Peter Lang.

Hunsinger, J. (2019) Personal communication. Email to C. M. Ess, September 23, 2019.

Hursthouse, R. and Pettigrove, G. (2018). Virtue Ethics. *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/archives/win2018/entries/ethics-virtue/

Jackson, D., Aldrovandi, C. & Hayes, P. (2015). Ethical framework for a disaster management decision support system which harvests social media data on a large scale. In N. Bellamine Ben Saoud et al. (Eds.), *ISCRAM-med 2015* (pp. 167–180), LNBIP 233. Doi:10.1007/978-3-319-24399-3_15.

Jones, S. (2019). Personal communication. Email to C. M. Ess, October 10, 2019.

Jones. S. (2014). People, things, memory and human-machine communication. *International Journal of Media and Cultural Politics, 10*(3), 245-258. Available online at http://www.intellectbooks.co.uk/journals/view-Article,id=19117/

Kaiko, K., Shuichi, N. & Shinichi, S. (2016). Can We Talk through a Robot As if Face-to-Face? Long-Term Fieldwork Using Teleoperated Robot for Seniors with Alzheimer's Disease. *Frontiers in Psychology 7* (July), article 1066, pp. 1-13. Doi: https://doi.org/10.3389/fpsyg.2016.0106610.3389/fpsyg.2016.01066

Karpf, D. (2012). Social Science Research Methods in Internet Time. *Information, Communication & Society, 15*(5), 639-661. https://doi.org/10.1080/1369118X.2012.665468

Kaufmann, K. (2019). Mobile Methods: Doing Migration Research with the Help of Smartphones. In K Smet, K Leurs, M Georgiou, S Witteborn & R Gajjala (eds.), *The SAGE Handbook of Media and Migration* (pp. 167-179). London: Sage.

Kitchin, R. & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951716631130, 2053951716631130.

Klastrup, L. (2017). Death, Affect and the Ethical Challenges of Outing a Griefsquatter. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 235-243). Berlin: Peter Lang.

Kotsios, A., Magnani, M., Rossi, L., Shklovski, I. & Vega, D. (2019). An Analysis of the Consequences of the General Data Protection Regulation (GDPR) on Social Network Research, 1–23. Retrieved from http://arxiv.org/abs/1903.03196

Lomborg, S. & Bechmann, A. (2014). Using APIs for Data Collection on Social Media. *The Information Society 30* (4), 256-265

Luka, M. E., Millette, M. & Wallace, J. (2017). A Feminist Perspective on Ethical Digital Methods. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 21-36). Berlin: Peter Lang.

Markham, A. (2006). Method as ethic, ethic as method. *Journal of Information Ethics*, *15*(2), 37–54.

Markham, A. (2012). Fabrication as ethical practice: Qualitative inquiry in ambiguous internet contexts. *Information, Communication and Society*, *15*(3), 334-353. doi: 10.1080/1369118X.2011.641993

Markham, A. (2016). OKCupid data release fiasco: It's time to rethink ethics education. *Data & Society Points* (, May 18, 2016)). Retrieved from: https://points.datasociety.net/okcupid-data-release-fiasco-ba0388348cd

Markham, A. (2018). Afterword: Ethics as Impact—Moving from Error-Avoidance and Concept Driven Models to a Future-Oriented Approach. *Social Media + Society*, *4*(3). https://doi.org/10.1177/2056305118784504DOI: 10.1177/2056305118784504

Markham, A. & Buchanan, E. (2012). Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0). http://www.aoir.org/reports/ethics2.pdf

Markham, A. & Buchanan, E. (2017). Ethical Concerns in Internet Research. In J. Wright, J. (ed.), *The International Encyclopedia of Social and Behavioral Sciences*, 2nd Edition (pp. 606-613). Amsterdam: Elsevier Elsivier Press.

Markham, A. & Pereira, G. (forthcoming). Experimenting with algorithmic memory-making: Lived experience and future-oriented ethics in critical data science. *Frontiers in Big Data Science*.

Markham, A., Tiidenberg, K. & Herman, A. (2018). Ethics as Methods: Doing Ethics in the Era of Big Data Research - Introduction. *Social Media + Society, 4* (3). https://doi.org/10.1177/2056305118784502

Massanari, A. (2016). When the researcher becomes the target: harassment, doxxing, and death threats – what to do? Association of Internet Researchers (AoIR) annual conference, October 6, 2016. Berlin, Germany.

Matzner, T. & Ochs, C. (2017). Sorting Things Out Ethically: Privacy as a Research Issue beyond the Individual. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 39-52). Berlin: Peter Lang.

McKee, H. A. & Porter, J. E. (2009). *The Ethics of Internet Research: A Rhetorical, Case-based Process*. New York: Peter Lang Publishing.

Moller, L. A. & Bechmann, A. (2019). *Research Data Exchange Solution*, SOMA report for the EU Commission, Brussels.: https://www.disinfobservatory.org/download/26541

Mukherjee, I. (2017). Case Study of Ethical and Privacy Concerns in a Digital Ethnography of South Asian Blogs against Intimate Partner Violence. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 203-212). Berlin: Peter Lang.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC.

Nature. (2019). Time to discuss consent in digital-data studies. Editorial, *Nature* (31 July). https://www.nature.com/articles/d41586-019-02322-z

NESH (The [Norwegian] National Committee for Research Ethics in the Social Sciences and the Humanities) (2018) *A Guide to Internet Research Ethics*. Oslo: NESH. [URL: https://www.etikkom.no/globalassets/documents/publikasjoner-som-pdf/forskningsetisk-veileder-for-internettforskning/guide-to-internet-research-ethics-2018-print.pdf

Neuhaus, F. & Webmoor, T. (2012). Agile Ethics for Massified Research and Visualization. *Information, Communication and Society, 15* (1),: 43-65. https://doi.org/10.1080/1369118X.2011.616519doi: 10.1080/1369118X.2011.616519

Obar, J. A. (2015). Big Data and The Phantom Public: Walter Lippmann and the fallacy of data privacy self-management. *Big Data & Society*. https://doi.org/10.1177/2053951715608876

Patterson, A. N. (2018). YouTube Generated Video Clips as Qualitative Research Data: One Researcher's Reflections on the Process. *Qualitative Inquiry, 24*(10), 759–767. https://doi.org/10.1177/1077800418788107.

Pentzold, C. (2015). Forschungsethische Prämissen und Problemfelder teilnehmenden Beobachtens auf Online-Plattformen. In A. Maireder, J. Ausserhofer, C. Schumann, & M. Taddicken (Eds), *Digitale Methoden in der Kommunikationswissenschaft*. Berlin (Digital Communication Research 2) (pp. 61-85). http://dx.doi.org/10.17174/dcr.v2.4

Pittman, M. & Sheehan, K. (2017). The Case of Amazon's Mechanical Turk. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 177-186). Berlin: Peter Lang.

Poor, N. (2017). The Ethics of Using Hacked Data: Patreon's Data Hack and Academic Data Standards. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 278-280). Berlin: Peter Lang.

Puschmann, C. (2017). Bad Judgment, Bad Ethics? Validity in Computational Social Media Research. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 95-113). Berlin: Peter Lang.

Puschmann, C. (2019). An end to the wild west of social media research: a response to Axel Bruns, *Information, Communication & Society,* 1582-1589. https://doi.org/10.1080/1369118X.2019.1646300DOI: 10.1080/1369118X.2019.1646300

Reilly, P., & Trevisan, F. (2016). Researching protest on Facebook: developing an ethical stance for the study of Northern Irish flag protest pages. *Information Communication and Society*, *19* (3), 419-435. https://doi.org/10.1080/1369118X.2015.1104373

Rensfeldt, A. B., Hillman, T., Lantz-Andersson, A., Lundin, M. & Peterson, L. (2019) A "Situated Ethics" for Researching Teacher Professionals' Emerging Facebook Group Discussions. In A Mäkitalo, T E Nicewonger & M Elam (eds.), *Designs for Experimentation and Inquiry: Approaching Learning and Knowing in Digital Transformation* (pp. 197-213). London: Routledge.

Robards, B. (2013). Friending Participants: Managing the Researcher-Participant Relationship on Social Network Sites. *Young 21*(3), 217–235. https://doi.org/10.1177/1103308813488815doi: 10.1177/1103308813488815

Robson, J. (2017). Situating the Researcher within Digital Ethnography. In M. Zimmer & K. Kinder-Kurlanda (eds.) *Internet Research Ethics for the Social* Age (pp. 195-202). Berlin: Peter Lang.

Sandvig, C., Hamilton, K., Karahalios, K. & Langbort, C. (2016). Automation, Algorithms, and Politics | When the Algorithm Itself is a Racist: Diagnosing Ethical Harm in the Basic Components of Software. *International Journal of Communication*, *10*(0), 19.

Seko, Y. (2006). Analyzing online suicidal murmurs. (AoIR) Internet Research 7.0: Internet Convergences. September 27-30, 2006, Brisbane, Australia.

Seko, Y. & Lewis, S. P. (2017). "We Tend to Err on the Side of Caution": Ethical Challenges Facing Canadian Research Ethics Boards when Overseeing Internet Research. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social* Age (pp. 133-147). Berlin: Peter Lang.

Sinnott-Armstrong, W. (2019). Consequentialism. *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2019/entries/consequentialism

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N. & Sanchez-Vives, M. V. (2006). A Virtual Reprise of the Stanley Milgram Obedience Experiments. *PLoS One 1*(1): e39. https://doi.org/10.1371/journal.pone.0000039doi: 10.1371/journal.pone.0000039

Stern, S. R. (2004). Studying adolescents online: A consideration of ethical issues. In Elizabeth A. Buchanan (Ed.), *Readings in virtual research ethics: Issues and controversies* (pp. 274-287). Hershey, PA: Information Science.

Thylstrup, N. (2019). *The politics of mass digitization*. Cambridge, MA: MIT Press.

Tiidenberg, K. (2018). Ethics in digital research. In U. Flick (Ed.) *The Sage Handbook of Qualitative Data Collection* (pp. 466-479). London: SAGE. doi: 10.4135/9781526416070.n30

Trevisan, F. & Reilly, P. (2014). Ethical dilemmas in researching sensitive issues online: lessons from the study of British disability dissent networks. *Information, Communication & Society, 17*(9), 1131-1146. https://doi.org/10.1080/1369118X.2014.889188.

Tromble, R. & Stockmann, D. (2017). Lost Umbrellas: Bias and the Right to be Forgotten in Social Media Research. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts* (pp. 75-91). Berlin: Peter Lang.

Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other.* New York: Basic Books.

Van Schie, G., Westra, I., & Schäfer, M. T. (2017). Get Your Hands Dirty. Emerging Data Practices as Challenge for Research Integrity. In: M. T. Schäfer & K. Van Es (eds.), *The Datafied Society. Studying Culture through Data* (pp. 183-200). Amsterdam: Amsterdam University Press.

Veltman, A. & Piper, M. (2014). Introduction. In A. Veltman & M. Piper (eds.) *Autonomy, Oppression and Gender* (pp. 1-11). Oxford: OUP.

VOX-Pol (VOX-Pol Project on Violent Online Political Extremism). (2018). *The Ethics of Terrorism Research*. Workshop, April 25-26, 2018. Swansea University, Wales, UK.

Warfield, K., Hoholuk, J., Vincent, B. & Camargo, A. D. (2019). Pics, Dicks, Tits, and Tats: negotiating ethics working with images of bodies in social media research. *New Media & Society*. https://doi.org/10.1177/1461444819837715.

Westlund, A. (2009). Rethinking Relational Autonomy. *Hypatia, 24* (4: Fall), 26-49.

Whelan, A. (2018). Ethics Are Admin: Australian Human Research Ethics Review Forms as (Un)Ethical Actors. *Social Media + Society* (April-June 2018), 1–9. https://doi.org/10.1177/2056305118768815

Zetter, K. (2016). Researchers Sue the Government over Computer Hacking Law. *Wired* (June 29). https://www.wired.com/2016/06/researchers-sue-government-computer-hacking-law/

Zevenbergen, B., Mittelstadt, B., Véliz, C., Detweiler, C., Cath, C., Savulescu, J. & Whittaker, M. (2015). *Philosophy meets Internet engineering: Ethics in networked systems research*. (GTC workshop outcomes paper). Oxford Internet Institute, University of Oxford. Retrieved from http://ensr.oii.ox.ac.uk/wp-content/uploads/sites/41/2015/09/ENSR-Oxford-Workshop-report.pdf.

Zimmer, M. (2016). OKCupid Study Reveals the Perils of Big-Data Science. *Wired*. May 14, 2016. Retrieved from: https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/

Zimmer, M. & Kinder-Kurlanda, K. (eds.) (2017). *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. Berlin: Peter Lang.

Zook M., Barocas, S., boyd, d., Crawford, K., Keller, E., Gangadharan, S.P. et al. (2017). Ten simple rules for responsible big data research. *PLoS Computational Biology 13*(3): e1005399. https://doi.org/ 10.1371/journal.pcbi.10053

6. Companion Resources: Topical Guidelines & Ethical Frameworks

6.1 AI and Machine Learning: Internet Research Ethics Guidelines
6.2 Corporate Data: Ethical Considerations
6.3 Feminist Research Ethics
6.4 An "Impact Model" for Ethical Assessment

AI and Machine Learning: Internet Research Ethics Guidelines (IRE 3.0 6.1)

Anja Bechmann, Aarhus University & Bendert Zevenbergen, Princeton University

Cite as: Bechmann, A. & Zevenbergen, B. 2020. AI and Machine Learning: Internet Research Ethics Guidelines, IRE 3.0 Companion 6.1, Association of Internet Researchers, https://aoir.org/reports/ethics3.pdf

1. Introduction: Outline of Ethics in the Context of AI and Machine Learning

1.1 The Focus of These Research Ethics Guidelines

The digital environment enabled by the Internet is a space for human interaction. Scholars in the social sciences and humanities use the technologies, platforms, and data collection capabilities offered by the internet as research tools or as an object of study and online surveys, web-scraping, and ethnographic methods in online communities are still popular techniques. As the technologies collectively known as machine learning (ML) and artificial intelligence (AI) are being deployed in society, scholars, too, are making use of their capabilities as tools for research or studying their (social) impact.

These research ethics guidelines focus specifically on the moral issues raised by the use of machine learning and AI models for Internet research (Floridi & Taddeo, 2016). We apply the structure and approach of the Association of Internet Researchers Ethical Guidelines 3.0 (here onwards IRE 3.0) specifically to the use of ML and AI models in internet research. The four main sections are divided into separate sections that contain relevant considerations and questions for researchers and reviewers. We hope this structured and inquisitive approach allows researchers to assess and justify their research methodologies and improve the trusted relationship to both knowledge generated through these models and the subjects involved (Taddeo, 2010, 2017). A secondary aim is to allow ethics committees at research institutions, journals, and conferences to assess research submissions through an informed lens and break down the needs for required and possible documentation.

