



**Selected Papers of #AoIR2021:
The 22nd Annual Conference of the
Association of Internet Researchers**
Virtual Event / 13-16 Oct 2021

AOIR ETHICS PANEL 1: PRACTICES AND ROADMAPS

Michael Zimmer
Marquette University, USA

Ylva Hård af Segerstad
University of Gothenburg, Sweden

Panel Rationale and Organization

Since its inception, the Association of Internet Researchers (AoIR) has fostered critical reflection on the ethical and social dimensions of the internet and Internet-facilitated communication and interactions. The AoIR Ethics Working Committee has been committed to not only ensuring the AoIR Ethics Guidelines remain helpful and relevant to researchers and ethical review committees, but also to ensure high-quality research focused on ethics is shared at the annual conference. This panel is one of two sessions organized by the AoIR Ethics Working Committee to highlight recent research engaging with the complexities of addressing ethics in our domain from various disciplinary perspectives, methods, and platforms.

This panel collects four papers exploring a broad (but shared) range of practices that present ethical challenges in internet research, while also providing possible roadmaps towards addressing these concerns. These include: “Problematizing Ethics and Individual Responsibility for Researchers Studying the Far Right” by [redacted] who highlights the risks and challenges researchers face when researching far-right communities online; “Ethical dilemmas in researching sexual crimes of children in the digital society” by [redacted] who explores the ethical dilemmas faced when researching online sexual crimes involving children; “Navigating the Ethical Grey Zone - Improvising Best Practice When Routines and Protocols for Management of Research Data are Missing” by [redacted] who explores practical challenges to conducting ethically responsible research; and “Mapping the Field of Data Ethics: A Roadmap for Educators” by [redacted] which provides a useful analysis of data ethics syllabi in order to provide a framework to guide how we teach data and internet ethics.

If presented live, the panel will be moderated by members of the AoIR Ethics Working Committee and will include a respondent from that group to spark further discussion across the contributors and among virtual attendees.

Suggested Citation (APA): Zimmer, M. and af Segerstad, YH. (2021, October). *AoIR Ethics: Practices and Roadmaps*. Panel presented at AoIR 2021: The 22nd Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from <http://spir.aoir.org>.

Simon Rogerson, Chief Editor *Journal of Information, Communication and Ethics in Society* (JICES) notes his publication aims to "...promote thoughtful dialogue regarding the wider social and ethical issues related to the planning, development, implementation, and use of new media and information and communication technologies." The *Journal* thereby offers "necessary interdisciplinary, culturally and geographically diverse works essential to understanding the impacts of the pervasive new media and information and communication technologies." These papers fit this objective, and are among those under consideration for publication in a special issue of the JICES) associated with the AoIR Ethics Working Committee and AoIR2021.

PROBLEMATISING ETHICS AND INDIVIDUAL RESPONSIBILITY FOR RESEARCHERS STUDYING THE FAR RIGHT

Antonia Vaughan
University of Bath, United Kingdom

Researcher safety in ethics, networked harassment, and academia

The development of social platforms and the far right online has coalesced to form a potent issue to the ethical conduct of research. With a strong history of targeting academia, the far right has continued to innovate harassment strategies on the internet, mostly significantly through networked harassment. Networked harassment is defined as the “form of online harassment against a target or set of targets which is encouraged, promoted, or instigated by members of a network, such as an audience or online community” (Lewis et al 2021, p.2). Veletsianos et al (2018) argue that greater attention needs to be devoted to this issue due to the numerous negative ramifications and uneven impact such as self-censorship and the loss of opportunities. This paper seeks to address the gap with an empirical and theoretical offering by illustrating and critiquing the problematic nexus in which scholars of the far right work. This juncture produces a significant risk that is currently insufficiently met through guidance, training, and funding and needs remedying in order to diversify scholarship and maintain equal access to research.

