

Are Taylor's Posts Risky?

Evaluating Cumulative Revelations in Online Personal Data

A persona-based tool for evaluating awareness of online risks and harms

Leif Azzopardi
leif.azzopardi@strath.ac.uk
University of Strathclyde
Glasgow, UK

Jo Briggs
jo.briggs@northumbria.ac.uk
Northumbria University
School of Design
Newcastle-upon-Tyne, UK

Melissa Duheric
melissa.duheric@northumbria.ac.uk
Northumbria University
School of Design
Newcastle-upon-Tyne, UK

Callum Nash
callum3.nash@northumbria.ac.uk
Northumbria University
School of Design
Newcastle-upon-Tyne, UK

Emma Nicol
emma.nicol@strath.ac.uk
University of Strathclyde
Glasgow, UK

Wendy Moncur
wendy.moncur@strath.ac.uk
University of Strathclyde
Glasgow, UK

Burkhard Schafer
b.schafer@ed.ac.uk
University of Edinburgh
Edinburgh, UK

ABSTRACT

Searching for people online is a common search task that most of us have performed at some point or other. With so much information about people available online it is often amazing what one can find out about someone else – especially when information taken from different sources is pieced together to create a more detailed picture of the individual, and then used to make inferences about them (leading to *cumulative revelations*). As such, the relevance of one piece of information is often *conditional and dependent* on other pieces of information found. This creates interesting and novel challenges in evaluating information *relevance* when searching personal profiles, posts and related information about an individual, as well as the potential *risks* that can arise from such revelations. In this demonstration paper, we present a tool designed to investigate how people assess and judge the relevance and potential risks of *small, apparently innocuous pieces of information* associated with fictitious personas, such as *Taylor Addison*, when searching and browsing online profiles and social media. The demonstrator also comprises a cyber-safety tool, which aims to provide education and raise awareness of the potential risks of cumulative revelations. It does so by engaging participants in different scenarios where the relevance of individual information items depends on the searcher and their particular underlying motivation.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SIGIR '22, July 11–15, 2022, Madrid, Spain

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-8732-3/22/07...\$15.00
<https://doi.org/10.1145/3477495.3531659>

CCS CONCEPTS

• **Security and privacy** → **Social aspects of security and privacy**; *Privacy protections*; • **Information systems** → *Information retrieval*; *Relevance assessment*; *Test collections*.

KEYWORDS

Information Revelation; Digital Traces; Data Self; Privacy; Cyber Safety; Training Tool

ACM Reference Format:

Leif Azzopardi, Jo Briggs, Melissa Duheric, Callum Nash, Emma Nicol, Wendy Moncur, and Burkhard Schafer. 2022. Are Taylor's Posts Risky? Evaluating Cumulative Revelations in Online Personal Data: A persona-based tool for evaluating awareness of online risks and harms. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*, July 11–15, 2022, Madrid, Spain. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3477495.3531659>

1 INTRODUCTION

In Information Retrieval, we often consider documents as discrete entities, materially and in terms of their content as independent of each other [24]. As such, documents are often judged based on their relevance, significance and usefulness, individually, without taking account of other, even closely proximate documents. However, in practice, information from documents is typically combined and used together, including to make greater inferences. Not only is the sum of all the information greater than its parts, through combination, any apparent “gaps” can be filled or inferred by the user of the information. This is particularly the case when looking for and retrieving information about individuals (such as from their online posts and personal profiles, or from what other sites report about them). Most of us have gone online and conducted a search for someone, whether it be to find out about:

- a colleague – *what have they been working on?*
- a friend – *what have then been up to lately?*
- a date – *what are they interested in?*
- or just someone we met online and are curious about – *are they a catfish?*

However, the sum of information posted by or about the individual comprises digital traces that can be used maliciously, by others, and result in harm, loss or detriment to the person.