The guiding question of this module is: *"How does the AoIR community ensure that we have an ethically sound approach to using AI technologies in our work?"*

1.2 AI Technologies Used in Social Research

A complete overview of the use of AI technologies in social science and the humanities is beyond the scope of these guidelines. In summary, though, it is useful to differentiate between (1) the use of AI technologies as part of a methodology for research and (2) studying the sociotechnical interactions between humans and AI technologies. This document focuses on the first.

The guidelines conceptualize machine learning (ML) as a certain type of models in the overarching framework of AI (Russell & Norvig, 2009). We specifically focus on models that are *built* and *trained* with the purpose of better recognizing patterns, clusters, and structures in data. In this way we have a broad understanding of what constitutes AI, which is not isolated to imitation games and human simulations but purely to the understanding of processing large amounts of data by models to learn to cluster and/or recognize patterns (Alpaydin, 2016). Examples of models that have been used in social science and humanities research within the broad field of Internet research:

- Neural networks (see for instance Krizhevsky, Sutskever & Hinton, 2012)

- Specific natural language processing (NLP) models such as Latent Dirichlet Allocation, Word2vec etc. (see for instance Chen, Li, Zhu & Chen, 2015)
- Naive Bayes (see for instance Ng & Jordan, 2002; Russell & Norvig, 2009)
- Support vector machines (SVM) (see for instance Cortes & Vapnik, 1995)
- As applied to network analysis for instance HAM (Pal et al., 2016)

1.3 AI Trends in Social Research

The recent uptake of AI models is caused by several factors, including increased computational capacity, an abundance of data, and funding available for experimentation with computational models. These factors also each can provide ethical dilemmas.

Dependence on third-parties

Depending on the scale of the project, researchers may be able to run their AI models on their own systems. In many other cases, researchers need to rely on the (1) computational capacity (e.g. High-Performance Computing facilities) and the (2) proprietary datasets that are owned by companies or government agencies. This has given rise to partnerships where academics make use of the systems, data, and funding from outside organizations, which may limit their ability to conduct their research freely and in a responsible manner (Bruns et al., 2018; King & Persily, 2019; Moeller & Bechmann, 2019).

Abundance of data

Open data, smart city initiatives, and open source datasets have generated a large pool of data that can be processed through ML models to understand patterns. These datasets may not contain personal data *per se*, though they may provide proxies for human subjects and their behavioural patterns when combined with other, auxiliary datasets.

Dependence on open source AI models

Internet researchers can develop their own AI models to process datasets and infer answers to their research questions. However, social scientists will typically use (pretrained) open source models. Researchers are left to rely on the documentation available for the models they decide to use. The decisions made in developing these models determine the extent to which the processing adheres to some fundamental principles (see also Bechmann & Bowker, 2019).

1.4 New Issues for Internet Research

As stated in the IRE 3.0, the research question, the research methodology, and the arising ethical issues are closely interlinked. The use of AI technologies allows researchers to apply new methodologies and to expand the scope of existing methodologies. These technologies have new and distinct characteristics from those used in more traditional Internet research. These characteristics may impact the moral issues that arise when AI technologies are applied in social science or humanities research.

Such issues typically relate to the intersection of models, (training, test and live) data and the surrounding contexts (human and social) that define the decision space in which the model is supposed to learn, act, evaluate and adjust. And specifically focus on how AI interacts with other models and technologies and actors in society, humans as well as autonomous agents.

When using ML models for internet research, questions around dataset accountability, bias in its training, and normalization in data cleaning become important to address. We address these issues in this guideline document. Our overarching aim is that this document lays the groundwork for more discipline-specific guidance by individual journals, conferences, and departments.

2. Working Responsibly with AI and ML in Research

The use of AI and ML models raises many ethical issues, which may vary depending on the focus and scope of the research project. We address some of the ethical questions and themes that arise across domains within internet research. Because different ethical dimensions may be in conflict with each other (e.g. transparency or vaster datasets to reduce biases versus more privacy and control for data subjects) several questions focus on addressing these tensions by making explicit – and thus give the opportunity to reflect on – the choices that have been made. In the following section, we separate these questions into different phases of a typical research study. However, the research process is often iterative and the questions posed below may therefore be relevant in different steps of the document.

2.1 Initial Research Design

The initial research design includes the steps such as gaining permission to conduct the study, applying for funding application, detailing the research study, and collecting the data.

Sociotechnical Context

The first step is to create an understanding of the AI, its operation, and social impact based on relevant contextual factors. The researcher must scrutinize the social, political, and economic context within which a technology operates, as much as the technology itself. This is particularly pertinent when dealing with AI as the learning models may show patterns that are unexpected. AI can also amplify already existing social hierarchies. The influence of stakeholders (e.g. developers in different communities and organizations inside and outside academia) may be strong when for instance cleaning data, using pretrained models or changing the models.

- How would you characterize the social context within which or about which the research is conducted?
- Who are the data subjects and affected stakeholders involved in this project (directly or indirectly)?
- How would the researcher characterize the norms (e.g. privacy, social hierarchy) and sensitivities in this social context?

- How have these norms and sensitivities influenced the application of AI in gathering data for this research project?
- How will the implementation of the AI system or use of the model affect the norms and sensitivities?

Research Aims and Risks of Harm

The aim, purpose, and perceived benefits of a project need to be clearly stated as a crucial first step of an ethical analysis. Emerging risks will be judged against these in a balancing exercise.

Some AI scholars would claim that stating a too narrow aim would prevent the model from performing properly as the strength of AI lies in processing large amounts of data inductively (Alpaydin, 2016; Shirky, 2005). Other critical algorithmic researchers would claim that we need to disclose potential power structures in these loose spaces of operation (Bechmann & Bowker, 2019; boyd & Crawford, 2012; Crawford & Calo, 2016; Sandvig, Hamilton, Karahalios & Langbort, 2016).

- Can the researcher articulate what their work aims to uncover?
- How will this research contribute to the state of the art in understanding Internet-related phenomena?
- How will the research benefit society and specific stakeholders?
- Will the research aims create potential risks of harm for the individuals and groups involved directly or indirectly?
- If a generic aim is used how may the researchers/developers influence the field and subjects directly and indirectly involved?

Legitimacy

Conducting research, especially with complex and often opaque ML models, confers a degree of power in the researcher. The researcher decides which technical features will be used, and which data flows will be enacted. This paternalistic approach raises questions about the extent to which the stakeholders or data subjects are aware and in agreement with the experiments or data collections conducted in their social domains.[26] Additional to questions from IRE 3.0, other questions are therefore relevant to highlight:

- How does the researcher justify the foreseen intervention in a social, economic, or legal context by technical means?
- Is the researcher able to understand and explain in an accessible manner how AI is to be used in the research?

---

[26] This is especially pertinent with research involving AI as it typically involves large datasets where gaining informed consent from all data subjects for the new research is not feasible. At the same time consent would not be informed as the understanding of the contribution would be too complex or generic for lay people due to the character of the models (Bechmann, 2014; Nissenbaum, 2011; Solove, 2004). Another dimension would be the dilemma that data subjects would consent on behalf of other people in the frequent case within internet research where data is interaction data belonging to several data subjects (or none) at the same time, scaling the number of consents exponentially per data unit.

- Is the researcher able to explain why the approach by technical means and the use of AI are better suited than any alternative methodology?
- Have the persons who will be affected by the AI system requested the research?
- If not, have they been informed and have they agreed?
- If gaining the informed consent of all data subjects is infeasible, can the researcher obtain proxy consent from a representative or institutional ethics board on their behalf?
- How has the balance between advantageous ends and individual freedom been struck?
- Which values did the organization decide to promote, and how?

Hypothesis or Explorative Research

Quantitative research is typically guided by stated hypotheses, which may be written down before data collection and analysis begins. In qualitative research, researchers often do not work with clear hypotheses to be tested but instead enter the field in order to learn something about the practice in a particular social setting. They conduct self-reflection and state their assumptions as well as the effect of their presence in the research domain. These processes add accountability and transparency to the research process.

In technical domains and data science, even though relying on massive amounts of data, it is also common for researchers to collect data over a period of time without a clear stated goal in order to find strong predictors, correlations or interesting cluster phenomena. We suggest that researchers who use AI systems in their research consider a hybrid of an *ex ante* hypothesis (still making room for exploration) as well as documenting actions taken and choices made throughout the process along with self-reflections on how this research practice has affected their findings and research questions.

- How do the research questions or hypotheses affect and control the outcome?
- If the researcher did not store a fixed hypothesis, 1) how has the researcher's choices been documented? And 2) how has the researcher affected the outcome by choosing this practice (e.g. discussing the presence of proxies and spurious correlations)?

Data Collection

Data collection for the model training and the research can be done either by using open data repositories or by directly recruiting participants to make a new dataset. The use of existing datasets raises issues around its intended openness, consent for reuse, and the change of context for which the data is used (Nissenbaum, 2001, 2009). Collecting new data raises issues around meaningful informed consent, whether the subjects are aware of what their data and the resulting research outputs will be used for, how this will affect them and others, and the representation of humans by a necessarily more limited model.

More general questions arise about privacy as a concept to allow data subjects self-determination and control over how data about them is used. Further, respect for autonomy ensures an individual's ability to make decisions for themselves, and to act upon them. Modern digital data collection (e.g. Application Programming Interfaces) and processing techniques have put the various concepts of privacy and autonomy under significant strain. It is therefore important for researchers to be mindful of ways to minimize the risk to research subjects' and any violations of privacy and autonomy by third parties. Further, applying technological solutions such as encryption are often mistakenly classed as efforts to improve privacy, while they instead provide more security. Similarly, not disclosing information is called confidentiality, not necessarily privacy.

General Data Collection:

- Are the identified data points necessary, relevant and not excessive in relation to the research aim?
- To what extent will data in the database identify individuals directly, or indirectly through inference?
- Do the datasets contain classifiers that are particularly sensitive or even protected classes? If so, what purpose do they serve? Can data points be used as proxies to reconstruct sensitive and protected classes? Is it possible to prevent the re-construction of sensitive and protected classes?
- How does the researcher protect the privacy of its users beyond security measures? For example, is data deleted after a certain amount of time? Is data that is not used for the purpose of the model deleted upon its inadvertent collection?

Existing Datasets:

- Is the existing dataset explicitly open for public or research, or was this dataset found without its reuse permissions being specified?
- Is the use of the existing dataset restricted by legal or other means?
- Could the data subjects (whether anonymized or not) in the existing dataset conceivably object to the new use of their data? Does the initial consent (informed or proxy) cover the intended re-use of the dataset?
- What are the limitations in the knowledge derived from the data in modeling individual and collective behaviour in its totality? How does this limit the generalizability of the findings of the study or the applicability of the precision and/or predictors found?

New Data Collection:

- Have data subjects consented to the collection of their data with a full understanding of what is being collected, for which purposes, and with an understanding of how the data will be used by the researcher? If not, have the collection processes gone through an ethical review board? And/or how has the research team reflected on how to otherwise gain proxy consent and the potential consequences of the proxy status?
- How could potential risks of harm be communicated to the research participants before entering the study?
- To what extent can researchers confirm whether people understand the consequences of derivative uses of their data in AI and ML, knowing from existing literature that the concept of 'informed' consent may not be meaningful for the data subjects?
- Has the organization decided how the privacy of data subjects is safeguarded?
- Does the system collect more information than it needs?

- Are data subjects empowered to decide which data is collected and which inferences are made about them?
- Can the data subject have access to their data? Can they choose to withdraw their data from the system?

Assessing Alternative Methodologies and Scope of Research

Internet experimentation projects can be scaled to a worldwide level (e.g. Kramer, Guillory & Hancock, 2014) and engineers are typically incentivized to deploy a project as widely as possible to maximize their reach. It is sometimes also just easier to let a project operate without limitations and to see later which data is collected, rather than limiting its scope artificially.

However, the knowledge gained using this collection method can have exposed some problems in specific political and cultural contexts. Risk levels can vary widely based on target countries or for particular target groups (see also Dwork, 2006). Therefore, trying to mitigate the risks and shifts in values in all areas will result in a race to appease the lowest common denominator (or: reduce the utility of the project to appease the context with the highest risk factors).

- How can the researcher limit the scope of the research questions and the project's aim to avoid some risks of harm or negatively affected values?
- How can the researcher limit the scope of stakeholders, by excluding particular groups or countries? If so, would the data collected still be a representative sample to answer the research question?
- Are any risks averted if the researcher limits the project duration to a shorter amount of operation time? And does this conflict with the ability of the researcher to conduct the research in question?

2.2 Research Analytical Process

The research analytical process includes selecting the training data, cleaning the data, developing the model through steps of training, evaluating, adjusting, re-training the model.

Source of Training Data

The inferences and predictions of an AI system are closely connected to the source of the training data and here especially issues on systemic discrimination or biases are interesting to disclose and reflect upon as many previous studies have shown such effects (Barocas & Selbst, 2016; Bechmann & Bowker, 2019; boyd & Crawford, 2012; Crawford & Calo, 2016; Kroll et al., 2017; Sweeney, 2013). The use of AI systems to uncover or predict social phenomena can thus be tainted by biases in the training data set on certain demographics or proxies thereof, which may lead to unfair and unjust outcomes.

- What is the cultural and sociodemographic profile of the datasets used by the researcher to train the models?

- To what extent does the cultural and sociodemographic profile of the training data allow for generalizability of the resulting findings or predictors from the research study?
- Are there particular groups which may be advantaged or disadvantaged, in the context in which the researcher is deploying an AI-system? What is the potential damaging effect of uncertainty and error-rates to different groups?
- How has the demographic profile of the researcher(s) affected the composition of the training data?
- How does the training data as ‘ground truth’ affect different demographic profiles and proxies thereof?

Data Cleaning

Data cleaning is the process of detecting, correcting, replacing and even removing inaccurate and incomplete records from a database and structuring the data in a consistent way that makes it processable in the model. Researchers typically find data cleaning a difficult, time-consuming, though necessary and important part of creating an AI-system. It is therefore tempting for some to cut corners or otherwise speed up the process, which can lead to concerns about the rigor and validity of the study because it is seldom accounted for in details. The time spent on cleaning a dataset and the assumptions that go into this process should be communicated more clearly in the resulting research paper. A descriptive analysis of the study datasets may help to identify missing information, incorrect coding, outliers, and misaligned data by the reader.