Online harassment as a phenomenon disproportionately impacts marginalized communities to the extent that it is an “normal part of online experience” (Marwick and Caplan 2018 p.545). This paper argues that academia only exacerbates the intensity of this “normal” experience by providing a spotlight or magnifying glass under which scholars work. Massanari (2018) critiqued the perceived necessity to be present online in how scholars become almost ‘microcelebrities’ through social engagement. This need is part of a valorization of impact promoted by neoliberal institutions and the marketisation of academia. Digital presence is understood as actively “desirable” (Veletsianos et al 2018), with the numerous benefits the internet offers including networking, opportunities, and the public dissemination of work. There are copious amounts of training available to scholars in this area, but these largely lack a consideration of the ramifications creating a tension between needing to be highly present online and the inevitability of networked harassment.

The techno-libertarian ethos of the internet

The inevitability of networked harassment leads to self-censorship, withdrawal, and the loss of opportunities (Veletsianos et al 2018). Exacerbating this issue is the structure of the internet and the techno-libertarian ethos of platforms. Predicated on this idea of a constraint-free technology with minimal guidance, emphasis on the value of free speech and the responsibility of the individual, the tools available both enable networked harassment and prevent protection (Lewis et al, 2021; Massanari, 2017). However, this does not reflect the nature of the harassment. Whereas before there may be one person

harass a user thousands of times, with networked harassment a user may have thousands of people harass them once or twice.

Problematizing the ethics of researcher responsibility.

Considering the significant risk that networked harassment by the far right poses, academics are increasingly being understood as a 'vulnerable population' (Mattheis and Kingdon), necessitating a full consideration in ethics applications for Institutional Review Boards. The current ethical approval process presumes the applicant as the expert, convincing the board that the project will be conducted ethically. Equally, ethical justifications are based on experience, personal relationships, and the currently available guidance. However, whilst the threat from the far right is easily identifiable, the solutions are not. Conway states that in guidance on ethics, there is an "almost total exclusion of potential harms to researchers" (2021, p.368).

In this paper I will illustrate the problematic nexus in which researchers operate that ultimately exacerbates structural racism and sexism, disproportionately impacting those at the sharp end of far right discourse.

Methods and preliminary results

This paper seeks:

- To analyse ethics from a more conceptual perspective, interrogating how covert research, researcher and participant safety, and the use of internet data intersect.
- To create a 10-point framework of best practice for ethical research on the far right.

The corpus puts interviews with academics in conversation with theories of ethics and publicly available guidance from institutions, funding bodies, and networks. By analysing the body of guidance available, the paper offers an understanding of the current corpus, and in particular the areas which are lacking. In order to democratise experiences of researching the far right and promote a diversity in scholarship, we need to meet the major gaps in support and guidance that currently exist. By illustrating the problematic juncture in which researchers of the far right work, this paper critiques the emphasis on individual responsibility promoted by the ethos of the internet and neoliberal institutions. The interviews highlight the individual experiences of developing ethical approaches and tackling the challenges to emotional and physical wellbeing.

The analysis takes a critical approach with qualitative underpinnings, utilising thematic analysis in rounds of inductive then deductive approaches. Specifically, the analysis seeks to highlight flaws with the status quo that primacies individual responsibility, engaging issues of seniority, gender, structural racism and sexism. Massanari (2018) and Veletsianos et al (2018) identify how the current approach significantly disadvantages those already at the sharp end of far right discourse, leading to a loss of opportunity and withdrawal from scholarship

Preliminary findings show that issues of safety are largely subsumed into sections on informed consent or in-person fieldwork sites; safety is framed as a justification for not getting informed consent rather than a fully-fledged section on the possibility for networked harassment, for example. Furthermore, the guidance that does exist fails to holistically account for the identities that exacerbate the danger, or the structural issues that problematise simple solutions. Through the valorisation of the individual, online platforms and neoliberal institutions exacerbate issues of structural racism and sexism by isolating researchers regardless of identity. Researchers are ill-equipped with the tools, guidance, and training to keep themselves safe online when researching the far right.