While the web, through social media, online networking sites, etc. presents an opportunity for people to build useful connections, construct personalized profiles and so on, where they can express their personality, thoughts, feelings and other personal values (e.g., interests, opinions, livelihood, place of work, relationship status, sexual orientation, religion, etc. [11, 17, 23]), using these platforms also leaves people open and vulnerable to potential exploitation and harm. This is often because small pieces of information shared online, across multiple networks and websites, individually seem innocuous. However over time, shared information may combine and link together as digital traces enabling greater inferences to be drawn about the individual. For example, Taylor may post messages that indicate that they live alone. Meanwhile their jogging GPS data posted online shows the routes and times that Taylor runs. Taken together, one can infer where Taylor lives, when and where Taylor routinely goes, and that no one is waiting for them at home! Thus, shared data over time can reveal cumulatively more about one’s identity, habits, work/life patterns, personality etc. than a person intends, which may result in loss of privacy, or worse. Clearly, such risks can have potentially negative and even disastrous consequences, for an individual such as stalking [15], identity theft [1], financial loss [2], damage to reputation [6], cyber-bullying [5]; for an employer (e.g., by creating opportunities for cyber-crime, damage to corporate reputation, etc.); or even for national security (e.g. by revealing deployment details, security access, etc.) [10].

Each of the risks above represent potential search scenarios and motivations that different actors could undertake. An employer may screen potential employees by searching through social media accounts to see if they can amass a picture of the candidate and whether they have a track record of (in)appropriate behaviour. Meanwhile, a hacker may be more interested in collecting details that could be used to socially engineer access to the person’s account or place of work. In practice, exploring and investigating such scenarios is particularly challenging for a number of reasons (e.g. privacy issues regarding sharing the data, ethical issues about exposing individuals, curating profiles that contain sufficiently rich relationships, etc.[18, 25]). To overcome these, we have developed a bespoke test collection (albeit small by contemporary social media standards) of fictitious personas. These contain curated personal posts, profiles, web pages, blog posts and so forth. Research participants are then given a brief describing a particular search scenario before being invited to inspect, explore and search the information presented. To further enable exploration, we have developed a tool that enables participants to search and browse the same collection for each scenario, rate and annotate individual pieces of information, first individually (evaluating each piece of information on its own), and then collectively (to draw inferences across the presented information). This process enables us to assess and evaluate each participant’s efficacy in identifying information that taken together

could increase someone’s risk of harm. Our demonstrator is broadly positioned towards raising awareness within the general public, but is also applicable for providing training to employees in workplace operational security and for educating young people in online risks more generally.

1.1 Motivation and Background

Personal online cyber-safety presents many challenges to individuals as their digital footprints span and encompass multiple different sites and platforms. While people say that they understand the need to protect their privacy and security online, they often do not take the necessary steps to do so [7, 12]. In addition, as people become accustomed to searching and browsing other people’s information online, they are likely to underestimate how their own online sharing behaviors “give off” insights about them to others, or even feel that such sharing practices are the norm [3, 4, 22]. Our research found that even among those participants who profess to be digitally literate, many struggled to recall what they had shared, or envision potentially harmful future scenarios emanating from their digital traces [20]. In other studies, even large-scale data violations e.g., Cambridge Analytica, which led to increased sensitivities around information sharing, did not significantly improve reported “digital hygiene” practices [22]. So; how can people use social media and other online platforms to enjoy their benefits while minimising their risk of negative or unintended consequences? One solution [10, 14] comprises a personal informatics system that enables people to examine and reflect upon details that they are sharing online, to increase awareness of their privacy risks. For example, *DataSelfie*¹ provided a browser plug-in for Facebook that warned individuals what their online interactions might reveal about them, while the *WASP* [16] personal web archive and search system prototype, integrated archiving, indexing, and reproduction technology into a single application. In our prior research, *DataMirror* [13], we aimed to use people’s social media content and invite them to inspect and explore aspects – such as the number and frequency of posts and any sentiment these conveyed, etc. However, such solutions are often technically problematic and require the consumption and ingestion of feeds across different APIs, and resulting data indexing and making sense of. Furthermore, while someone’s online information can comprise thousands of individual posts that have been shared over many years, it may comprise only a small proportion of instances from widely recognised potential risks. Subsequently, enabling self-reflection on participants’ online information, while potentially insightful, may provide insufficiently holistic awareness of wider risks, leaving them vulnerable, and undermining a tool’s intent. Moreover, multiple ethical concerns arise around managing personal data and resulting acquired collections. These could be used to automate discovery of security vulnerabilities in individuals’ digital traces with the potential to be exploited more widely by nefarious actors. We addressed this by creating fictitious personas and curated collections of their online information. Our tool has dual motivations: exploring people’s information-seeking and sharing behaviours and practices and their assessment of risks stemming from combining pieces of information (*cumulative revelations*); and to support reflection on personal online practices.