- How would you characterize the datasets and their cleaning processes? For which variables was the cleaning process optimized? (Features, labels etc.)
- How have (small) adjustments to the training data to make data fit into a model logic potentially influenced the outcome of the model calculations and predictions?
- If the researcher used the raw data to train the model, to what extent could the resulting model be inaccurate, inappropriate, or dysfunctional?
- Specifically, which actions have been taken by the research team in the process of cleaning the dataset and what potential consequences do these choices have on the predictions and/or findings made in the study?
- How do the data cleaning actions normalize data and what are the potential consequences of taking out outliers in terms of minority representation in the model?
- To what extent does the data cleaning process reflect the character of the data collected and the context in which it was provided?
- What actions have been taken to anonymize/pseudonymize the data and to what extent is it possible to de-identify data subjects? Does the anonymization prevent certain types of analysis and what is the argument for the decisions taken?
- How has the data been stored in order to safeguard the privacy of the data subjects?
- If the research team consists of multiple parties and/or distributed calculations how has access to data been negotiated and established in a safe space solution for data subjects?

Model

The researcher's model, based on cleaned training data, will likely have utility in predicting behaviours, or finding correlations in datasets. Such inferences may not be tailored to individuals or be based on anonymized data. Ethical issues may still remain, however, with regard to (1) the privacy considerations of groups on their collective behaviour and the resulting shifts of power balances, (2) the automation of inferences or decision-making, and (3) biases as well as errors in the output data. These issues may also arise if researchers choose to work with a pretrained model on different datasets, for instance open source models.

Group Privacy and Power Dynamics

- Can the knowledge that is generated and inferred from the model shift power balances with regard to specific communities and societies in the training data or as data subjects in terms of predictive power over their behaviour?
- Could the increased power be operationalized maliciously if the model or inferred data was shared with a third-party, and how could such problems be mitigated?
- Could the predictors identified by the model be operationalized maliciously by a third party when published and how could such use potentially be mitigated?
- To what extent is the organization or the AI-system making decisions for data subjects?

Automation

- To what extent is human deliberation being replaced by automated systems and what consequences does it have for the research results?
- Can the researcher override the automated inferences procedure, how will this be documented and justified for later reproducibility?
- Are the automated inferences explainable?
- Is there a strong incentive for the researcher to take the automated inferences as a base truth? How was the ground-truth identified and is this ground-truth adequate to predict the whole spectrum of the problem and/or population behaviour?
- Can the data subjects influence the reach of the AI-system on their lives, their physical environments, and their own decisions? Should the researchers provide such functionality?

Biases and Errors

- To what extent has the researcher accounted for false positives and false negatives in the output of the model, and to what extent can the researcher mitigate their negative impacts?
- Can the researchers use their model to infer social biases and communicate them?
- How have steps of re-training the model to improve accuracy influenced the outcome and what considerations on representation/non-representation have been made in this practice?

- If the research team uses a pretrained model, are the datasets well-documented and how can the character of the datasets influence the predictions of the research in question and the study of another context/practice?

Model Training

- How many instances of re-training have taken place, what was the reason for each re-training and the result, what were the choices made for changing the settings, and what was the specific type of data added to the training loop?
- How do the re-training choices align with the cultural and sociodemographic profile of the research group, and how does this affect the robustness/generalizability of the predictions and/or the findings of the study?
- What would be the consequences of manually tweaking certain weights in the model and feeding the model with different training data? How would this affect the predictions of the model?

2.3 Publication and Other Dissemination

Publication and dissemination of research findings serve several purposes. First, the results need to be communicated to other scientists and the public. Second, the study must be reproducible and replicable, which are two central concepts of the scientific method. Finally, it may be required (or encouraged) to publish the collected datasets and models in an open, machine readable format, in compliance with emerging ideals of open access, open science, and open research. These actions of dissemination raise some dilemmas for researchers, which are partly the result of the chosen scope of data collection and inherent in the characteristics of the AI system.

Reproducibility & Replicability

The scientific requirements of reproducibility and replicability demand from researchers to describe their experiment in such a way that another person could achieve at least similar results. For social science and humanities research using AI tools, this includes for instance making available the training data, the model and test prediction results if deemed safe for the data subjects (Zimmer, 2010).

- Can the researcher make datasets available without violating the privacy of data subjects or revealing other sensitive information?
- To what extent would rigorous anonymization of research data affect the utility of the data to allow for reproducibility and replicability?
- What exact version of the model did the researcher use, was this model pre-trained and if so, what are the precise specifications of that particular dataset(s), what is the cultural and sociodemographic profile of the dataset?
- Has the journal/conference in question established procedures for uploading material to safe space solution or other process for safe access for review purposes? Are there established repositories that offer sharing solutions appropriate for the specific material that might be used?

- Does the journal/conference offer guidance for how to safely and adequately document AI based research?
  (The NeurIPS conference for example requires completion of a "reproducibility checklist," https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf)

Transparency & Explainability

The scientific method requires a high degree of transparency and explainability of how research findings were derived. This conflicts to some extent with the complex nature of AI technologies (Ananny & Crawford, 2018; Calo, 2017; Danaher, 2016; Wachter, Mittelstadt & Floridi, 2017; Weller, 2017). Indeed, it can be too complex for a researcher to precisely state how research data was created given the way a neural network with potentially hidden layers operates. However, the concepts of transparency and explainability do not mean strictly understanding the highly specialized technical code and data of the trained neural networks. For example, a richer notion of transparency asks researchers to explain the ends, means, and thought processes that went into developing the code, the model, and how the resulting research data was shaped. Similarly, explainability does not need to be exact, but can learn from interpretations in philosophy, cognitive psychology or cognitive science, and social psychology (Miller, 2017).

- Can the researcher give an account of how the model operates and how research data was generated?
- Has the researcher explained how the model works to an institutional ethics board, and have they understood the reasons and methods of data processing?
- What roles have the developers and researchers played and what choices have they made in constructing the model, choosing and cleaning the training data and how has this affected the results and prediction?
- What kind of negotiations have taken place in the decision-making around model selection, adjustments and data modelling in the research process that can affect the result and prediction?

Downstream Responsibility

Models can be used in a variety of ways, or they may influence others to create similar models for other ends. Research ethics frameworks, however, typically require the review process to limit itself to the immediate impact on research stakeholders and not necessarily assess the potential long-term impacts of the research outputs (Zevenbergen, Brown, Wright, & Erdos, 2013).

This may be problematic for omni-use technologies such as AI models. Innovations in AI technologies and their inferences on social and human dynamics may be used for a multitude of purposes for instance tailoring microtargeted communication and thus potentially undermining democracy (e.g. issues of fair election & voting discrimination). Models or datasets that were designed to produce positive social ends can be used towards malicious and destructive ends (e.g. facial recognition to clamp down on political gatherings/dissent). Even if the research aims are beneficial for a wide group of stakeholders, research methods and models may need to be published along with the research outcomes and thus set a standard or precedent and initiate function creep and unintended consequences.

Once an AI system has left the hands of the original researchers, they may not have any control over how their models are used by others. The same is true for the generated research data: once it has been freely published, it will be difficult to contain its further uses.

Legal and constitutional checks and balances on the exertion of power or the use of data may differ around the world (Knapp & VandeCreek, 2007). While it is beyond the scope of an ethics review to assess the political governance in countries around the world, it is useful for researchers to be mindful that their data and models may contain personal and sensitive data that could be used directly or indirectly against individuals in other countries or political systems.

Researchers should thus engage actively with the reality that their methods and models may be misused by others, and find ways to mitigate risks and harms. It is ultimately the responsibility of researchers – in dialogue with ethics boards and other stakeholders in a specific project – to agree on the limitations based on a thorough understanding of the project weighed heavily against the knowledge production it produces.

- What could be the downstream consequence for data subjects for erroneous identifications, labelling, or categorization?
- To what extent is the researcher sensitive to the local norms, values, and legal mechanisms that could adversely affect the subjects of their research?
- To what extent can the researcher foresee how the data created through research project inferences may be used in further, third-party systems that make decisions about people?
- Is it foreseeable that the methodologies, actions, and resulting knowledge may be used for malicious ends in another context than research and to what extent can this be mitigated?
- Which actors will likely be interested to use the methodology for malevolent purposes, and how?
- Can the inferred data be directly useful to authoritarian governments who may target or justify crackdowns on minorities or special interest groups based on (potentially erroneously inferred) collected data? Can this be mitigated without destroying the findings for the specific research project?
- Is it possible to contain the potential malevolent future uses by design?
- Up to which hypothetical data reuse moment is it appropriate to hold the researcher responsible?

2.4 Close of the Project

The close of the project includes data archiving and model storage for future access, deployment and development, and for third parties to replicate similar results on other datasets or reuse the dataset (e.g. with permission) on other research questions. Established repositories such as data archives can assist researchers in post-research data governance.

Post-research Data Governance

Systems and research design will never be as robust as intended. To mitigate unforeseen risks, researchers must be prepared and manage the unknown, also after the project has been completed. For example, when a dataset containing sensitive information is disclosed by a third-party unexpectedly, researchers must alert data subjects so they can take precautions.

- Are the datasets and models stored securely?
- Are some datasets more sensitive than others and do they warrant special security precautions?
- Will the data be destroyed at a specific date? How will this data be destroyed? Or will they be anonymized and archived at a specific date?
- How might the data and model be accessed through an application process and what potential harm to data subjects and/or society might this future access have?
- Is there a containment policy for unexpected breaches or malicious uses and what does it oblige the researcher to do or will this responsibility go to the archival organization?
    - Will researchers contact the data subjects and/or the relevant privacy regulator directly about a breach?
    - To what extent does this depend on the seriousness of the disclosure or the sensitivity of the data?
    - How will harmed data subjects or stakeholders be compensated?

3. Concluding Comments

This set of guidelines has been developed as a starting point for research ethics evaluation of Internet research studies that employ the various methodologies and technologies that fall within the broad category of artificial intelligence. As with any novel technological approach to research, many benefits as well as risks emerge as the research community becomes accustomed to the new methodologies, data collection, and their processing. Fortunately, much has already been written about the general use of AI technologies in research, so these guidelines did not need to reinvent the wheel.

The guidelines followed the structure of the AoIR Ethics 3.0 initiative. By following the steps and answering the relevant questions from top to bottom, researchers will have addressed some fundamental issues in the use of AI technology to datasets that contain data about humans or traces of human behaviour. Research ethics review committees can use these guidelines to supplement their own formal procedures to ensure they are asking the correct questions. University departments, corporate R&D, and other discipline specific organizations may also use these guidelines to draft the rules that are specifically useful to them.

We hope this will be a living document, where the people using these guidelines will also give feedback on what did and did not work for them, and why. This particular document should not be considered as being set in stone, but rather an invitation to discuss the research ethics of AI technologies on a slightly higher level. Only through collaboration and iteration will this document maintain its usefulness. For feedback or any other comments and suggestions, please reach out to the AoiR Ethics Working Group, Bendert Zevenbergen (benzevenbergen@princeton.edu) or Anja Bechmann (anjabechmann@cc.au.dk).

4. References

Alpaydin, E. (2016). *Machine Learning: The New AI.* Cambridge, MA: The MIT Press.

Ananny, M. & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Barocas, S. & Selbst, A. D. (2016). Big Data's Disparate Impact. *104 California Law Review*, *671*.

Bechmann, A. (2014). Non-Informed Consent Cultures: Privacy Policies and App Contracts on Facebook. *Journal of Media Business Studies*, *11*(1), 21–38. https://doi.org/10.1080/16522354.2014.11073574

Bechmann, A. & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, *6*(1), 1–11. https://doi.org/10.1177/2053951718819569

boyd, d. & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Bruns, A., Bechmann, A., Burgess, J., Chadwick, A., Clark, L. S., Dutton, W. H., … Zimmer, M. (2018). Facebook shuts the gate after the horse has bolted, and hurts real research in the process. *Internet Policy Review*. Retrieved from https://policyreview.info/articles/news/facebook-shuts-gate-after-horse-has-bolted-and-hurts-real-research-process/786

Calo, R. (2017). Artificial Intelligence Policy: A Primer and Roadmap. *U.C. Davis Law Review*, *51*, 399–436.

Chen, J., Li, K., Zhu, J. & Chen, W. (2015, October 29). *WarpLDA: A Cache Efficient O(1) Algorithm for Latent Dirichlet Allocation*. Retrieved from http://arxiv.org/abs/1510.08628

Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Crawford, K. & Calo, R. (2016). There is a blind spot in AI research. *Nature News*, *538*(7625), 311. https://doi.org/10.1038/538311a

Danaher, J. (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy & Technology*, *29*(3), 245–268. https://doi.org/10.1007/s13347-015-0211-1

Dwork, C. (2006). Differential Privacy. *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II* (pp. 1–12). https://doi.org/10.1007/11787006_1

Floridi, L. & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), 1–5. https://doi.org/10.1098/rsta.2016.0360

King, G. & Persily, N. (2019). A New Model for Industry-Academic Partnerships. *PS: Political Science and Politics*. Retrieved from https://gking.harvard.edu/partnerships

Knapp, S. & VandeCreek, L. (2007). When values of different cultures conflict: Ethical decision making in a multicultural context. *Professional Psychology: Research and Practice*, *38*(6), 660–666. https://doi.org/10.1037/0735-7028.38.6.660

Kramer, A. D. I., Guillory, J. E. & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Retrieved from http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G. & Yu, H. (2017). Accountable Algorithms. *University of Pennsylvania Law Review*, *165*, 633–699.

Miller, T. (2017). Explanation in Artificial Intelligence: Insights from the Social Sciences. *ArXiv:1706.07269 [Cs]*. Retrieved from http://arxiv.org/abs/1706.07269

Moeller, L. A. & Bechmann, A. (2019). *Research Data Exchange Solution* [Report for the EU Commission]. Brussels: SOMA.

Ng, A. Y. & Jordan, M. I. (2002). On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems 14* (pp. 841–848). Retrieved from http://papers.nips.cc/paper/2020-on-discriminative-vs-generative-classifiers-a-comparison-of-logistic-regression-and-naive-bayes.pdf

Nissenbaum, H. (2001). Securing trust online: Wisdom or oxymoron? *Boston University Law Review*, *81*(3), 635–664.

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Retrieved from https://www.bookdepository.com/Privacy-in-Context-Helen-Nissenbaum/9780804752374?redirected=true&utm_medium=Google&utm_campaign=Base2&utm_source=DK&utm_content=Privacy-in-Context&selectCurrency=DKK&w=AF4ZAU9SB230V3A8038M&pdg=kwd-293946777986:cmp-1597361031:adg-58873824845:crv-303010908953:pid-9780804752374:dev-c&gclid=CjwKCAiA_P3jBRAqEiwAZyWWaGes5fE-4mkm7tgbEDU4MBnKIC-DZVSGqqAyo6IOZuf-iu098bnCUBoCBisQAvD_BwE

Nissenbaum, H. (2011). A Contextual Approach to Privacy Online. *Daedalus*, *140*(4), 32–48.