Conclusion

The intersection between networked harassment, academia, the neoliberal ethos of institutions, and the techno-libertarian ethos of the internet provides a juncture of vulnerability in which scholars work. Exacerbating this vulnerability is the critical lack of consideration of researcher safety in research ethics guidelines, which challenges the attempts of scholars to conduct their projects safely and ethically. The online environment has produced issues of context collapse and a tilt in the power dynamics between researchers and participants. This particularly disadvantages those already at the sharp end of far right discourse as they are already more likely to receive online harassment, and thus suffer from the emphasis on individuality in front of a coordinated threat. The neoliberal insistence on impact and the commodification of research provides a tension between being visible, or successful, and being safe. Continuing this, the techno-libertarian ethos of the internet and the neoliberal ethos of institutions emphasize individual responsibility whilst insufficiently meeting this framing with sufficient tools and support. The particular issue of networked harassment problematizes this framing as it isolates and magnifies one particular individual in front of a malicious audience. Instead, research ethics must conceptualize a framework that incorporates the identities that add vulnerability and this threat landscape.

This paper offers both a theoretical reconceptualization of research ethics and a new practical offering by putting interviews in conversation with current best practice in order to encourage the diversity of scholarship.

References

Conway, M. 2021 Online Extremism and Terrorism Research Ethics: Researcher Safety, Informed Consent, and the Need for Tailored Guidelines, *Terrorism and Political Violence*, 33(2) 367-380

Lewis, R., Marwick, A.E., Clyde Partin, W., 2021 "We Dissect Stupidity and Respond to It": Response Videos and Networked Harassment on YouTube *American Behavioural Scientist*

Mattheis, A.A., and Kingdon, A. 2021 preprint Does the Institution Have a Plan for That? Researcher Safety and the Ethics of Institutional Responsibility

Marwick, A., Blackwell, L., & Lo, K. 2016 *Best Practices for Conducting Risky Research and Protecting Yourself from Online Harassment* (Data & Society Guide). New York: Data & Society Research Institute.

Marwick, A & Caplan, R. (2018). Drinking Male Tears: Language, the Manosphere, and Networked Harassment *Feminist Media Studies* 18(4), pp.543-559

Massanari, A. (2017) '#Gamergate and The Fapping: How Reddit's algorithm, governance, and culture support toxic technocultures' *New Media & Society* Vol.19(3) pp.329-346

Massanari, A.L. (2018). Rethinking Research Ethics, Power, and the Risk of Visibility in the Era of the "Alt-Right" Gaze *Social Media & Society* 1(1), pp.1-9

Veletsianos, G., Houlden, S., Hodson, J., Gosse, C., 2018 Women scholars' experiences with online harassment and abuse: Self-protection, resistance, acceptance, and self-blame *New Media & Society* 20(12) pp.4689-4708

RESEARCHING SEXUAL OFFENCES AGAINST CHILDREN IN THE DIGITAL SOCIETY

Marie Eneman
University of Gothenburg, Sweden

In this paper, I briefly describe my research approach and reflect on ethical issues that can arise when exploring offenders' use of digital technologies for sexual offences against children, with a focus on child abuse material and sexual grooming. Child abuse material, sometimes referred to as child pornography, refers to material depicting the sexual exploitation of children ranging from images of children posing (usually naked) to material portraying physical sexual abuse (Howitt & Sheldon, 2007; Eneman, 2020; 2020) and sexual grooming refers to the process where an adult seeks to establish contact with a child for sexual purposes (Craven et al, 2006; Eneman et al, 2010).

Neither child abuse material nor sexual grooming are new problems in society, but the development of new technologies has facilitated for individuals with a sexual interest in children to seek contact with potential victims (sexual grooming) and to produce, distribute, view and download child abuse material (Howitt & Sheldon; Eneman, 2020). Aside from being criminal offences in many Western countries, child abuse material and sexual grooming are illustrative examples of one the most problematic areas of harmful use of digital technologies. Also characterized as socially unacceptable and stigmatized phenomena and as research topics characterized as emotive and sensitive (Lee, 1993; Dickson-Swift, 2009; Hilário & Augusto, 2020).

One of the first challenges you have to deal with as a researcher when you conduct research on offenders' behavior in relation to child abuse material and sexual grooming is the question of access to empirical material. Accessing the domain of sexual crimes against children is difficult and involves both certain practical and ethical considerations (Hollway and Jefferson, 2012). This includes, for example, how to gain access to people who have experience from this domain, as they can be characterized as a hard-to-reach group for people from the outside (Brantsaeter, 2001).