¹<http://dataselfie.it/>

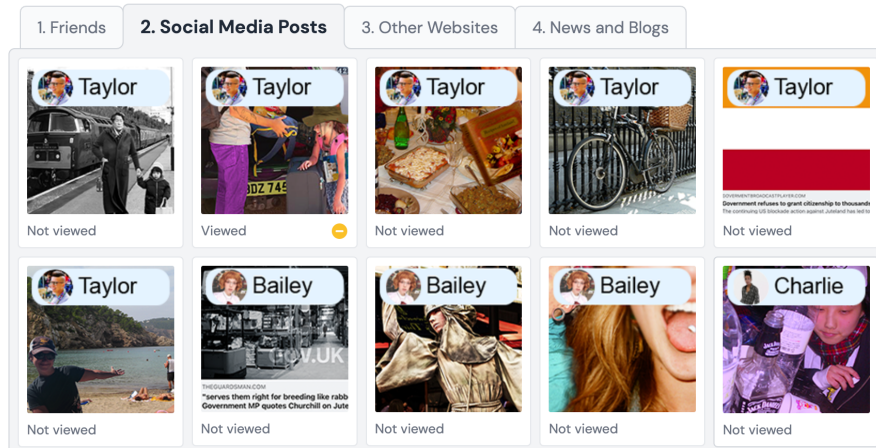


Figure 1: Search result page in the demonstrator containing verticals of (1) friends and followers, (2) social media posts (3) other websites, and (4) news and blog sites. The participant can click on the thumbnails to view the result page.

2 DEMONSTRATOR

The scenarios and tool design build upon outcomes from our previous work [19–21]. This involved a data narrative inquiry into people’s awareness of risks that stem from everyday online information sharing practices across their personal and working lives. During interviews, participants described their practices and also conceptualised them in ideographic form, by sketching on paper. The study found that many participants adopted incomplete risk models when assessing dangers, associating these with individual pieces of information, rather than arising from their connections or summation [19, 20]. When challenged to think about how a motivated hostile actor could make them vulnerable to future risks, participants found it difficult to envision any such scenario. Relatedly, we found that some described their online self as boring, rationalising that their online information was of no interest or value to other actors.

The data narrative study findings and those from our earlier prototype [13] confirmed that people find it difficult to reflect upon their own digital traces or readily recognise existing or conceptualise potential future risks. We thus decided that it could be more powerful to fabricate bespoke personas, enabling a richness of scenario that curated the required range of common risks for participants to reconstruct. Moreover, this avoided ethical and privacy issues associated with managing personally identifiable digital trace data, providing a responsible research design, while ensuring sufficient variety of content and risks to enable an engaging participant experience and sense of discovery.