Pal, S., Dong, Y., Thapa, B., Chawla, N. V., Swami, A. & Ramanathan, R. (2016). Deep learning for network analysis: Problems, approaches and challenges. *MILCOM 2016 - 2016 IEEE Military Communications Conference* (pp. 588–593). https://doi.org/10.1109/MILCOM.2016.7795391

Russell, S. & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3 edition). Upper Saddle River, NJ: Pearson.

Sandvig, C., Hamilton, K., Karahalios, K. & Langbort, C. (2016). Automation, Algorithms, and Politics | When the Algorithm Itself is a Racist: Diagnosing Ethical Harm in the Basic Components of Software. *International Journal of Communication*, *10*(0), 19.

Shirky, C. (2005). Ontology is Overrated: Categories, Links, and Tags. Retrieved May 1, 2018, from http://shirky.com/writings/herecomeseverybody/ontology_overrated.html

Solove, D. J. (2004). *The Digital Person: Technology and Privacy in the Information Age*. New York: NYU Press.

Sweeney, L. (2013). Discrimination in Online Ad Delivery. *Communications of the ACM*, *56*(5), 44–54.

Taddeo, M. (2010). Modelling Trust in Artificial Agents, A First Step Toward the Analysis of E-Trust. *Minds and Machines*, *20*(2), 243–257. https://doi.org/10.1007/s11023-010-9201-3

Taddeo, M. (2017). Trusting Digital Technologies Correctly. *Minds and Machines*, *27*(4), 565–568. https://doi.org/10.1007/s11023-017-9450-5

Wachter, S., Mittelstadt, B. & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, *7*(2). Retrieved from https://ora.ox.ac.uk/objects/uuid:6c90bb41-093c-417b-852a-7716054e44bb

Weller, A. (2017). Challenges for transparency. *ArXiv Preprint ArXiv:1708.01870*.

Zevenbergen, B., Brown, I., Wright, J. & Erdos, D. (2013). *Ethical Privacy Guidelines for Mobile Connectivity Measurements* (SSRN Scholarly Paper No. ID 2356824). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=2356824

Zimmer, M. (2010). “But the Data is Already Public”: On the Ethics of Research in Facebook. *Ethics and Information Technology*, *12*(4), 313–325. https://doi.org/10.1007/s10676-010-9227-5

Academy/Industry partnership and corporate data: Ethical considerations

(IRE 3.0 6.2)

Elisabetta Locatelli, Università Cattolica del Sacro Cuore

Cite as: Locatelli, E. 2020. Academy/Industry partnership and corporate data: ethical considerations, IRE 3.0 Companion 6.2, Association of Internet Researchers, https://aoir.org/reports/ethics3.pdf

1. Introduction

Academy/Industry partnerships are not something new in research. Scholars from several disciplines have reflected on it, unfolding the actors involved and identifying the factors influencing it. On the one side, Academy/Industry partnership is something that may threaten academic research with the potential loss of its nature and aims, while on the other side, it is something to pursue to open research to new paths and horizons.

While it should not be taken for granted, it is worth reflecting on the opportunities and risks of Academy/Industry partnership in the field of internet research, to disentangle the underlying questions and to try to answer to them from the point of view of the research ethics.

While, as noted above, Academy/Industry partnership is not a new phenomenon, the contemporary social, political and economic scenario is nonetheless fast changing and requires the academy to step up or to slow down in order to find its pace within it. In a context in which academic research grants and funding have been diminishing and University-Industry collaboration "is considered a relevant economic driver" (Rajalo & Vadi, 2017, p. 42), external funding and cooperation with corporations, institutions, and foundations represent, for example, an interesting perspective for academic research. Academics are thus required to become entrepreneurs with a role in shaping the use of data (see for example the model of Utrecht Data School (Schäfer & van Es, 2017, p. 17)). They are also expected to nurture civil society, disseminating results among public institutions, stakeholders, and companies, and promoting with them fruitful relationships according to the approach that evidences the presence of the "Third Mission" of the university that directly refers to the contribution that university can give to society (Zomer & Benneworth, 2011).

In the field of social media research, an element that added complexity to this scenario is the change in social media platforms' policies that put new limits and constraints on accessing data. This occurred through the change of the Terms of Service (ToS) and of the API access policies (Freelon, 2018; Kotsios, Magnani, Rossi, Shklovski, & Vega, 2019; Halavais, 2019), causing a major turn in social media research to the point that Axel Bruns calls it the "APIcalypse" (Bruns, 2019). The ways in which these limitations can be overcome have recently been topic of debate inside the academic community, as, for example, in the pages of *Information, Communication & Society* by Bruns (2019) and Puschmann (2019), who discuss precisely the possibility of undertaking partnerships between academic scholars and the corporations that own social media platforms, such as Facebook and Twitter. Their debate will be recalled also later in this essay. Partnering with a corporation that provides funding or that makes available data, in fact, can be an interesting opportunity for developing academic research and for enabling knowledge or technology transfer. However, having a corporation or a public institution that finances academic research or that provides data poses several ethical questions regarding the independence of the research or the nature of data, for example. Other ethical questions are raised when the object of the study is a corporation itself and the contents it posts online.

This essay will try to unfold this area with a view towards ethics, tracing first a brief theoretical background of the relationship between academic and administrative research from two points of view: the traditional sociological debate that took place in social sciences after

Lazarsfeld's reflection and the investigation about models of University/Industry relationship. It will then address the main ethical issues in this field, presenting and discussing some best practices and relevant topics. In the conclusion, a list of issues, suggestions, and questions are proposed in order to sum up the main relevant points of attention.

2. Theoretical Background

2.1 Academic vs Administrative Research

Although Academy/Industry relationship is a relatively recent field of investigation and has been developed in several forms during the recent years (Perkmann et al., 2013), reflections about this topic are not new in social sciences. The most relevant and popular debate is certainly the dichotomy between academic and administrative research. Lazarsfeld gave a crucial contribution to settling the boundaries between the two with his 1941 essay, *Remarks on Administrative and Critical Communication Research*, together with later essays (Hardt, 1992; Simonson & Weimann, 2003). Recently, the debate about administrative and critical research was resumed in audience studies and political sciences in the special issues of *Participations. Journal of Audience & Reception Studies* (Barker, Mathijs & Turnbull, 2015) and the *Journal of Information Policy* (Jayakar, Schejter & Taylor, 2016).

Outlining briefly the origins of the debate: the dichotomy between academic and administrative research and the ways to use empirical data can be traced back to the debate between empiricists and rationalists (Melody & Mansell, 1983) while origins of critical research can be traced back to neo-marxist philosophy (Katz & Katz, 2016).

In Lazarsfeld's reflections, on the one side there is administrative research that is research conducted "in the service of some kind of administrative agency of public or private character" (Lazarsfeld, 1941, p. 8) and focuses on specific questions such as "size and composition of media audiences" and "whether specific types of intended effects are achieved" (Napoli & Friedland, 2016, p. 42). Price (2015) adds that administrative research usually coincides with market or commercial research and that generally in this kind of research methodological choices are minimally explained; there are no references to former literature; and the dissemination of results is done by reports (Price, 2015). On the other side, there is critical research that, again in Lazarsfeld's words, investigates "the general role of our media communication" and "develops a theory of the prevailing social trend of our times" (Lazarsfeld, 1941, p. 9), addressing questions about media organization and control and how they threaten human values.

Price specifies that critical research is conducted by academic scholars, is based on previous literature, is disseminated through journal papers or conferences, and has a clear explanation of methodological choices (Price, 2015).

Other researchers have argued about further distinctions between the two, showing that they also differ in the "ideological orientation of the researcher" (Smythe & Dinh, 1983, p. 117) and "in the allegiance of researchers to the status quo [administrative] versus [changes in existing political and economic institutionalized relations [critical]" (Melody & Mansell, 1983, pp. 109-110).

Administrative research about social media done by commercial research institutes or digital communications agencies (i.e. for marketing purposes) seems, then, to differ from academic/critical research about social media – not primarily in terms of methods, which are often similar, but rather in goals, theoretical assumptions, and outcome. This is of no little importance, since these aspects are crucial to academic work. Academic or critical research has, in fact, the function of questioning the current status quo of media, communication, and/or technology (just to give examples in the field of internet research); to put in evidence the non-neutrality of technology; to highlight questions about power; and to promote development in society (Melody & Mansell, 1983).

Particular attention has to be given to administrative research done inside academy, as a form of cooperation and partnership with corporations or institutions.

2.2 Models of Academy/Industry Partnership

As above mentioned, research about the modes and models of Academy/Industry relationship has increased and refined over the last years, focusing on more systemic and organizational aspects and producing several models of cooperation.

According to Ankrah and Al-Tabbaa, “universities—industry collaboration (UIC) refers to the interaction between any parts of the higher educational system and industry aiming mainly to encourage knowledge and technology exchange” (Ankrah & AL-Tabbaa, 2015, p. 387) and can take different forms such as “Joint Ventures, Networks, Consortia, and Alliances” (p. 390). Perkmann et al. introduce the concept of “academic engagement” that is “knowledge-related collaboration by academic researchers with non-academic organizations. These interactions include formal activities such as collaborative research, contract research, and consulting, as well as informal activities like providing ad hoc advice and networking with practitioners" (Perkmann et al., 2013, p. 424). According to the authors, academic engagement is different but related with academic entrepreneurship and commercialization. Academic entrepreneurship “is the founding of a firm with the objective to commercially exploit a patented invention, or in some cases, a body of unpatented expertise” (Perkmann et al., 2013, p. 424) while commercialization is a “technology transfer” (p. 424) or an action in which “an academic invention is exploited with the objective to reap financial rewards” (p. 424).

Thus, Academy/Industry partnerships may have different forms and goals. Moreover, partnerships may occur at both an institutional and personal level. Bodas et al. document two main areas of governance: “institutional” and “personal contractual” (Bodas Freitas, Geuna, & Rossi, 2013), and also highlight that while usually the focus is on the former, the latter is also very present and thus must be considered. Salleh and Omar, in their analysis of University-Industry Collaboration Models, add another important factor, namely, the presence of the local government, which may support or even promote forms of academic-industry partnerships for the development of the country (Salleh & Omar, 2013).

In this multi-faceted context, studies show that several factors may influence these collaborations. The literature on the antecedents of academic engagement and commercialization highlights the importance of individual characteristics (such as seniority, academic quality and success, or grant funding), organizational context (like the quality of the university or the department of affiliation), and institutional context (like disciplinary affiliation or institutional pressures) (Perkmann et al., 2013). Among the evaluation parameters of the quality of the cooperation there are collaboration, knowledge sharing, culture, financial support, communication, and barriers (Ivascu, Cirjaliu, & Draghici, 2016).

As shown above, the literature usually focuses on understanding the institutional or individual factors and constraints that may influence the quality and the output of the collaborations or propose a model for organizing the subjects involved. Less attention is devoted to more structural topics, such as the independence of academic research or the importance of research ethics. Maybe this is due to the fact that these aspects are specific for each scientific sector and may be taken for granted due to the nature of academic research.

Adopting the approach of ethics as method (Markham & Buchanan, 2012) and applying it to three levels of antecedents of academic engagement, it is possible to unfold some of the ethical issues that arise. At the institutional level the scientific sector and its regulations may be considered, since each scientific sector may have specific rules and guidelines about ethics and public policies may be different. Considering organizational factors, other questions should be posed, for example how ethics will be developed during the research project, how data will be stored, how ethics will be developed during the course of the whole research project and when researchers have to face unforeseen events. There are, then, individual factors, such as how ethics will be managed by the members of the research team or the seniority of the researchers, and the individual ability to identify and address the ethical issues during the research process. Ethics will also interact with the outputs and dissemination of the research projects, whether in academic, educational, or commercial outputs.

To be more practical and to give examples of the ethical issues applied to internet/social media research, the essay will continue with two case studies of research entities that adopt a particular attention to ethics and with two thematic areas interesting for the questions they pose.

3. Academy/Industry **C**ollaborations and **W**orking on **C**orporate **D**ata: **C**ase **S**tudies

3.1 Working on Corporate Data: a Brief Introduction

One interesting point of discussion for the ethics of internet and social media research is raised by research that works on online data produced or about a corporation. Among such research there are examples of data from corporate websites or data posted on branded social media profiles.

Here, one of the main ethical questions is whether companies should be treated like human *persons* or not. Questions raised are, for example, whether informed consent is required to analyse data or if there are privacy protection concerns. Kotsios et al., 2019 note that the GDPR (EU, 2108) does not apply when social network analysis nodes are represented by companies (Kotsios et al., 2019, p. 8), so one may guess that some restrictions do not apply to data about companies. However, data about companies often includes users' data (for example, comments on social media posts), making some ethical questions still relevant. In these cases, the suggestion should be to properly identify the data needed, to minimize the data retrieved and address the ethical questions raised by each type of data. Ess and Hård af Segerstad suggest the principle of data minimization according to which "one should only retrieve the data needed to pursue a specific research question" (2019, p. 177). They also suggest to build ad hoc tools for the project, that may help to retrieve only the data needed avoiding the involuntary download of other data, and to take initiatives aimed at protecting sensitive and personal data like anonymizing them (2019; Hård af Segerstad, Kasperowski, Kullenberg, & Howes, 2017). In the case of the analysis of social media corporate data, a solution can be to build a tool that downloads only data about the corporation or the corporations to be analysed, excluding data produced by users that may include personal sensitive data, such as comments and shares or, when downloading the data, to immediately encrypt sensitive data such as commentator's names.

3.2 Working on Corporate Data: the Case of Social Science One

Other occasions to work on corporate data may consist in data provided by a corporation. Among the projects that work on this kind of corporate data, Social Science One is a project for academic-industry partnership that put a specific attention to ethics:

> Social Science One implements a new type of partnership between academic researchers and the private sector to advance the goals of social science in understanding and solving society's greatest challenges. Our mutually incentive-compatible approach enables academics to analyse and use the increasingly rich troves of information amassed by companies to address societal issues, while protecting their respective interests and ensuring the highest standards of privacy and data security (Social Science One, n.d.).

As King and Persily (2019) describe, the structure of the project is quite complex, in order to guarantee in each stage the independence of the researchers and the respect of ethical principles. Their paper offers a detailed description of the structure of Social Science One. Here only the main point will be highlighted.

First, it is interesting to note that Social Science One separated the various stages of the research in order to reduce conflicts of interest and to guarantee full independence of the academic researchers. Thus, for example, data, funds, review processes, and research projects' submission are separate stages and come from different entities. The main components of the structure of Social Science One are: the company that provides data; foundations that provide finances; the commissions of "respected scholars" (King & Persily, 2019, p. 5) that cooperate with the company for identifying the topic(s) of research and defines an "open grant competition for independent academic experts to receive funding to take on this work" (p. 7); independent scholars who participate in the competition with their proposals that are evaluated by independent peer reviewers and then by the commission that "makes final decisions about which grants to award" (p. 7). After grants are awarded, the procedure is similar to other "university procedures for sponsored research" (p. 8).