Due to the fact that many actions surrounding child abuse material and sexual grooming are illegal in many countries, it is most unlikely that people involved in these activities would be willing to discuss their activities prior to conviction. It would also be an extremely difficult and complex issue, practically, ethically and legally, for a researcher to try to gain access to non-convicted people who have experience activities related to child abuse material or in relation to sexual grooming where offenders have contacted (potential) victims (Ray et al, 2010). An alternative approach is therefore to focus on offenders who have been involved in these types of crimes. I therefore choosed to conduct individual interviews with offenders convicted of child abuse material (the Swedish legislation is entitled 'child pornography') and/or sexual grooming in combination with analysis of court records and police reports. Research interviews

with offenders have been described as an invaluable source to better understand criminal behaviour (Presser, 2010; Morgan & Lamble, 2018).

It should be noted that the empirical material, consisting of interviews with offenders together with court and police records represent only individuals that have been arrested and convicted of the production, possession and/or distribution of child abuse material and/or sexual grooming. This means that the material must be interpreted carefully since arrested and convicted offenders represent only a fraction of all offenders engaged in these activities, and that sexual crimes of children have a very low reporting rate (Howitt & Sheldon, 2007).

To access offenders that would be willing to be interviewed I contacted the Swedish Prison and Probation Service and was asked to send an information letter to them describing the research project including the interview guide illustrating the themes and questions I wanted to focus on during the interviews with the offenders. The information letter also contained a description of research ethics in relation to my research. The letter was then distributed within their organization to a prison psychologist working with the types of offenders that I wanted to interview. The prison psychologist then contacted me for a further discussion about the selection of offenders for the interviews. The selection of offenders for the interviews was thus made by the prison psychologist who, in addition to taking into account my formulated criteria, also considered two other important aspects, namely whether the individual offender was in a suitable phase in his therapy and whether the offender was considered a suitable respondent for me as a researcher to interview in the light of my personal safety.

Conducting interviews with offenders about their involvement in sexual offences against children constitute certain complex challenges. Previous research has shown that it can be difficult to get people to talk about their criminal or socially unacceptable behavior (Lee, 1993; Brantsaeter, 2001; Eneman et al, 2010). Therefore, certain ethical recommendations have been proposed when interviewing offenders on sensitive topics which guided me in my research (Hollway & Jefferson, 2012). A prerequisite for conducting research that focuses on sexual offences against children is that the project has been ethically approved. My research has been ethically approved by the Swedish Ethical Review Authority on a number of occasions and also other precautions have been taken in order to protect the integrity of those involved in this study, for example by avoiding very unique descriptions that could reveal individuals. Researchers have an ethical obligation to protect participants/respondents from harm particularly when asking about sensitive topics such as criminal offending behaviour in sexual offences against children (Ray et al, 2010). The ethical issues of informed consent, confidentiality, voluntary participation were clearly communicated in advance together with contact information where the respondent could turn afterwards if questions arose regarding their participation in the study.

When interviewing offenders, the interviews will most likely take place in a prison environment (with sexual offenders), which can feel both unusual and uncomfortable (Noaks & Wincup, 2017). It is often emphasized that it is important to create an environment where the respondent feels safe and secure, but I want to emphasize that it is at least as important for the researcher. Perhaps even more important when it

comes to interview people convicted of crimes. For my own part, I started the interviews in a visiting room in the prison (was referred there by the prison psychologist) but during a meeting with a respondent who when I asked what he was convicted of also answered (in addition to sexual crimes against children) rape of an adult woman, then I got scared and realized that I had not found out any safety routines, for example if there was an alarm button, etc. The interview went well but after that I asked to change room close to the psychologists.

Finally, it should be emphasized that my motivation to conduct research within this area is driven by my ambition to develop knowledge that increases awareness and can contribute to regulative measures such as legislation and policies that further strengthen the protection of children in the digital society. I hope that my experiences of conducting research on a highly emotive and sensitive topic surrounded by a number of ethical issues in the digital society can be an inspiration and support other researchers approaching these kinds of sensitive themes.