2.1 Scenarios

To date, we have developed two personas with associated collections of posts, pages, blogs, news articles, etc. as though created and shared by the person of interest, or by their associated others. Each collection consists of approximately 50 items, enabling around 2500 (50 × 49) possible combinations of dependent judgements – each post potentially combined with each other post enabling greater inference. The scenarios meanwhile employ ambiguity, narrative and

gameplay to provoke curiosity and encourage exploration across the individual items. While the total number is relatively small at 50, their possible combinations is large enough, and manageable for individual research participants to explore without becoming cognitively overloaded. For the two personas, we developed different scenarios or topics. For Taylor Addison these comprise:

- **Identity Theft:** Taylor has discovered usual activity on one of their accounts and wonders if they have been the victim of a hacker. Taylor asks you for help in indicating what public information, such as Taylor’s date of birth, etc., could have been used by hackers to access Taylor’s accounts.
- **Unwanted Attention:** Taylor is feeling paranoid, looking over their shoulder sensing someone is physically following them. You are invited to investigate to what extent Taylor’s movements can be gleaned from across the online platforms Taylor regularly uses.
- **Lost Employment Opportunity:** Taylor has received a call saying they weren’t offered a coveted job. Feeling perfect for the role, Taylor is left wondering if the recruiter had surreptitiously searched Taylor’s online posts, and found something that didn’t align with company values. You are enlisted as a critical friend to evaluate Taylor’s traces through the eyes of a potential employer.
- **Political Victimisation:** Taylor has strong beliefs about people’s right to live in a free and democratic world. While Taylor thinks that trolls and paid internet commentators have been targeting their friends, leading to Taylor’s online harassment, you think Taylor’s recent apparent persecution might also stem from Taylor’s own online activities and check these out to see.

For each scenario, the participant is challenged to search through all the profiles, posts and pages to identify individual items of potential risk. Each scenario additionally contains items that can be linked, enabling specific greater inferences. These exposing connections between items are the gold standard judgement outcome for each task, and used as the basis of our task-evaluation.

2.2 Interface

Below, we describe the main pages of the demonstrator.

Search Result Page Participants are presented with a search result page (see Figure 1) that divides and presents posts, pages, profiles, and sites into groups: friends and followers; social media posts (from e.g. “Friendbook”, “Tweeta”, “InLinked”, etc.), other sites (e.g. government websites with open data, open fitness app data, etc.) and news (e.g. articles from online newspapers, blogs, etc.).

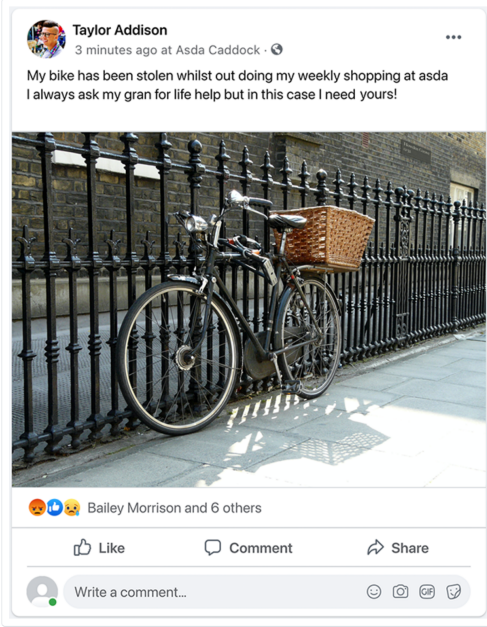
Result Annotation Page On clicking on a result, participants are taken to a page where they are invited to inspect and assess the selected item (see Figure 2), before rating its relevance as a matter of concern (i) to the posed scenario or (ii) for some other reason. For both questions, participants rate the item: (0) no, (1) possibly or (2) yes. We included a middle ground of possibly – rather than relying on binary relevance – so that participants who are unsure can flag their concern. This is because the nature of relevance in these scenarios is conditional – a post may only become relevant in light of subsequent information found. Participants can re-visit and revise their ratings in light of the new information. We have thus included a tool function that tracks the number of times each post is visited, and whether participants later change their rating of a particular item. We have also included a free text box where participants can note the nature of their concern.