Second, the process is based on an accountability model, since each stage and relationship is regulated by contracts, and Social Science One itself "has the *obligation* to report to the public" (p. 7, italics in original).

Third, the attention to ethics is developed in two more aspects. To begin with, Social Science One collaborates with researchers of the PERVADE group (p. 9). Second, it has established nine ethical processes to be followed. Among them there are instructions for ensuring that the research projects submitted follow the most rigorous standards and also the principle of respecting the privacy of the subjects whose data are object of research, using for example the principle of "differential privacy" (p. 14). About the mechanism of differential privacy in Social Science One project see also the work of Messing et al. (Messing et al., 2019). The goal pursued, thus, is to not violate privacy in both the research project and in potential following stage that may occur later, for example when different datasets are matched together potentially creating a situation in which single users can be identified.

The Social Science One project has been recently the topic of debate among academic scholars, especially about the first active collaboration involving Facebook.

Bruns (2019) raises doubts about the real independence of academic scholars and the excessive power of the co-chairs of the project in deciding which research projects should be financed. According to his view, projects like these are forms of "corporate data philanthropy" that "appears designed predominantly to benefit the corporation" (p. 8) and that "leaves precious little room for actual, meaningful, socially responsible, and impactful research, especially at a time when there are growing calls for scholars to embrace a more 'data-activist research agenda' (Kazansky et al., 2019, p. 244) in order to combat pressing societal problems" (p. 9).

One further point of attention is that this approach switches the responsibility for ethics from the researchers as individuals to the collective of people and entities involved (King & Persily, 2019, p. 13). This approach recalls the earlier concepts of distributed morality (Floridi, 2013), distributed responsibility (Ess, 2013), and relational concepts of selfhood (Ess, 2014; cf. in this document pp. 6; 8) in which the value of not only the individual, but also of all the subjects involved, is recognized. In this light, we also may suggest that the approach adopted belongs neither to utilitarianism nor to deontology, of which the former maximizes the good and the latter anchors ethics to human autonomy or freedom – but to virtue ethics (Ess, 2014). In Ess' proposal, virtue ethics should be a candidate for global ethics in which the focus is on the relational self and on the complete growth of the individual, even if with a specific attention to privacy protection (Ess, 2014).

3.3 Academic Entrepreneurship and Ethics: The case of the Utrecht Data School

In an environment that encourages academic entrepreneurship (Perkmann et al., 2013), universities are fast adapting to this scenario. The Utrecht Data School is a best practice of connecting academic entrepreneurship and ethical issues (Utrecht Data School, n.d.-c). The Utrecht Data School has the possibility of establishing partnerships and collaborations with companies, institutions, governments and other entities for research goals (Utrecht Data School, n.d.-a). It is thought of as "research and education platform" (Schäfer & van Es, 2017,

p. 17) that involves researchers and students. It has also the goal of initiating a debate with the relevant stakeholders about data, thereby being an active part of society in promoting an ethical approach to data research.

This strong commitment to ethics is mainly summarized by DEDA (Data Ethics Decision Aid) that offers several tools among which workshops with their data ethics consultants and a worksheet that "addresses the various phases of a project and the different ethical issues that might emerge" (Utrecht Data School, n.d.-b).

The Utrecht Data School offers a systematic approach to ethics in which the role of academics and universities in creating a culture about ethics of research, inside and outside university, is magnified. It also proposes the academic world as a relevant stakeholder for helping to identify and resolve ethical issues also in other sectors of society as well. The Utrecht Data School is still working not only with corporate data but also with data from governments (e.g., social welfare data, employment data, educational records, records of residential burglaries, just to name a few).

3.4 Working on Corporate Data: Ethical and Innovative Approaches

The study of corporate data may be complex, as the ones described above or may be simpler and conducted internally by the firms, with or without the support of third-party agencies to achieve specific goals.

Ajunwa, Crawford, and Ford (2016) reflect on medical data taken from wellness programs. While these kinds of programs are at the moment mainly diffused in the USA, we cannot exclude that they may become more popular outside the USA, and thus insights also about this kind of data should be helpful for the whole community of researchers.

The approach of the authors is interesting because they address the ethical issues raised by wellness programs during all the data collection processes such as: informed consent; data collection and control; and potential for employment discrimination (Ajunwa, Crawford, & Ford, 2016, pp. 476–478). These elements are required in order to provide ten "core promises" (p. 479) for building ethical collection of data among wellness programs. The researchers also propose innovative approaches in which employees not only give data but also cooperate to develop wellness programs.

Another innovative approach is the one illustrated by Halavais (2019) that proposes a model in which researchers are partnering not with companies but with users (Halavais, 2019). Halavais collected projects that adopted this approach, such as crowdsourcing, donating data for science, or the open sharing of web activity through a browser plugin (Halavais, 2019, pp. 9–10). From an ethical point of view, this approach insists on transparency and trust between researchers and subjects (Halavais, 2019, p. 11).

4. Conclusions: Some Remarks and Summarizing Questions

This essay has focused on the ethical issues raised by corporate data in two forms: when academic partners with corporations to work on data or when a research is done on corporate data. It first offered two theoretical backgrounds of the relationship between academy and corporations, recalling the debate between academic/critical and administrative research and then models of Academy/Industry collaboration.

Considering the debate that followed Lazarsfeld's essay from the perspective of the ethics of the research, it is possible to highlight the differences between academic and administrative research regarding purpose, theoretical assumptions, method, outcome, and dissemination. A focal point that should be considered under this perspective is the independence of the researchers and the freedom to address wider questions also when partnering with a company.

The debate about the models of University/Industry partnership showed the great complexity of this growing field, making it possible to identify the subjects involved and the forms of influence. University/Industry relationships emerged as a multifaceted phenomenon, with many nuances and subjects involved. From the perspective of ethics, the studies in this field helped to break down the actors and situations in which ethical issues may arise and how to address them. The case histories described offered the opportunity to discover different ways to address ethical issues.

Social Science One proposed a model of partnership between academy and industry in which the different entities involved are connected but also kept separate (as examples, the selection of research topics, data, funding, research projects) in order to preserve researchers' independence and to treat data ethically (for example, by preserving subjects' privacy). The Utrecht Data School offered a model in which ethics is a core principle of the Academy when proposing external partnerships. These cases also help to overcome the distinction between academic and administrative research by showing the potentialities of academic researchers to make companies and other entities aware of the relevance of ethical issues and to create a culture in this sense. Other emerging fields, such as the analysis of data present in wellness programs or online data produced by corporations, showed that there are new fields and areas where ethical issues arise and must be addressed. Other approaches, such as the ones analysed by Halavais (2019), suggest building a cooperation with users, involving them actively into the research process.

Regarding social media research and direct partnership between universities and social media corporations, the discussion is open and leaves room for very different points of view. Bruns (2019) raises several doubts about this kind of partnership, considering it more an operation of "data philanthropy" (p. 2) and public relations for social media companies than something of which academy would benefit from. Bruns is also very clear in recalling the strong role of academic research that should be a "critical, independent, public-interest scholarly inquiry" (p. 15) that addresses the relevant and global contemporary challenges, in which social media platforms are involved, such as "the disruption of democratic processes by deliberate disinformation campaigns; in the growing polarisation of public debate; and in the re-emergence of naked fascism even within the most powerful democratic nation in the world" (p. 15). Pushmann (2019), although considering the risks and the still unclear outcome of such partnerships, is more optimistic in his view, highlighting that it is worth trying because of the role that social media assumes nowadays globally in social communication dynamics; because of the absence at the moment of alternative models that can protect both the academic standards and the privacy of the subjects involved; and because it is a possibility to understand how these companies work and "and, if need be, contribute to holding them accountable" (p. 8).

If turning back to Lazarsfeld's dilemma between critical and administrative research, we can guess that nowadays academic research needs more than ever to be critical research, focusing on the challenges of the contemporary media environment, understanding its logic, and being a counterpart for corporations, even when cooperating with them, in order to nurture society and to empower users.

Lessons learned from the literature and cases analysed suggest that the *ethics as method* approach (Markham & Buchanan, 2012) and the principle of *accountability* (boyd & Crawford, 2012; Locatelli, 2013) are very useful if applied in contexts where several subjects and issues are present. Another suggestion is that it is helpful is to break down the research process in its components in order to better address the ethical issues that may arise during each step, from funding to dissemination. This last approach was encouraged in the second set of AoIR guidelines on Internet Research Ethics (Markham & Buchanan, 2012) and is further expanded in the current IRE 3.0 documents (cf. 2.2., 2.3. above).

Finally, it will be useful to summarize the main topics treated in some questions that, together with AoIR guidelines and other models (such as the above mentioned DEDA worksheet), may help researchers in identifying the ethical issues raised in the cooperation between academy and corporations or while working on corporate data:

- Who are the stakeholders involved?
- Who is the subject that is providing financial support?
- What are the goals of the research? Are the researchers answering only to the financer's needs and/or do they have also their own research questions to pursue?
- Which is the relationship between the funder(s) and the researchers? And between the different stakeholders involved? Is it a relationship on equal terms or are there some power imbalances?
- Who decides the methodological approach(es)?
- What is / are the methodological approach(es)?
- Who is the owner of the data obtained during the research process and where do the data come from (i.e. originally produced/retrieved for the research and/or already extant corporate data provided by the funder)?
- What are the forms of dissemination of the research (i.e. academic, corporate, educational)?
- Are there risks for the researchers or for the academic institution in carrying on such research (i.e. reputation, data management, ethics)?
- Do the corporate data involve any human subject data? Where are the data stored? Which are the terms of use of the platform studied?

Note and acknowledgement

This essay is a living document that hopes to be useful in raising a debate concerning an emerging field that challenges the ethics of internet research. All comments and suggestions are therefore welcome, especially case studies that may enrich section 3 devoted to case histories and innovative approaches.

I want to thank Charles M. Ess, for his kindness, cooperation, and patience, Fabio Giglietto for his precious comments, and aline shakti franzke for her support in developing this essay.

5. References

Ajunwa, I., Crawford, K. & Ford, J. S. (2016). Health and big data: An ethical framework for health information collection by corporate wellness programs. *Journal of Law, Medicine and Ethics*, *44*(3), 474–480. https://doi.org/10.1177/1073110516667943

Ankrah, S. & AL-Tabbaa, O. (2015). Universities-industry collaboration: A systematic review. *Scandinavian Journal of Management*, *31*(3), 387–408. https://doi.org/10.1016/j.scaman.2015.02.003

Barker, M., Mathijs, E. & Turnbull, S. (2015). Getting Beyond Paul Lazarsfeld, at last? *Participations: Journal of Audience & Reception Studies, 12*(2), 1–6.

Bodas Freitas, I. M., Geuna, A. & Rossi, F. (2013). Finding the right partners: Institutional and personal modes of governance of university-industry interactions. *Research Policy*, *42*(1), 50–62. https://doi.org/10.1016/j.respol.2012.06.007

boyd, d. & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Bruns, A. (2019). After the 'APIcalypse': social media platforms and their fight against critical scholarly research. *Information Communication and Society*, *22*(11), 1–23. https://doi.org/10.1080/1369118X.2019.1637447

Ess, C. (2013). *Digital Media Ethics, 2nd Edition*. Cambridge: Polity Press.

Ess, C. (2014). Selfhood, moral agency, and the good life in mediatized worlds? Perspectives from medium theory and philosophy. In K. Lundby (Ed.), *Mediatization of communication* (pp. 617–640). Berlin/Boston: De Gruyter, Inc.

Ess, C. & Hård af Segerstad, Y. (2019). Everything old is new again: the ethics of digital inquiry and its design. In Å. Mäkitalo, T. E. Nicewonger, & M. Elam (Eds.), *Designs for Experimentation and Inquiry: Approaching Learning and Knowing in Digital Transformation* (pp. 176–196). London: Routledge.

EU. (2108). GDPR. General Data Protection Regulation, (GDPR) Regulation EU 2016/679. Approved 27 April 2016, implemented May 25 2018. Retrieved September 25, 2019, from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679.

Floridi, L. (2013). Distributed Morality in an Information Society. *Science and Engineering Ethics*, *19*, 727–743. https://doi.org/10.1007/s11948-012-9413-4

Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, *35*(4), 665–668. https://doi.org/10.1080/10584609.2018.1477506

Halavais, A. (2019). Overcoming terms of service: a proposal for ethical distributed research. *Information, Communication & Society*, *22*(11), 1–15. https://doi.org/10.1080/1369118X.2019.1627386

Hård af Segerstad, Y., Kasperowski, D., Kullenberg, C. & Howes, C. (2017). Studying Closed Communities On-line Digital Methods and Ethical Considerations beyond Informed Consent and Anonymity. In M. Zimmer & K. Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social Age* (pp. 213–225). Berlin: Peter Lang.

Hardt, H. (1992). *Critical Communication Studies: Essays on Communication, History, and Theory*. London: Routledge.

Ivascu, L., Cirjaliu, B. & Draghici, A. (2016). Business Model for the University-industry Collaboration in Open Innovation. *Procedia Economics and Finance*, *39*, 674–678. https://doi.org/10.1016/s2212-5671(16)30288-x

Jayakar, K., Schejter, A. & Taylor, R. (Eds.). (2016). Administrative versus Critical Research: Implications for Contemporary Information Policy Studies. [Special Issue]. *Journal of Information Policy*, *6*.

Katz, E. & Katz, R. (2016). Revisiting the Origin of the Administrative versus Critical Research Debate. *Journal of Information Policy*, *6*, 4–12. https://doi.org/10.5325/jinfopoli.6.2016.0004

Kazansky, B., Torres, G., van der Velden, L., Wissenbach, K., & Milan, S. (2019). Data for the social good: Toward a data-activist research agenda. In A. Daly, S. K. Devitt, & M. Mann (eds.), *Good data* (pp. 244–259). Amsterdam: Institute of Network Cultures. Retrieved from http:// networkcultures.org/wp-content/uploads/2019/01/Good_Data.pdf

King, G., & Persily, N. (2019). *A New Model for Industry-Academic Partnerships*. Retrieved from https://gking.harvard.edu/files/gking/files/partnerships.pdf

Kotsios, A., Magnani, M., Rossi, L., Shklovski, I. & Vega, D. (2019). An Analysis of the Consequences of the General Data Protection Regulation (GDPR) on Social Network Research, 1–23. Retrieved from http://arxiv.org/abs/1903.03196

Lazarsfeld, P. F. (1941). Remarks on Administrative and Critical Communications Research. *Studies in Philosophy and Social Science*, *9*(1), 2–16. https://doi.org/10.5840/zfs1941912

Locatelli, E. (2013). Analizzare la comunicazione nei social network tra ricerca accademica e amministrativa. *Comunicazioni Sociali*, *3*(3), 359–369.