References

Brantsaeter, M (2001) Møter med menn dømt for seksuelle overgrep mot barn. Doctoral Thesis, Universitetet i Oslo.

Craven, S., Brown S & Gilchrist, E (2006) Sexual Grooming of Children: Review of Literature and Theoretical Considerations, *The Journal of Sexual Aggression* 12:3.

Dickson- Swift, V., James, E., Kippen, S & Liamputtong, P (2009) Researching Sensitive Topics: Qualitative Research as Emotion Work, *Qualitative Research*, 9:1.

Eneman, M (2020) Crime investigations of 'child abuse material' - Challenges and opportunities posed by digital technology, *The 21th Annual Conference of the Association of Internet Researchers (AoIR)*, Dublin, Ireland.

Eneman, M., Gillespie, A. & Stahl, B (2010) Technology and Sexual Abuse: A Critical Review of an Internet Grooming Case, *Proceedings of the International Conference on Information Systems (ICIS)*, Missouri, USA.

Hilário, A & Augusto, F (2020) Practical and ethical dilemmas in researching sensitive topics with populations considered vulnerable, *MDPI Societies*.

Hollway, W & Jefferson, T (2012) *Doing Qualitative research differently*, SAGE Publications Ltd.

Howitt, D & Sheldon, K (2007) *Sex Offenders and the Internet*, John Wiley Sons Inc.

Lee, R (1999) *Doing research on sensitive topics*, SAGE Publications Ltd.

Morgan, S & Lamble, I (2018) Understanding men who access sexualised images of children: exploratory interviews with offenders, *Journal of Sexual Aggression*, 25:1.

Noaks, L & Wincup, W (2017) *Qualitative Research in Criminology*, Routledge.

Presser, L (2010) Collecting and Analysing Stories of Offenders, *Journal of Criminal Justice Education*, 21:4.

Ray, J., Kimonis, E & Donoghue, C (2010) Legal, ethical, and methodological considerations in the Internet-based study of child pornography offenders, *Behavioral Sciences & the Law*.

NAVIGATING THE ETHICAL GREY ZONE - IMPROVISING BEST PRACTICE WHEN ROUTINES AND PROTOCOLS FOR MANAGEMENT OF RESEARCH DATA ARE MISSING

Ylva Hård af Segerstad
University of Gothenburg, Sweden

Stefan Nilsson
University of Gothenburg, Sweden

Maria Olsson
University of Gothenburg, Sweden

Introduction

On May 25, 2018 the world's strongest set of data protection rules, the General Data Protection Regulation, or GDPR, came into force across Europe. GDPR was designed to "harmonize" and modernize data privacy laws across all of the EU countries as well as to provide greater protection and rights to individuals (<https://gdpr.eu/>). GDPR has now been in place for three year and has prompted regulatory bodies and universities to develop technical solutions and data handling protocols for secure management of personal data for research purposes. Even so, in striving to follow correct procedures for managing sensitive research data as ethically responsible researchers, it became strikingly clear that neither technical solutions nor routines for managing Level 3 data were in place at one of the largest universities in Sweden. This lack can have serious consequences, both for the protection of the research subjects and for the ability to conduct ethically responsible research.

In this paper, we address the consequences of missing such tools and procedures. We describe how we improvised best practice under the circumstances on the fly in collaboration between researchers, legal officers and an IT architect at the university.

Practical Challenges to Being Able to Conduct Ethically Responsible Research

Vitak et al. (2017) described that IRBs often seem to apply overly strict guidelines to low-risk research protocols, while simultaneously lacking the expertise to effectively evaluate technical proposals. These insights, coupled with rapid technological development and requirements to adapt regulations to new legislation, such as GDPR, makes it hard for the slow processes in academia to keep pace when it comes to developing technical solutions and protocols for legally and ethically correct and secure procedures.

Data Collection and Data Classification

In the context of a research and innovation project, we conducted 16 dyadic interviews with young cancer survivors via videoconferencing platform Zoom. Zoom has been

extensively used for research purposes (Lobe et al., 2020), and the platform was chosen for being compliant with the Health Insurance Portability and Accountability Act (HIPPA), and for supporting video-recording with local storage. The Swedish ethics review authority granted us approval to conduct the interviews using Zoom.