Post Scenario Annotation Page After completing the scenario, for those posts that participants marked as concerning a subsequent rating page is presented (see Figure 3). Here, participants are asked to grade the relevance of items when taken together, thus providing an assessment of risk across items both individually and cumulatively. Participants can also input a free text description of their concerns regarding the sum of the information found.

Relevance and Risk Assessments The tool allows us to capture the order that each participant inspected each item, the amount of time spent reviewing each item, how many times participants visit and then perhaps revisit an item, along with participants’ relevance and risk assessments (including changes to their decisions). While our demonstrator focuses on scenarios specifically for risks associated with online cyber-safety, it could also be used to capture assessments for other scenarios where relevance is conditional and dependent. For our scenarios, we are particularly interested in understanding how participants explore and rate items during interactive search tasks, and specifically, how well they are able to identify sets of relevant (risky) items. After the scenario is completed, participants are provided with a debriefing page explaining the relationships between posts according to our gold standard judgements.

3 SUMMARY AND FUTURE WORK

People’s online profiles, posts and pages, whether constructed by themselves, friends or others, create digital traces that can be searched and explored by other actors. This motivates people-based search tasks – an area largely under-investigated in Information Retrieval, except perhaps in the context of Expert Search [8] and Celebrity Profiling [9]. Research and development in this area, which involves the processing of personally identifiable data, is fraught with technical, legal and ethical challenges. An additional challenge arises as an individual’s collection of profiles, pages, posts



The image shows a social media post by Taylor Addison, posted 3 minutes ago at Asda Caddock. The post text reads: "My bike has been stolen whilst out doing my weekly shopping at asda I always ask my gran for life help but in this case I need yours!". Below the text is a photograph of a bicycle with a basket, parked against a black metal fence. The post has received 6 likes from Bailey Morrison and 6 others. Below the post, there is an annotation form with two questions and a text box for comments.

Do you think this post might reveal information that could be used to hack Taylor's account or steal Taylor's identity?
 No Possibly Yes

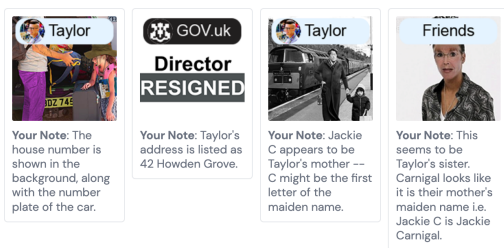
Do you think that this post is concerning for other reasons?
 No Possibly Yes

If you think this post is of concern, please explain below.

Figure 2: Page Annotation. Participants flag and note the concerns with the post – which they can revise during the course of the task.

etc may not fully represent the spectrum of risks created via cumulative revelation, reducing the value of using a person’s own data for exploratory purposes. Questions therefore remain regarding how we can create and build large-scale re-usable test collections to examine, explore and investigate how people search for and about people.

In our presented work, we have begun to probe and examine this phenomenon, albeit using a carefully curated collection. Our approach not only allows us to explore how well people can piece information together to form cumulative revelations, but also to examine the concept of “*conditional and dependent relevance*” from another perspective. Meanwhile, the designed tool aims to provide a novel and engaging cyber-safety educational experience by raising awareness of potential risks, consequences and harms, through gameplay and, to some extent, gamification of the search and annotation process. Raising awareness and understanding of these issues is important both at an individual and societal level. With this tool, we believe that we can measure individuals’ competencies around identifying potential risks, taking in both online and offline contexts.



Given these posts that you found, please tell us why you think they may risk giving away information about Taylor's identity to hackers and thieves.

How risky would you rate these posts, on average?
 Not Risky Extremely Risky

Taking these posts together, how risky would rate these posts, overall?
 Not Risky Extremely Risky

Figure 3: Post Scenario Annotation. Participants then rate the posts collectively on a graded scale in terms of risk.