Markham, A. & Buchanan, E. (2012). *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0)*. Retrieved from https://aoir.org/reports/ethics2.pdf

Melody, W. H. & Mansell, R. E. (1983). The Debate over Critical vs. Administrative Research: Circularity or Challenge. *Journal of Communication*, *33*(3), 103–116. https://doi.org/10.1111/j.1460-2466.1983.tb02412.x

Messing, S., Degregorio, C., Persily, N., Hillenbrand, B., King, G., Mahanti, S., … Wilkins, A. (2019). *Facebook Privacy-Protected URLs light Table Release*.

Napoli, P. M. & Friedland, L. (2016). US Communications Policy Research and the Integration of the Administrative and Critical Communication Research Traditions. *Journal of Information Policy*, *6*, 41–65. https://doi.org/10.5325/jinfopoli.6.2016.0041

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., … Sobrero, M. (2013). Academic engagement and commercialisation: A review of the literature on university-industry relations. *Research Policy*, *42*(2), 423–442. https://doi.org/10.1016/j.respol.2012.09.007

Price, S. M. (2015). Academic and commercial research: Bridging the gap. *Participations: Journal of Audience and Reception Studies*, *12*(2), 168–173.

Puschmann, C. (2019). An end to the wild west of social media research: a response to Axel Bruns. *Information, Communication & Society*, *22*(11), 1–8. https://doi.org/10.1080/1369118x.2019.1646300

Rajalo, S. & Vadi, M. (2017). University-industry innovation collaboration:

Reconceptualization. *Technovation*, *62–63*(April), 42–54. https://doi.org/10.1016/j.technovation.2017.04.003

Salleh, M. S. & Omar, M. Z. (2013). University-industry Collaboration Models in Malaysia. *Procedia - Social and Behavioral Sciences*, *102*(IFEE 2012), 654–664. https://doi.org/10.1016/j.sbspro.2013.10.784

Schäfer, M. T. & van Es, K. (2017). *The Datafied Society. Studying Culture through Data*. Amsterdam: Amsterdam University Press. https://doi.org/10.5117/9789462981362

Simonson, P. & Weimann, G. (2003). Critical Research at Columbia: Lazarsfeld's and Merton's 'Mass Communication, Popular Taste and Organized Social Action. In E. Katz, J. D. Peters, T. Liebes, & A. Orloff (Eds.), *Canonic Texts in Media Research* (pp. 10–38). Cambridge: Polity Press.

Smythe, D. W. & Dinh, T. (1983). On Critical and Administrative Research: A New Critical Analysis. *Journal of Communication*, *33*(3), 117–127.

Social Science One. (n.d.). Social Science One. Building Industry-Academic Partnerships. Retrieved from https://socialscience.one

Utrecht Data School. (n.d.-a). About Us. Retrieved September 30, 2019, from https://dataschool.nl/about-uds/?lang=en

Utrecht Data School. (n.d.-b). Deda.worksheet. Retrieved September 30, 2019, from https://dataschool.nl/deda/deda-worksheet/?lang=en

Utrecht Data School. (n.d.-c). Utrecht Data School. Retrieved September 30, 2019, from https://dataschool.nl

Zomer, A. & Benneworth, P. (2011). The Rise of the University's Third Mission. In J. Enders, H. F. de Boer, & D. F. Westerheijden (Eds.), *Reform of Higher Education in Europe* (pp. 81–101). SensePublishers. Retrieved from https://link.springer.com/chapter/10.1007/978-94-6091-555-0_6

Feminist Research Ethics (IRE 3.0 6.3)

aline shakti franzke,University Duisburg Essen

Cite as: franzke, a. s. 2020. Feminist Research Ethics, IRE 3.0 Companion 6.3, Association of Internet Researchers, https://aoir.org/reports/ethics3.pdf

> *Feminism doesn`t need a doctrine of objectivity that promises transcendence, a story that loses track of its mediations just where someone might be held responsible for something, and unlimited instrumental power. We don't want a theory of innocent powers to represent the world, where language and bodies both fall into the bliss of organic symbiosis. We also don't want to theorize the world, much less act within it, in terms of Global Systems, but we do need an earth wide network of connections, including the ability partially to translate knowledge among very different – and power- differentiated- communities. We need the power of modern critical theory of how meanings and bodies get made, not in order to deny meanings and bodies, but in order to build meanings and bodies that have a chance for life. (Haraway, 1988, p. 580)*

There is not one single tradition of feminist history. It is more than just one movement. It is not an ideology, nor just one discourse. It is a multiverse of stories, lives and perspectives. It is too often a story of deaths and injustices, abuses and struggles within oneself, the community, the society, the system. So many different voices of what it means to care, to keep going, to resist. When do we start to act for what reasons? Who is benefiting? How can we legitimize what we are doing?

The internet functions increasingly as a new ecosystem of knowledge production, which lets internet studies flourish. Three strands of research occurred: Firstly, feminist media studies, which focus on critical discourse analysis with an emphasis on the audience, the text or infrastructures. This body of work foregrounds the relations of power and contestation.

Research interests might focus on profile pictures, blogs, games, videos and fan fiction (for example in Shepherd, 2015). Secondly, cyberfeminist approaches, which focus on the myriad experience of users with the help of cyberethnographic research methods (for example Gajjala, 2004). And thirdly, feminist technoscience: these scholars are interested in the materiality of technology and conceptualize their approaches for example with Latour's actor-network theory (for example Bivens, 2015) (Leurs, 2017, p. 137).

Ethics is using moral theory in order to make well informed, hence well-balanced judgements based on certain frameworks. For feminist ethicists, the analytical category of gender plays an important role in the reflection on how to decide. It is not limited, however, to gender-related questions but disagreements and agreements caused by the theoretical and practical question of what social, political and economic consequences follow from being a woman*. In the same moment, it also adds "perspectives of race, class, ability, sexuality, and immigrant status, and many more" (D'Ignazio & Klein, 2019a). Intersectional feminism is about the experience of intersecting power relations- in the sense of experience privilege on the one hand and oppression on the other- that form one's personal experience, that may produce marginalization (see D'Ignazio & Klein, 2019b).

This paper maps out some relevant work that has been done on Ethics of Care in the context of Feminist Internet Research Ethics by providing a generic overview. It is structured in four pillars. Pillar one will set out the scene by adding some historical perspective on feminism. It is never possible to provide a complete overview, so some might be missing. By caring for our bodies, our communities and our research subjects, we are acting ethically. We balance right and wrong along the way by reflection on our context and thereby gain 'situated knowledge'. This will be the topic of pillar two. Thirdly, nothing comes without its problems, therefore some concerns and objections will be listed. Not too many, not too detailed. A starting point. Raise your voice if something is missing. Fourthly, we turn from more philosophical to more practical concerns by addressing some data ethical suggestions. Those gain importance since in internet research the relation between the subject and the researcher might be blurred by (big) data, platforms, logics, materialities. I will end with a summary and some practical implications for the research process and everyday life.

1. On the Shoulders of Giants

Not so long ago, ideas of the ideal democracy were brought forward and with them the Enlightenment flourished. Addressing questions of what it means to be human and who was considered human enough can be described as the beginning not only of humanism in which the core power of 'Man' namely rationality was brought to life through self-regulating powers of the perfect male bodies. By cultivating a picture of the ideal male rationality European humanism was ever since intertwined not only with gender struggles and feminism* but also with European culturalism (Braidotti & Gilroy, 2016, p. 2). Next to philosophers like Immanuel Kant, Rousseau prominently addressed in his book "Origin by discourse of inequality" (1755) important inequalities without seeing women as full members of the human community. This was critically addressed by Mary Wollstonecraft already in the 18th century, who laid, next to authors like John Stuart Mill, Catharine Beecher, Charlotte Perkin Gire and Cady Stanton, important foundations for later feminist reflection by fighting for the right to vote and participate in political life. Given the historically 'young' right to vote for women, we can be quite proud to look back to the achievements of former feminists. What in retrospect sometimes is called the first wave of feminism, was followed by the second wave, which expanded the focus from legal and property rights to topics concerned with family and sexuality. All these issues were considered to be topics of the so-called private realm. By addressing them, the so-called private realm was problematized and its political relevance kept being an important and ongoing debate. Simon de Beauvoir (1968) and her most famous quote "one is not be born, but becomes a woman" can be seen as an important cornerstone that unleashed debates on female and male inequalities, which later expanded to questions of gender identity, their construction, and their political relevance (Butler, 2011).

Within media and communication studies feminist ethics of care gained broader attention especially related to questions regarding methodologies. Feminist ethics and ethics of care emerged early on as especially appropriate frameworks for, e.g., participant observation methodologies in which researchers often felt close bonds with and thereby heightened moral obligations to their research subjects (e.g., Hall et al., 2003; Walstrom, 2004; Buchanan & Ess 2008, pp. 276 f.). Note that participant observations, however, is not used as a method only in Media Studies only but most prominently in Social Anthropology, Sociology, and Biology. McKee and Porter (2010) included care along with five other qualities defining the ethos of feminist research methodology – along with commitments to social justice and improvement of circumstances for participants, critical reflexivity, flexibility, dialogical and transparency (155 f.). (These directly overlap and resonate with the feminist principles of empowerment, reflexivity, and reciprocity identified by Kingston, this volume.) Both care ethics and feminist approaches more broadly are increasingly applied to Big Data research (e.g., Leurs, 2017, Fotopoulou, forthcoming; Lupton, 2018; Luka & Milette, 2018; Rambukkana 2019) and other more quantitatively-oriented research (e.g., Tiidenberg, 2018; Suomela et al., 2019).

2. Ethics of Care and Situated Knowledge

Beginning with the 1960's authors like Charlotte Perkin Gilman, spread the idea of the necessity of so-called 'female values' such as care. In her utopian Novel "Herland" (1915), she describes an isolated state, where only women live, a land where the so-called female* values like caring, educating and development could unfold to create a flourishing and perfect world without oppression. Authors like Carol Gilligan challenged the idea of Kohlberg, who stated that there were six stages of moral development within humans. She pointed out that Kohlberg had only researched white males and illustrated that people in moral decision making are not favouring the abstract over the concrete, the principle over the relationship and the cognitive over the affective, but the contrary. Humans try to make meaningful decision with multiple values involved, which are not perfect but wise for the moment (Murphy & Gilligan, 1980). Nel Noddings was one of the first authors that formulated care ethical principles that have set the fundament for feminist ethics of care, which endorse the experience of being cared for and care for others, in short: relationships over individual principles (1988). "Unlike previous ethical theories that start from the position of an independent rational subject thinking about how to treat other equally independent rational subjects, the Ethics of Care starts with the real experience of being embedded in relationships with uneven power relations" (Suomela et al., 2019, p.2).

Care became a central value of feminist reflection and Joan Tronto and Berenice Fisher delivered a generic definition. Care is understood as "everything that we do to maintain, continue and repour >our world< so that we can live in it as good as possible. That world includes our bodies, ourselves, and our environment, all of which we seek to interweave in a complex life-sustaining web" (Tronto, 1994, p.103, as emphasized by Maria Puig de la Bellacasa, 2012, p. 198).

Next to the care for the own selves, bodies and worlds other feminist arguments have recognized that care is a "non- normative obligation" (Puig de la Bellacasa, 2010) and something that seems unavoidable between reliant and vulnerable beings (Kittay & Feder, 2003). Care is therefore understood not only as vital necessity for flourishing relations. Since interdependency is a fact, an overall nostalgic or romanticized idea of care should be avoided. Feminist inspired visions do not imply a longing for a harmonious ideal world, but want to focus on vital everyday practices that engage with the troubles and interdependent struggles (Puig de la Bellacasa, 2012, p. 199). For a feminist internet research, the power of individual struggles should gain visibility as part of the research process. Questions regarding how we gain knowledge and how knowing can be understood as a relational practice plays a central role within feminist research ethics.

'Situated knowledge', a concept that was introduced by Donna Haraway, asks what counts as knowledge and how this knowledge is gained (Haraway, 1988). For Haraway situated knowledge is about communities and not about isolated individuals (Haraway, 1988, p. 590). The body is understood as an agent and not merely as a resource. Care involves care for one's body, one's community and the relationship towards research participants. For research ethics reflection on the nature of knowledge are important in the sense that knowledge in itself is neither good nor bad, but good or bad in the relation between its content and the context. For what purpose was knowledge gained? What is the relation between knowledge and the knower? Feminist objectivity as understood by Haraway simply means situated knowledge, as a counterprogram of the gods trick, referring to a pseudo objective science that seems to operate from a point of nowhere, from a distance and pretending to be the all-seeing objective one and only truth (Haraway, 1988, p. 581). Similar claims have been made by authors like Harding (1987), who pointed towards the political and relational aspect of knowledge. Liberal feminists like Tong (1998) have promoted research that supports social change and transformation in a way that society becomes a better place for women to live in beyond oppression (Tong, 1998, p. 230).

3. Objections

Contextualizing one's perspective does not mean to fall into relativism but to gain passionate detachment to one's own perspective and to seek self-knowledge. Situated knowledge in its subjectivity is multidimensional, never complete, finished or perfect but therefore an invitation for others to cooperate and add to the complexity. The danger of using care as the only principle could be dangerous, since care and emotions have played an important role in the suppression of women over centuries, where they need to be the emotional caregiver. Martha Nussbaum has therefore critically added the suggestion that we need to focus on human dignity as grounding principles in order to avoid political reproduction of inequalities (Nussbaum, 2013).

Taking care can easily lead to the feeling of being overly responsible for everything. This problem has already been addressed by sociologist Ulrich Beck. It is important to take one's own responsibility seriously but also seek to join others in order to support systematical change. Not everything can be solved by just focusing on care.

4. Data Feminism

Ongoing technological developments and people's interaction with those technologies generate an array of digitized information that can be used to create profound insights about people's bodies, their habits and preferences and their social relations. Small data sets that focus on one particular aspect of life are becoming big datasets once the datasets are aggravated. Data-driven research, however, poses severe issues of how to interpret and make sense of data, how to collect it, cook it, share it and store it, given the fact that other agencies, firms and actors do frequently collect, access and exploit user's data without people's knowledge or consent (Lupton, 2018, p. 1).

The importance of data has given raise for data feminism, which can be described as "a way of thinking about data and its communication that is informed by direct experience, by a commitment to action, and by the ideas associated with intersectional feminist thought" (D'Ignazio & Klein, 2019a). Caring for oneself can be a starting point for situated knowledge by bringing the body and its needs back into the process of knowledge production. Clarification about the definition of data feminism is needed. Several layers can be found. Firstly, from the beginning on it was criticized how woman's achievements in the development of the Internet have been made invisible (See for example Gürer, 2002). Further, it has been criticized in which sense the binary thinking of one and Ceros might reproduce gender ideals. Thirdly, it has been asked in which sense the internet and its materiality might reproduce biases, particularly in AI and its training sets (On Bias in AI see for example Crawford, 2016). Other research has pointed out in which sense working with grey data might cause conflict (see for example Rambukkana, 2019).