Data Storage and Processing

Data security levels are defined differently in different countries. The videorecorded interview sessions generated for the project are classified as Level 3 data in Sweden, i.e., sensitive personal data including information about the participants' health. We password protected the videorecorded interview sessions and uploaded the files to a secure repository provided by the university. Only the researchers in the project had access to the files. The interviews were to be transcribed for further analysis. In order to keep up with the pace of the other work packages in the project, we decided to make use of the services of a transcription company, procured by the university.

This is where we encountered serious challenges which neither we, nor the ethics approval authority, had foreseen (cf. Franzke et al., 2020, Markham, 2018). The storage repository that the university provided turned out to only be approved for storage of data up to information Level 2. Moreover, the routines that the transcription company had established with the university for transferring data to process was to use the university approved productivity platform. However, like the storage repository, this platform is also only approved for up to information Level 2.

Improvising and Solving the Dilemma on the fly

In collaboration with the legal administration officers and an IT architect at the university, we developed a protocol for best practice under the circumstances, that allowed us to securely store, transfer and process the sensitive material in accordance with current regulations. Joining our competences and skills to meet current regulations, we improvised by combining technical affordances of a number of devices, platforms and procedures that were deemed to do the job securely. The procedure we settled on is as follows;

Zoom automatically also creates audio-only files of the recordings. Audio-only were deemed to contain less identifying information than the corresponding video files. We compressed the audio-only files and protected the resulting zip-file with 20-character password. The zip-file was uploaded to the designated area on SharePoint, and from there downloaded by the transcriber. The researchers and the transcriber met in a Zoom meeting and the password was sent to the transcriber using a fully-encrypted application for money transfer (similar to PayPal). The transcriber downloaded and unzipped the files, and worked with them on a computer without internet connection. The interview material was de-identified in the transcription process by substituting identifying information such as names, place names, diagnosis etc. When the job was done, the reverse procedure allowed the researchers to download the transcribed files and begin the process of analyzing the material. The raw data, i.e., the video and audio recordings of the interviews, along with the password-protected files providing the key to

re-identification of the research participants on two password-protected and fully encrypted hard-drives. These in turn were stored in a safe at the university.

Navigating the Ethical Grey Zone

Despite the fact that three years has passed since the GDPR regulations came into force, it seems that it is not at all uncommon that even government run educational institutions have not been able to keep pace with current regulations. At the time, sufficient technical platforms or routines were still not in place for secure data management in accordance with these. We found ourselves being stuck in the research process, as there were no secure procedures in place for transferring the interview files to the transcriber, or to manage the data during the transcription process.

Missing routines and tools to meet changing demands and regulations - legal, practical and ethical - is challenging. Most university researchers currently find themselves in an ethical grey zone. Some are not even aware of the problem, as they rely on the university to provide what is needed, and those who are aware are left without a choice as there are no other options available to them.

Our particular case of developing a best practice - under the circumstances - to solve ethical and methodological challenges contributes to our collective knowledge base and AoIR's ongoing work on ethical guidelines. Furthermore, the case highlights the need for a process approach to research ethics (Franzke et al., 2020). It emphasizes that research ethics has to be case and context sensitive, and that researchers and the research design has to be flexible and adaptive throughout to mitigate possible harm that might arise during the research process (Markham, 2006).

References

- FRANZKE, A. S., BECHMANN, A., ZIMMER, M., ESS, C. & RESEARCHERS, T. A. O. I. 2020. Internet Research: Ethical Guidelines 3.0.
- LOBE, B., MORGAN, D. & HOFFMAN, K. A. 2020. Qualitative Data Collection in an Era of Social Distancing. *International Journal of Qualitative Methods*, 19.
- MARKHAM, A. N. 2006. Ethic as Method, Method as Ethic: A Case for Reflexivity in Qualitative ICT Research. *Journal of Information Ethics*, 37-54.
- MARKHAM, A. N. 2018. Afterword: Ethics as Impact—Moving From Error-Avoidance and Concept-Driven Models to a Future-Oriented Approach. *Social Media+ Society*, 4, 2056305118784504.
- VITAK, J., PROFERES, N., SHILTON, K. & ASHKTORAB, Z. 2017. Ethics regulation in social computing research: Examining the role of institutional review boards. *Journal of Empirical Research on Human Research Ethics*, 12, 372-382.