By presenting this demonstrator, we hope to: (i) elicit valuable feedback towards developing and applying the collection and tool to other scenarios and contexts (ii) gain new insights into how people search for and about people, and (iii) better understand the conditional and dependent nature of relevance when embedded in such contexts. In future, with respect to our primary goal of understanding how well individuals can identify risks in people's digital traces, we plan to conduct further user studies evaluating people's search behaviours and their ability to connect information together and, moreover, understand whether engaging with such scenarios leads to greater awareness and long term changes to people's information behaviours and practices, towards their improved cyber-safety.

ACKNOWLEDGMENTS

Cumulative Revelations of Personal Data This project is supported by the UKRI's EPSRC under Grant Numbers: EP/R033889/1, EP/R033889/2, EP/R033897/1, EP/R033854/1, EP/R033870/1. The personas in this work are fictional. Any similarity to actual persons, living or dead, or actual events, is purely coincidental.

REFERENCES

- [1] Alessandro Acquisti and Ralph Gross. 2009. Predicting social security numbers from public data. *Proceedings of the National academy of sciences* 106, 27 (2009), 10975–10980.
- [2] Angeliki Aktypi, Jason RC Nurse, and Michael Goldsmith. 2017. Unwinding Ariadne's identity thread: Privacy risks with fitness trackers and online social networks. In *Proceedings of the 2017 on Multimedia Privacy and Security*. 1–11.
- [3] DM Boyd. 2008. *Taken out of context: American teen sociality in networked publics*. Ph.D. Dissertation. <https://search.proquest.com/openview/9cc930ef134daf46c17434d2992e8251/1?pq-origsite=gscholar&cbl=18750>
- [4] Danah Boyd. 2009. Why youth (heart) social network sites: The role of networked publics in teenage social life. *papers.ssrn.com* (2009). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1518924