It seems to be hard or even impossible to separate online and offline life, since the world, our work, private and public is already frequently depending on the use of the Internet (Floridi, 2015). The broad body of feminist literature and frequent claims for quality, visibility and many more need to be translated and critical theory of data economics need to be added. Data, the so-called new gold of the industry, is also important for research, since the internet, understood as a research object can bring new insights and plays a fundamental role in everyday life. New internet-based methodology, such as data scraping, data visualization and data archives, on the other hand, add complexity and challenge epistemology (See for example: When do we start to trust the machine).

Ethical reflection that has focused on the computer and its regulation (Johnson, 1985; Moor, 1985) has turned into reflection on the nature of information (Floridi, 2015) and in recent years has turned towards reflection on data (Kitchin, 2016; Richards & King, 2014; Mittelstadt & Floridi, 2016; Crawford et al., 2014; Markham, 2013; Metcalf & Crawford, 2016). It is still a long way to fully grasp the inequalities that are reproduced by the information age and how to fully engage in order to take care of one's research community. In order to be capable to gain reflexive judgement, researchers need to be proceeded not determinatively but evaluate the web of relations (Vallor, 2006, p. 105). Even though it might be hard to determine where context starts or where it ends, it is important to reflect on the context (Ess, 2014, pp. 211 f.).

The following principles have been developed for a feminist ethics of care:

- A basic starting point is examining the project from an internal perspective by describing the relationships between the many people who are actively engaged in data collection, analysis, writing, and archiving.
- The next step is to check on the relationships between the research project and the subject community that is involved. Are the research subjects being consulted or respected? Are the researchers avoiding potential harms?
- A final step is to look at external relationships beyond the subjects of the study, often this means looking outward to the wider research community for suggestions and openings for dialogue (Suomela et al., 2019).

Zooming into the data-related aspects Koen Leurs gathered important debates and formulated the following:

1) "People are more than digital data"

- People do not equal data traces; data traces are limited, often ahistorical and decontextualised representations of people.
- 'Big data' are only meaningful in interaction with in-depth 'small data' that value human subjectivities and meaning-making.
- Rather than only extracting data, collaborative data analysis values users as experts over their own practices and attempts to maintain connections between researchers and people under study

2) Context-specificity

- Digital data are never 'raw' but are always already 'cooked' (Lupton,
- 2016, p. 4). The process of cleaning data to ensure software readability does not take place in a vacuum but reflects and shapes intersecting gender, race and classed power relations.
- Situations define moral problems (access, disclosure, privacy).
- User-generated content is generally published in informal spaces that users often perceive as private but may strictly speaking be publicly accessible. In any case,

researchers are rarely the intended audience of user-generated content. Although often mentioned in

- Terms of Service agreements, users may be unaware about the possibility and reality of their data being mined for (commercial) research purposes. Researchers are responsible to inform users about how and why their data are gathered, coded, analysed, stored and possibly shared, and under which conditions.
- Digital data are performative, subjects are known in data-contexts through repetitive enactments and are slotted into known parameters (Day, 2014). These processual relationships between software/hardware/humans are bounded but not determined by sociotechnical configurations (Wajcman, 2004; Barad, 2007) and can only be understood as part of wider visual/audio/textual/hyperlinked/coded/algorithmic narratives.

A feminist ethics of care demands attention to human meaning-making, context-specificity, inter/dependencies, temptations, as well as benefits and harm. A moral focus is on relationality, responsibility inter-subjectivity and the autonomy of the research participants. These concerns offer new ways to theorise and empirically sustain the propositions that digital data cannot be expected to speak for themselves, that data do not emerge from a vacuum, and that isolated data on their own should not be the end goal of a critical and reflexive research endeavour. What a situated ethics of care for data analysis might look like will be further developed in the following substantive empirical section. This endeavour brings its own challenges, and I will discuss opportunities and constraints of striving for a feminist and postcolonial ethics in data analysis. In particular, I relate the guidelines outlined above to assess the 'micropolitics' (Bhavnani, 1993) of creating and analysing Facebook friendship visualisations with young people as part of my aim to develop anti-oppressive digital data research feminist data studies in practice. Reading most journal articles reporting on data-driven research, one gets the impression that gathering user-generated data is a straightforward process of using a software tool to scrape online texts. What often gets silenced in the methods sections of journal articles is how gathering digital data is a context-specific and power-ridden process similar to doing fieldwork offline.

3) Dependencies and relationalities

- Digital data-environments, like social media sites, are characterised by distinctive 'platform values' (Leurs & Zimmer, 2017).
- Presupposing inevitable dependencies. Dependencies include human and non-human actors including users, communities, networks, software, interfaces, algorithms, corporations, governments and fellow academics.
- Dependencies and relationalities are 'asymmetrically reciprocal' (Young, 1997). Encounters between researchers and research participants reflect uneven relations of power and knowledge, which also characterise complicated relations of dependency between human and non-human actors in data analysis.

4) Temptations

- Over-investment with digital personae under study may lead to misunderstandings of a multiplicity of selves.
- Researchers might fail to recognise that human subjects in research hold autonomy and authority over data, have the right to opt out, and can refuse to be coded.
- Over-investment in politics or outcomes might lead to over-identification with those studied.

5) **Benefits and harm**
   - Researchers benefit in various ways from carrying out their studies (personally, career-wise, social standing). Researchers are accountable for creating exploitative relationships with research participants. Rather than ‘do no harm’, research should be beneficial to the people involved.
   - Research participants may benefit in some way from collaborating or participating (personally, professionally or otherwise), causing harm when connections are broken at different points in a study. Those volunteering their data may feel betrayed when the researcher moves on to another project (Leurs, 2017, p. 139 f.).
   -

5. Conclusion

The history of gender equality is still a relatively short one, looking for example at the right to vote. Feminist theory can achieve a lot and our fight to shape reality can be successful. By bringing in the body as an agent, for that we care, our relationships and struggles, we honour the historical aspect of second-wave feminism. And still, there are also political aspects that should not be forgotten. Practical questions of who can examine how ethical research can look like, who will have time for doing what is still undecided. We need collective reflection on what it means to apply situated knowledge to our publishing processes, our research. What would a research environment look like where healthy people can do healthy research? We need to organize structures of support and cooperation instead of fighting against each other for research grants.

Situated knowledge can be a good starting point for an internet ethics of care not only regarding of our process but also in relating to research subjects. Data blurs harm. We need to gain a better understanding of inequalities produced by neoliberal technocratic systems. Further research needs to be done in order to fully grasp in which sense our knowledge production and technological world is producing real harms. New knowledge for change is needed! By adding the body, selfcare and context to our research, it might be easier to further reflect on the bodies and contexts of our research subjects. Ethics of Care does not mean to outsource ethical reflection in order to further accelerate research processes but to take overwhelming struggles in data collection, usage and storage serious. This can serve as a starting point for collective reflection in which sense research might contribute to intersectional feminist thoughts.

Acknowledgments
We would like to express our deep gratitude to the members of the AoIR list who suggested many of the resources incorporated here:

> Jill Walker Rettberg, Elizabeth Losh, Aristea Fotopouplou, Koen Leurs, Heidi McKee, Nicholas Proferes, Deborah Lupton, Elysia Guzik, Katrin Tiidenberg, Shaka Mc Glotten, Stefania Milan, Nathan Rambukkana, Evelina Lilliequist

Special thanks also to Morgane Boidin, Carolin Brendel, Christopher Smith Ochoa for encouraging talks and feedback.

This research was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia.

6. References

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Bhavnani, K. K. (1993). Tracing the contours: Feminist research and feminist objectivity. In *Women's Studies International Forum, 16*(2), 95-104.

Beauvoir, S. D. (1968). Das andere Geschlecht. *Sitte und Sexus der Frau*, *2*.

De la Bellacasa, M. P. (2012). 'Nothing comes without its world': thinking with care. *The Sociological Review*, *60*(2), 197-216.

Bivens, R. (2015). Under the hood: the software in your feminist approach. *Feminist Media Studies, 15*(4)714–717.

Braidotti, R. & Gilroy, P. (Eds.). (2016). *Conflicting humanities*. Bloomsbury Publishing.

Buchanan, E. & Ess, C. (2008). Internet Research Ethics: The Field and Its Critical Issues. In K. Himma and H. Tavani (eds.), *The Handbook of Information and Computer Ethi*cs, 273-292. John Wiley & Sons.

Butler, J. (2011). *Gender trouble: Feminism and the Subversion of Identity*. Routledge.

Crawford, K., Gray, M. L. & Miltner, K. (2014). Big Data critiquing Big Data: Politics, ethics, epistemology. Special section introduction. *International Journal of Communication*, *8*, 1663–1672.

Crawford, K. (2016). Artificial intelligence's white guy problem. *The New York Times*, *25*.

Day, R. E. (2014). Indexing it All: The Modern Documentary Subsuming of the Subject and its Mediation of the Real. *iConference 2014 Proceedings* (pp. 565 – 576). doi:10.9776/14140

D'Ignazio, C., & Klein, L. (2019a). *Our Values and Our Metrics for Achieving Them.* MIT Press Open. Retrieved from https://bookbook.pubpub.org/pub/zkzi7670

D'Ignazio, C., & Klein, L. (2019b). Introduction. MIT Press Open. Retrieved from https://bookbook.pubpub.org/pub/dgv16l22

Ess, C. (2014). Trust, social identity, and computation. In Richard Harper (ed.), *The Complexity of Trust, Computing, and Society* (pp. 199-226). Cambridge: Cambridge University Press

Fotopoulou, A. (forthcoming) Understanding citizen data practices from a feminist perspective: embodiment and the ethics of care. In Stephansen, H. and Trere, E. (eds.) *Citizen Media and Practice*. Taylor & Francis/Routledge: Oxford.

Floridi, L. (2015). *The Online Manifesto: Being Human in a Hyperconnected Era*. Heidelberg: Springer.

Gajjala, R. (2004). *Cyber selves: Feminist ethnographies of South Asian women*. Oxford: Rowman Altamira.

Gürer, D. (2002). Pioneering women in computer science. *ACM SIGCSE Bulletin*, *34*(2), 175-180.

Harding, S. G. (Ed.). (1987). *Feminism and methodology: Social science issues*. Indiana University Press.

Hall, G. J., Frederick, D. & Johns, M. D. (2003). “NEED HELP ASAP!!!” A Feminist Communitarian Approach to Online Research Ethics, in M. Johns, S.L. Chen, and J. Hall (eds.), *Online Social Research: Methods, Issues, and Ethics* (pp. 239-252). New York: Peter Lang.

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies, 14*(3: Autumn), 575-599.

Johnson, D. G. (1985). *Computer Ethics*. Englewood Cliffs (NJ): Prentice Hall.

Kitchin, R. (2016). The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), https://doi.org/10.1098/rsta.2016.0115

Kittay, E. F., & Feder, E. K. (Eds.). (2003). *The subject of care: feminist perspectives on dependency*. Rowman & Littlefield Publishers.

Leurs, K. (2017) Feminist data studies. Using digital methods for ethical, reflexive and situated socio-cultural research. *Feminist Review*, *115*(1), 130-154. https://link.springer.com/article/10.1057/s41305-017-0043-1

Leurs, K. & Zimmer, M. (2017). Platform values: an introduction to the# AoIR16 special issue. *Information, Communication and Society, 20*(6), 803-808. https://doi.org/10.1080/1369118X.2017.1295464

Luka, M. E, & Millette, M. (2018). (Re)framing Big Data: Activating Situated Knowledges and a Feminist Ethics of Care in Social Media Research. *Social Media + Society*, 4(2), 1-10. https://doi.org/10.1177/2056305118768297

Lupton, D. (2016). Digital companion species and eating data: Implications for theorising digital data–human assemblages. *Big Data & Society*, *3*(1), 1-5. https://doi.org/10.1177/2053951715619947

Lupton, D. (2018) How do data come to matter? Living and becoming with personal data. *Big Data & Society* (July–December 2018): 1–11. https://journals.sagepub.com/doi/full/10.1177/2053951718786314.

Markham, A. (2013). Undermining ‘data’: A critical examination of a core term in scientific inquiry. *First Monday*, 18(10). https://doi.org/10.5210/fm.v18i10.4868

Metcalf, J. & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. Big Data & *Society*.(January-June), 1-14. https://doi.org/10.1177/2053951716650211

Moor, J. H. (1985). What is computer ethics? *Metaphilosophy*, *16*(4), 266-275.

Mittelstadt, B. D. & Floridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. *Science and engineering ethics*, *22*(2), 303-341.

McKee, H. A. & Porter, J. E. (2010). Rhetorica Online: Feminist Research Practices in Cyberspace. In Eileen E. Schell and K. J Rawson (eds.) *Rhetorica in Motion: Feminist Rhetorical Methods & Methodologies* (pp. 152-170). Pittsburgh: University of Pittsburgh Press.

Murphy, J. M. & Gilligan, C. (1980). Moral development in late adolescence and adulthood: A critique and reconstruction of Kohlberg’s theory. *Human development*, *23*(2), 77-104.

Noddings, N. (1988). An ethic of caring and its implications for instructional arrangements. *American journal of education*, *96*(2), 215-230.

Nussbaum, M. C. (2013). *Political emotions*. Harvard University Press.

Puig de la Bellacasa, M. (2010). Ethical doings in naturecultures. *Ethics, Place and Environment:A Journal of Philosphy and Geography, 13*(2), 151-169.

Puig de la Bellacasa, M. P. (2012). 'Nothing comes without its world': thinking with care. *The Sociological Review*, *60*(2), 197-216.

Rambukkana, N. (2019) The Politics of Gray Data: Digital Methods, Intimate Proximity, and Research Ethics for Work on the "Alt-Right," *Qualitative Inquiry, 25*(3), 312–323.

Richards, N. M. & King, J. H. (2014). Big data ethics. *Wake Forest Law Review*, *49*, 393-204.

Rousseau, J. (1755). *Un Discours sur l'Origine et les Fondemens de l'Inégalité parmi les Hommes* (A Discourse on the Origin of Inequality; Second Discourse).

Shepherd, T. (2015). Mapped, measured, and mined: The social graph and colonial visuality. *Social Media+ Society*, *1*(1). https://doi.org/10.1177/2056305115578671

Suomela, T., Chee, F., Berendt, B. & Rockwell, G. (2019). Applying an Ethics of Care to Internet Research: Gamergate and Digital Humanities. *Digital Studies/Le champ numérique*, 9(1), 4. https://www.digitalstudies.org/articles/10.16995/dscn.302/

Tiidenberg, K. (2018). Research Ethics, Vulnerablity, and Trust on the Internet. In J. Hunsinger, L. Klastrup, M. Allen (eds.) *Second International Handbook of Internet Research* (pp. 1-15). Dordrecht: Springer.