MAPPING THE FIELD OF DATA ETHICS: A ROADMAP FOR EDUCATORS

Tian Zheng
Columbia University, USA

Isabelle Zaugg
Columbia University, USA

Jonathan Reeve
Columbia University, USA

Data Ethics Education

From the spread of disinformation via social media, to class-biased dynamic pricing, to racial profiling in online systems that lead to “real-world” harms, teaching data ethics has never been more urgently needed. This paper explores current topics within data ethics syllabi. It analyzes the overlaps and divides between various approaches to teaching data ethics, from emphasizing FAT (Fairness, Accountability, Transparency), to the Public Interest Technology movement, to calls to reimagine “digital justice” from Critical Race and Digital Studies scholars, to applying AoIR’s ethical Internet research guidelines, to the way philosophical texts on ethics from different cultural contexts are being applied to the digital age. We present a semantic, linked open data graph describing the relations between texts, courses, professors, and universities involved in teaching data ethics. While patterns in data ethics education will to some degree emerge organically from the data, we also use a “human in the loop” approach to identify and label these patterns.

Informed Course Development

We will then use this tool to inform our own approach to teaching data ethics. As a collaborative team that spans the humanities, social sciences, arts, and STEM, we believe that data ethics education is most meaningful when it draws upon transdisciplinary perspectives to both understand and, importantly, act on data ethics. Our approach will not only familiarize students with data ethics on a theoretical level, but will push them to recognize a horizon of possible solutions and build computational skills to explore solutions in practice.

Computational Problem-solving

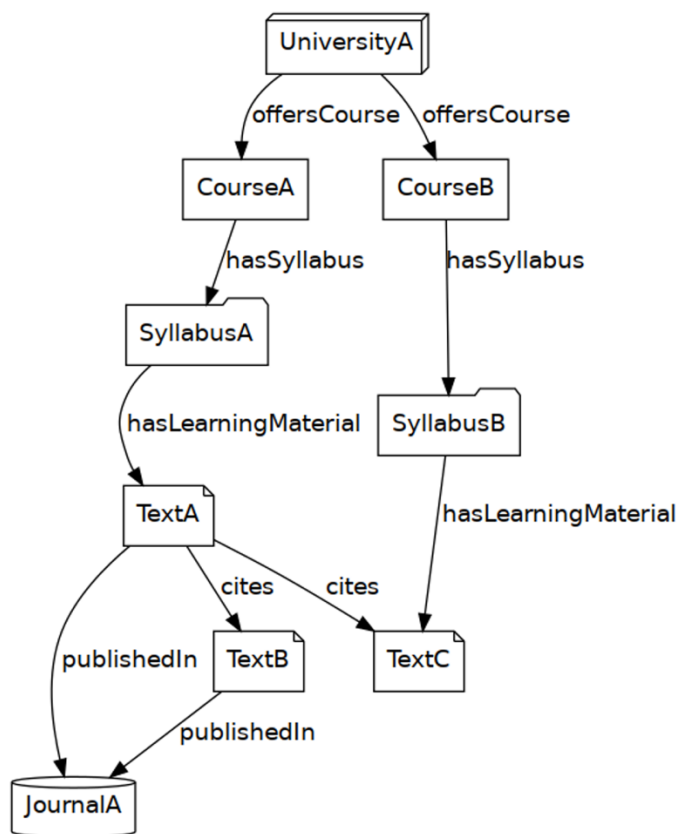
As such, as we look at the field of data ethics coursework and its literature, we are interested in identifying whether these courses help students make the leap into computational “problem-solving.” Even if this is not happening at the level of a course, identifying whether these courses are embedded or “required” within computational programs, will help us understand the degree to which data ethics education is reaching future data scientists.