- [5] Xiongfei Cao, Ali Nawaz Khan, Ahsan Ali, and Naseer Abbas Khan. 2020. Consequences of cyberbullying and social overload while using SNSs: A study of users' discontinuous usage behavior in SNSs. *Information Systems Frontiers* 22, 6 (2020), 1343–1356.
- [6] Hongliang Chen, Christopher E. Beaudoin, and Traci Hong. 2016. Protecting oneself online: The effects of negative privacy experiences on privacy protective behaviors. *Journalism and Mass Communication Quarterly* 93, 2 (2016), 409–429.
- [7] Lynne M. Coventry, Debora Jeske, John M. Blythe, James Turland, and Pam Briggs. 2016. Personality and Social Framing in Privacy Decision-Making: A Study on Cookie Acceptance. *Frontiers in Psychology* 7 (2016). <https://doi.org/10.3389/fpsyg.2016.01341>
- [8] Nick Craswell, Arjen P De Vries, and Ian Soboroff. 2005. Overview of the TREC 2005 Enterprise Track.. In *Trec*, Vol. 5. 1–7.
- [9] Walter Daelemans, Mike Kestemont, Enrique Manjavacas, Martin Potthast, Francisco Rangel, Paolo Rosso, Günther Specht, Efstathios Stamatatos, Benno Stein, Michael Tschuggnall, et al. 2019. Overview of PAN 2019: bots and gender profiling, celebrity profiling, cross-domain authorship attribution and style change detection. In *International conference of the cross-language evaluation forum for european languages*. Springer, 402–416.
- [10] Judson C Dressler, Christopher Bronk, and Daniel S Wallach. 2015. Exploiting military OpSec through open-source vulnerabilities. In *MILCOM 2015-2015 IEEE Military Communications Conference*. IEEE, 450–458.
- [11] Oliver L. Haimson, Jed R. Brubaker, Lynn Dombrowski, and Gillian R. Hayes. 2016. Digital footprints and changing networks during online identity transitions. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, 2895–2907.
- [12] Cormac Herley. 2009. So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In *Proceedings of the 2009 Workshop on New Security Paradigms Workshop (NSPW '09)*. Association for Computing Machinery, New York, NY, USA, 133–144. <https://doi.org/10.1145/1719030.1719050>
- [13] Amal Htait, Leif Azzopardi, Emma Nicol, and Wendy Moncur. 2020. *DataMirror: Reflecting on One's Data Self: A Tool for Social Media Users to Explore Their Digital Footprints*. Association for Computing Machinery, New York, NY, USA, 2125–2128. <https://doi.org/10.1145/3397271.3401398>
- [14] Danesh Irani, Steve Webb, Kang Li, and Calton Pu. 2011. Modeling unintended personal-information leakage from multiple online social networks. *IEEE Internet Computing* 15, 3 (2011), 13–19.
- [15] Puneet Kaur, Amandeep Dhir, Anushree Tandon, Ebtesam A. Alzeiby, and Abeer Ahmed Abohassan. 2021. A systematic literature review on cyberstalking. An analysis of past achievements and future promises. *Technological Forecasting and Social Change* 163 (2021), 120426. <https://doi.org/10.1016/j.techfore.2020.120426>
- [16] Johannes Kiesel, Arjen P de Vries, Matthias Hagen, Benno Stein, and Martin Potthast. 2018. WASP: web archiving and search personalized. (2018).
- [17] Hanna Krasnova, Sarah Spiekermann, Ksenia Koroleva, and Thomas Hildebrand. 2010. Online social networks: Why we disclose. *Journal of Information Technology* 25, 2 (jun 2010), 109–125.
- [18] Luis A. Leiva, Ioannis Arapakis, and Costas Iordanou. 2021. My Mouse, My Rules. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. ACM. <https://doi.org/10.1145/3406522.3446011>
- [19] Callum Nash, Daniel Carey, Emma Nicol, Amal Htait, Burkhard Schafer, Jo Briggs, Wendy Moncur, and Leif Azzopardi. 2022. Making Sense of Trifles: Data Narratives and Cumulative Data Disclosure. In *Proceedings of RIIS22 INTERNATIONALES RECHTSINFORMATIK SYMPOSIUM*. 8p.
- [20] Emma Nicol, Jo Briggs, Wendy Moncur, Amal Htait, Daniel Carey, Leif Azzopardi, and Burkhard Schafer. 2022. Revealing Cumulative Risks in Online Personal Information: A Data Narrative Study. *PACM HCI* (2022).
- [21] Emma Nicol, Amal Htait, Leif Azzopardi, and Wendy Moncur. 2021. Towards identifying, understanding and controlling cumulative revelations in social media. *Proceedings of the Association for Information Science and Technology* 58, 1 (13 Oct. 2021), 798–800. <https://doi.org/10.1002/pri.2566> 84th Annual Meeting of the Association for Information Science and Technology (ASIST) ; Conference date: 29-10-2021 Through 03-11-2021.
- [22] OfCom. 2019. *Ofcom Adults' Media use and attitudes report*. Technical Report. https://www.ofcom.gov.uk/_data/assets/pdf_file/0021/149124/adults-media-use-and-attitudes-report.pdf
- [23] Ning Xia, Han Hee Song, Yong Liao, Marios Iliofotou, Antonio Nucci, Zhi-Li Zhang, and Aleksandar Kuzmanovic. 2013. *Mosaic: Quantifying Privacy Leakage in Mobile Networks*. 564 pages.
- [24] ChengXiang Zhai, William W Cohen, and John Lafferty. 2015. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *Acm sigir forum*, Vol. 49. ACM New York, NY, USA, 2–9.
- [25] Steven Zimmerman, Alistair Thorpe, Jon Chamberlain, and Udo Kruschwitz. 2020. *Towards Search Strategies for Better Privacy and Information*. Association for Computing Machinery, New York, NY, USA, 124–134. <https://doi.org/10.1145/33343413.3377958>