Tong, R. P. (1998). Postmodern feminism. *Feminist Thought: A More Comprehensive View*, Second Edition, (pp. 193-211). Boulder, CO: West View Press.

Tronto, J. C. (1994). *Moral boundaries: A political argument for an ethic of care*. New York: Routledge.

Tronto, J. C., & Fisher, B. (1990). Toward a Feminist Theory of Caring. In E. Abel, & M. Nelson (Eds.), Circles of Care (pp. 36-54). Albany, NY: SUNY Press.

Vallor, S. (2006). An enactive-phenomenological approach to veridical perception. *Journal of Consciousness Studies*, *13*(4), 39-60.

Wajcman, J. (2004). Technofeminism. Cambridge, UK: Polity.

Walstrom, M. (2004). Ethics and Engagement in Communication Scholarship: Analyzing Public, Online Support Groups as Researcher/Participant-Experiencer. In E Buchanan (ed.), *Readings in Virtual Research Ethics: Issues and Controversies* (pp. 174-202). Hershey, Pennsylvania: Information Science.

Young, I. M. (1997). Feminism and the public sphere: Asymmetrical reciprocity: On moral respect, wonder, and enlarged thought. *Constellations*, *3*(3), 340-363.

An "Impact Model" for Ethical Assessment
(IRE 3.0 6.4)

Annette Markham, Aarhus University

Cite as: Markham, A. 2020. An "Impact Model" for ethical assessment, IRE 3.0 Companion 6.4, Association of Internet Researchers*,* https://aoir.org/reports/ethics3.pdf

Levels of Impact

All research has impact at different levels, likely a combination of negative and positive. Especially in times of rapidly transforming technological capabilities, an impact model of ethics (Markham, 2018) can be a useful assessment tool to break down ethical considerations into more granular units, focusing on the possible outcomes of research design, data collection, management, and storage strategies, analytical choices, and dissemination practices. As these may be obvious or non-obvious, deliberate or accidental, IRE 3.0 includes focus on levels or arenas of possible impact. Drawing from Markham (2018) these include at least:

Level 1: Treatment of People

While researchers interact with people all the time. People may be considered 'human subjects' in some cases, but might also be identified as interface testers, piece workers (i.e. Mechanical Turkers), volunteers, end users, or more abstractly classified as digital signals, user profile, or data points. Considerations of impacts on this level emerge from psychology and sociology domains, where concern for human subject is paramount. Includes the possible impact of, e.g., manipulating news feeds to test the results of certain system inputs, building system pings to prompt Mechanical Turkers to work faster; building features to prompt people to produce data. Most classically, this level of impact includes potential harms commonly associated with human subject research such as psychological, physiological interventions without debriefing, intentional triggering, deception without cause. Less immediate considerations include impact of research on broader communities, both those associated with the target of research and existing or future research communities, and the researcher's own wellbeing and safety.

Level 2: Side Effects

Most research is accompanied by unintended side effects and consequences. These can be obvious, provoked, e.g., during prototype testing, or non-obvious, e.g., if these effects occur later, after the immediate project ends. May be small or large-scale side effects linked directly to particular research project or resulting from processes that combine unexpected elements, e.g., data aggregation by third parties, or at broad infrastructural or cultural levels. Ethical consideration might usefully focus on, e.g., how interface design or implementation might unexpectedly cause users to feel stupid; how a search engine result displaying a person's tweet, thus transferring data from one context to another; when data or algorithms collected or designed for one audience or purpose are analysed or used by others without permission. As Markham notes, "This arena builds on the strengths of science and engineering domains, where a precautionary principle aids in the assessment of short and long term social and environmental effects of scientific developments" (2018).

Level 3: Use of Data After or Beyond Initial Analysis

This level of impact considers how the collection, storage, analysis, aggregation, and further use may generate inferences or conclusions in ways that will influence social classification and categories. Data findings are used as grounds for building specific technologies, creating particular structures, policies, and laws. Issues arise when data sorting and categorizing mechanisms exert or reinforce harmful structures of privilege and marginalization, e.g., genetic data being used to pre-determine job qualifications; insurance companies using quantified self-tracking data to calculate risks and rates; automated profiling through aggregation or multi-source analytics. A strong example of this concern is how datasets and analytic processes can obscure the origin points of police investigations, essentially building predictive systems that bypass (or strip) citizen protections against many forms of warrantless surveillance (Brayne, 2017). This arena is well discussed in feminist, critical, and cultural studies domains, where the processes and politics of categorization, marginalization, and power are central concerns.

Level 4: Future-Making

Any successful research, design, and development will have impacts. This category of impact is less obvious, longer term, possibly unavoidable. Drawing from futurists, researchers can use long term forecasting and speculative methods to consider potential social changes associated with research protocols or processes. With swiftly changing data formats, researchers may think about long term future data loss or build flexible or adaptive platforms for archives. Includes considerations about ecological cost or carbon footprint of continuous or unnecessary Big Data processing, storage, and maintenance.

7. Appendices

7.1 Operational Security: Central Considerations

Keith Douglas
Programmer-Analyst and IT Security Educator
Statistics Canada (*)
philosopher.animal@gmail.com
(*) Employer given for recognition purposes only. This product is on my own initiative.

Cite as: Douglas, K. 2020. Operational Security: Central Considerations, IRE 3.0 Appendices 7.1, Association of Internet Researchers, https://aoir.org/reports/ethics3.pdf

Meta-considerations and Cross-Cutting Topics

There are several people and organizations and their properties that one is protecting when one engages in IT security activities. The first of these is the researchers themselves and their professional and indeed sometimes personal reputational, mental health, financial status, etc. Another is the institution and indeed academia and research integrity more generally. Assets to consider include the computers, operating systems, other software (particularly things like web browsers and email clients) and networking hardware in use. As many of these items can be shared, any decision to perform certain activities has a potential impact on the groups beyond the specific researcher.

A general principle is that online activities should be mediated by an IT Security team with appropriate special responsibility and training. (Ideally, such a team would at least have several individuals with well recognized IT security certifications: the CISSP being the most valuable in this context.)

In many academic contexts, this unit, even if it exists, may be far removed from the academic departments. Such a group usually helps design important documents like “network use policies”. Even though academic or NGO networks are traditionally more open compared to those (say) in for-profit corporations or government/public service, they still generally have rules to abide by. These can be the terms of service of their networks, ISP regulations, and sometimes more general public service or public money requirements.

We shall meet more details on specifics on all these topics as we go. The next sections address 3 topics which came out of experience at AoIR, at Statistics Canada and doing technology awareness discussions at various civil society and NGO organizations.

IT Operational Security

(This should not be confused with OpSec in the military sense)

In order to perform certain activities various software, hardware and social precautions can be taken. We discuss a few as introductions to what they are and why they may be necessary or useful.

VPN

VPN, that is, “Virtual Private Network” is a way to partially isolate network traffic to and from a given machine. These will allow access of resources on the destination (e.g., your institution library holdings) as if present locally. If used exclusively they offer modest protection against “drive-by” attacks but not launched from malicious carriers or other networks (for example not much against café wifi providers, etc.)

Ad blockers

These are add-ons to web browsers which prevent ads of certain kinds being downloaded. This protects against some malicious code and improves download time where this can help. A side effect is often that some sites will fail to work with the blocker enabled. A good blocker (consult your IT Security team for details) will be available (or not) on an ad hoc basis.

TOR

Tor, or “the onion router”, is used for two distinct activities of interest to AoIR: as a local proxy for network traffic and also to access special hosts only available via these protected mechanisms. It is generally used as an anonymization procedure, but some researchers might be wanting to do research on the later aspects of internet (“Dark Net”). In which case, operational security should be a bit stronger due to the criminal associations some activity has. We have no guidance on using TOR this way at this time.

Shared Resources

Most AoIR researchers will be using a university network or similar to do their research (as opposed to their own personal ISP contracted one, etc.) As such, they should be conscious that they are sharing a resource with potentially thousands of other people. Even outbound browsing of web pages can potentially be dangerous, for example, if it invites retaliation. Without special tools (e.g., TOR) researchers should be aware that their affiliation is visible to the provider of any service (e.g., any online forum) and likely also to many of the other users of said service.

In particular, researchers located in the EU and/or collaborating with colleagues located in the EU are thereby required to meet the GDPR requirements for protecting personal data.

Moreover, many organizations may use a so-called web proxy, which will make all (or a large number) affiliated computers look the same. This has emergent effects that should be recognized. The well-known “I am not a robot” from Google services results from the emergent effect of all people in a broad class of users.

This class can be quite large indeed. If I am user@lab.department.university.edu, an abuse notification that fails to reach “lab” may escalate to “department” or “university” quite quickly. Note that even if there is no DNS record for your machine and you simply appear on (e.g.) a web forum as from (say) 127.0.0.1, services like WHOIS can be used to determine (perhaps) the affiliation of the poster.

Firewall

A (computer) firewall is a program (sometimes available as a special hardware appliance) designed to inspect network traffic and deny or allow it based on certain rules. An IT security practitioner may be able to help you set up one for your computer (called "host-based" at that point) if you need extra protection for some activity. We recommend you consult based on business need and see what the expert tells you might help rather than the other way around.

Anti-exfiltration

There is software which is supposed to protect against documents leaving an organization via networks or via attached disks. An IT security practitioner may be able to help you set up one for your computer if you need extra protection for some activity. We recommend you consult based on business needs and see what the expert tells you might help rather than the other way around. Your organization may have a policy on where documents go and how they are stored which may also play a role here; in fact, the IT security task here may be to help compliance with said policy.

Email

Email is a postcard. This slogan should serve as a reminder that email is not encrypted or protected in any way in transit in general terms and can be read by any number of administrators along the way when sent. To set up email to be encrypted either requires sharing a secret out of band (e.g., in person) or a directory of public keys for use with (e.g., PGP). A future version of this document will discuss encryption in more detail.

Additionally, if respondents to a survey or other activity are expected to contact the researcher by email, we recommend using an appropriately dis-associated email account to avoid creating a target at a researcher's permanent account. Contact your IT unit for such; we do not recommend setting up your own email server. Using free external email (e.g., Gmail)[27] might be suitable but this will depend on your institutional policies for such activities. (Many places require official mail for all official business.)

Insecure protocols

Telnet and FTP (proper) require sending credentials in the clear and should never be used except on a very local network if at all. SSH and FTPS/STFP can be used instead with a proper setup.

Similarly, HTTPS is encrypted; HTTP is not. Pay attention to this web browsing. However, there are still other things that can go wrong with HTTPS. Browsers may warn that the certificate issued to the remote site cannot be trusted, for example. Certificates are a way of brokering trust: you can (or not) trust that a place is what is claimed by the certificate by the certificate being issued by a certification authority you trust which is in turn trusted by the manufacturer of your browser (say). Your organization may have its own authority for internal sites also. There is also a lot more to certificates and trust chain that can be covered immediately in this document. (A future version may have more.)

---

[27] Some examples of alternative mail providers: https://www.techjunkie.com/most-secure-email-providers/ or https://www.privateproxyguide.com/best-private-email-providers/)

Chat
Various chat and messenger programs exist, with varying degrees of security[28]. Traditional IRC is completely interceptable; the security stance of almost every other protocol is not well understood. Consult an expert if one's communications are sensitive. Additionally, encrypted communications (especially if not web-based) are frowned upon (or, rarely, outright illegal) in many places and may result in increased rather than decreased state interest.

Data Disposal
Many places have data disposal policies. An important fact to realize is "file deletion" is normally not any more than a file "unlink" – i.e., the data is marked as over-writable but is not gone until sometime later when something else overwrites it. There are data disposal (and recovery) experts who can help.

Community Standards
AoIR researchers often perform anthropological or other "participant observation" research. It should be noted that violating community standards in many online communities may result in retaliation, which may not be limited to the one researcher violating. (See above about shared resources.)

IT Security Virtues
Confidentiality of data, integrity (of systems and data, though also of practitioners), availability (of systems and data). Privacy is distinct, but often related. IT security incidents affect at least one of these. Everyone should work to see what they can do to ensure them as well, which is why they are included here.

Software Development and Security
[TODO: 2.0 of this document]
Vigilantism and Unauthorized Software Testing
[TODO: 3.0 of this document]

---

[28] see, for instance, https://www.securemessagingapps.com/

7.2 Members of the AoIR Ethics Working Group 3.0

Adrienne Massanari (University of Illinois, Chicago, Illinois, USA)
aline franzke (Duisberg-Essen Universität, Germany)
Ane Kathrine Gammelby (Aarhus University, Denmark)
Anja Bechmann (Aarhus University, Denmark)
Anna Jobin (ETH, Zürich, Switzerland)
Annette Markham (Aarhus University, Denmark)
Axel Bruns (Queensland University, Australia)
Bendert Zevenbergen (OII, Oxford)
Carsten Wilhelm (Université de Haute-Alsace, Mulhouse, France)
Charles Ess (University of Oslo, Norway)
Corinne Cath (OII, Oxford / Alan Turing Institute, London)
Cornelius Puschmann (Hans Bredow Institute for Media Research, Hamburg, Germany)
David Brake (Independent Researcher, St. John's, Newfoundland, Canada)
Elisabetta Locatelli (Catholic University of the Sacred Heart, Milan, Italy)
Heidi McKee (Miami University, Ohio, USA)
Ioana Literat (Teachers' College, Columbia University, New York, USA)
Jennifer Stromer-Galley (Syracuse University, New York, USA)
John Murray (formerly SRI, Stanford, California, USA)
Karine Nahon (Interdisciplinary Center, Herzliya (IDC), Israel)
Kate Crawford (Microsoft Research)
Katharina Kinder-Kurlanda (Leibniz Institute for the Social Sciences, Köln, Germany)
Katrin Tiidenberg (Tallinn University, Estonia)
Lauri Kanerva (Facebook)
Mary Gray (Microsoft Research)
Michael Zimmer (University of Wisconsin-Milwaukee)
Michele White (Tulane University, New Orleans, Louisianna, USA)
Mutlu Binark (Hacettepe University, Ankara, Turkey)
Natalie Hendry (Deakin University, Melbourne, Australia)
Natasha Whiteman (Leicester University, UK)
Nele Heise (Independent Researcher, Germany)
P. J. Reilly (Sheffield University, UK)
Soraj Hongladarom (Chulalongkorn University, Bangkok, Thailand)
Steve Jones (University of Chicago, Illinois, USA)
Stine Lomborg (Copenhagen University, Denmark)
Sun Sun Lim (Singapore University of Technology and Design)
Vanessa Dennen (Florida State University, Tallahasee, Florida, USA)
Vivien Chen (Nanyang Technological University, Singapore)
Ylva Hård af Segerstad (University of Gothenburg, Sweden)
Yukari Seko (Holland Bloorview Kids Rehabilitation Hospital, Toronto, Canada)