Widening the Pipeline

Another possible goal of data ethics coursework is to attract new cohorts of students to engage with computational fields. In terms of data ethics, this is important on two fronts. One, it is important to bring students with ethical insights from other disciplinary backgrounds into the field of computation, and transdisciplinary data ethics coursework can facilitate that. Two, by using data ethics as a bridge to open a new pathway into computational fields, there may be a potential to attract students from under-represented backgrounds who might hesitate to take traditional, foundational computational coursework. Both potential “pipeline” outcomes have the promise to address the calls data ethicists have made regarding the need to diversify the computational workforce in terms of disciplinary expertise, cultural background, and lived experience. While we can only speculate about the profiles of individual students taking these courses as well as the long-term outcomes of their learning, we hope this project will shed light on how data ethics education is engaging with the larger disciplinary structures of universities, and particularly whether these courses may be providing an entryway into computational practice for students rooted in other disciplines.

Our Approach

The semantic web, also known as “Web 3.0” or linked open data, is a relatively new system of conventions for standardizing and encoding graph data, such that it is universally interoperable, in a language known as [RDF, or the Resource Description Framework](#). Some of the most well-known projects in the field include DBpedia, the set of parsed and inferred data from Wikipedia, and Wikidata, the data set which proposes to be the knowledge basis for Wikipedia. At its most basic, RDF data may be represented as a series of subject-verb-object triples, where each node has a stable URL. Social relationships between people, for instance, may be described as <Bob> <is friends with> <Alicia>, where the angle-bracketed entities resolve to URIs. There exist a number of *ontologies*, or pre-defined sets of relations, which may be used to describe entities within their domains. We use a number of ontologies in conjunction: the [Curriculum Course Syllabus Ontology \(CCSO\)](#) describes relations between courses, universities, syllabi, professors, and learning materials such as texts; the [Bibliographic Ontology \(Bibliontology\)](#) describes metadata for articles, books, videos, and other media; and the [Citation Typing Ontology \(CiTO\)](#) describes citation relations between texts. [[@Fig:graph](#)] shows an example directed graph visualization, illustrating relations between these entities.



We collect data in a quasi-automated fashion, often beginning with course lists, such as the [tech ethics curriculum list provided by Casey Fiesler et al. \(2020\)](#). Given a course syllabus URL, we are able to automatically extract bibliographic references, and resolve them to stable identifiers at a number of bibliographic databases, such as [CrossRef](#). We then resolve universities and departments to their Wikidata identifiers, which allows us to retrieve additional information with which we can organize our data, for instance the geographic coordinates of the university and its founding date. We resolve instructors and authors to their [ORCID](#)s, which allows us to retrieve an author’s other publications, and their past and present institutional affiliations. This allows us to answer questions such as:

- What are the most-cited works in the field of data ethics?
- Which are most assigned in courses?
- What books are only cited in one geographic region (e.g., California)?
- What courses are cross-listed in multiple departments?

Sharing Our Roadmap

We are building a website to visualize these connections, as a force-directed network visualization in JavaScript, so that it may be explored by a wider user base. We hope this resource will be useful for others designing data ethics coursework, as our data visualization will allow one to quickly identify both valuable patterns in texts assigned, as well as outliers. Our visualization prioritizes users’ engagement with both consensus and outliers, which is important considering that efforts to “decolonize curricula” have

highlighted how the process by which texts gain importance is not always meritocratic and in some cases “outsider” voices deserve to be centered. This is all the more true in a field such as data ethics where critical voices are challenging established perspectives, practices, and institutions. This semantic web approach also allows us to be multilingual by default, since much of this data, such as that gleaned from Wikipedia, is available in many languages.

Further Outcomes

We hope to build a mechanism to submit users’ data ethics syllabi to our database; this way, our literature review will always stay up to date. A further step will be to generalize this framework so that it may be used to map any academic discipline, given a list of courses and their syllabus URLs.

As data ethicists ourselves, we care about openness and transparency, and so we have open-sourced this data, so that other researchers can use our work to answer their own questions. We hope that our framework may be used to help map the institutional knowledge structures of even more disciplines.

References

Fiesler, Casey, Natalie Garrett, and Nathan Beard. “What Do We Teach When We Teach Tech Ethics?: A Syllabi Analysis.” In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 289–95. Portland OR USA: ACM, 2020.