Introduction

Metadata is not value-free. Access to metadata and the way it is used or misused has a profound effect on the opportunities and rights of citizens. The creation and use of metadata raises issues of governance, privacy, security and human rights (Haynes 2018).

Understanding metadata

Metadata has become a somewhat loose term recently and there is some debate about the distinction between metadata and data. All metadata is data, but not necessarily the reverse. If we make the distinction between data about things, and data about information resources, metadata, we have a working definition. A catalogue of books in a library, for instance, is a collection of metadata, because a library book is an information resource. Other commentators such as Zeng and Qin (2015) propose a wider, more inclusive definition of metadata where the distinction between data and metadata is less clear.

Examples of metadata

When we talk about metadata we could be talking about (cataloguing) data that describes resources such as bibliographic materials. We could be talking about the results of social media activity such as postings on Facebook or completion of an online survey or quiz. We could also be talking about information-containing artefacts such as museum objects. A museum object has to have documentation associated with it to verify its authenticity (is it what it purports to be?) and its context (how does this fit into a cultural or historical movement?). Then there are records of digital transactions. The data about phone calls, for instance, could be classed as metadata. It could also apply to data about our online activity. Let us look into some into some of these examples of metadata use in more detail.

Bibliographic materials

In what is perhaps a more familiar context for many readers, the library catalogue can be seen as a repository for metadata. The electronic publications services and secondary information services for abstracting and indexing journal articles are all about metadata. The information resources being described may include physical and digital materials. As well as data to help users retrieve information resources, there is also the metadata used to manage these resources, including rights management and transactions associated with individual library users.

Security and intelligence services

The security and intelligence services make use of metadata as has been famously reported during the Edward Snowden revelations (Greenwald 2013). Indeed Privacy International (2017) has identified the following types of metadata that is gathered or could be gathered by security agencies: Location; Device used; Date/time; Sender; Recipient; and Length of call. By requiring telephone companies to hand over records (the metadata) of international telephone calls made by US citizens, the National Security Agency in the United States unleashed a major controversy over the scope of the Fourth Amendment (United States n.d.).

Other service providers have been required to hand over transactional data to the security and intelligence services and the requests themselves have been subject to gagging orders, although some companies have tried to report levels of information requests from security services (Apple Inc. 2013). Another concern is the combination of data from different sources. "*Metadata in aggregate is content*" as Jacob Appelbaum observed during the Wikileaks controversy (Democracy Now 2013). In other words when metadata from different sources is aggregated it can be used to reconstruct the information content of individual communications.

Advertising and metadata about online activity

Advertisers make extensive use of metadata generated by the use of cookies and other tracking technology (Sipior et al. 2011; Gervais 2014; Pierson & Heyman 2011). This maps the characteristics of an individual with their online behaviour – which sites they visit, and what transactions they follow on the site. This metadata is the basis of revenue generation for the largest corporations in the world. It provides the economic model for platforms such as Facebook and Twitter. Google, Bing, Yahoo! and other search engines also keep details of every search done via their search engine (Waller 2011). This metadata is associated with individuals, and is available on request if an EU citizen puts in a data subject access request.

Ethical issues

Metadata use raises several distinct ethical issues from an individual perspective. They are: privacy, power, and information inequality. There are ethical issues that are focused more on a corporate perspective – particularly information governance.

Privacy

The association between personal data and transactional metadata raises information privacy as an important ethical consideration. This was a problem faced by libraries in the Unites States in the aftermath of 9/11. The Homeland Security services wanted to identify who was reading books that they considered were indicative of sympathy for the terrorists cause. The American Library Association fought and campaigned hard to make sure that librarians were well-briefed about privacy rights and the legality and extent of requests for information about the reading habits of individuals (ALA 2017). On-going work by CILIP and the Carnegie UK Trust have increased awareness of the responsibilities of librarians as well as helping to educate individuals about more general online safety issues (Carnegie UK Trust 2018). Campaigning groups such as Privacy International (Anon n.d.), the Open Rights Group (2018) and the Electronic Frontier Foundation (2018) actively campaign around online safety issues.

Power

The power difference between consumers and service providers or governments is another ethical challenge. This is a theme that has been picked up elsewhere (Marques & Pinheiro 2014). The collection, control and use of transactional metadata by social media platforms and search engine providers gives them an immense amount of control over the user's online experience. Eli Pariser's (2011) book '*The Filter Bubble'* describes how we can end up in our own 'bubble' of news items and search results that reinforce our existing world-view and prejudices. Apparently we do not like to have our views challenged when we are online. The types of ads displayed, the news items that are highlighted and even the search results themselves are all a result of the behind-the-scenes configuration (or manipulation) of your online experience. Much of the time this may be seen as a way of making the experience more convenient and palatable. Julia Angwin (2014) talks eloquently about her experiences of extricating herself from Google and the consequence loss of convenience in her book '*Dragnet Nation'*.

The recent Windrush scandal, where it became evident that the Caribbean-born children of the Windrush immigrants were being systematically discriminated against because they could not prove their right to remain in the UK, is another aspect of the power of metadata (Gentleman 2018). There was a tremendous injustice against someone who came to the UK as a ten-year old, as British subjects and then having to prove that they were legitimately in the UK. Part of the problem was that many of the landing records of their arrival had been disposed of and the Home Office had not kept any metadata to 'prove' the existence of the records, let alone their content. Lack of access to information is one way in which people can be marginalised.

Quality

This brings us to the other major ethical challenge that we face, that of the disparity between information-rich and information-poor communities. At one time that was focused on access to the internet, and that is probably still true in many parts of the world.

Even in societies where there is widespread access via the public library system, such as in the UK for instance, there is still a problem of lack of access to good quality information. There is no universal quality mark to say that something is from a reliable source and is therefore likely to be true. We build up our own internal list of trusted sources. However when we pick up news items from social media sites or delivered to us by search engines, it is easy to take this at face value. This was precisely the problem that has occurred in recent elections. Of course it would be naïve to expect impartial and factual reporting by supporters of different political parties, but when the news item apparently comes from independent news sites, it is very easy to assume that they are impartial. Many of the sites making more outlandish claims about candidates in the US presidential elections in 2016 were probably focused on trying to generate traffic through their site to maximise advertising revenue, rather than supporting a particular candidate. However their activity had a significant effect on the election outcome.

Initiatives such as CILIP's Facts Matter campaign and Dave Clarke's Post-Truth Forum have started to provide some of the tools needed to address these issues (CILIP 2017; Clarke 2018). Bibliographic tools such as citation analysis and web analytics can also be used to detect possible fake news by looking at the patterns of connections and transactions associated with a news story. This is particularly relevant in forums such as Twitter. There is still a lot of work to be done on information quality and it will be interesting to see what research and initiatives might be directed towards establishing a methodology for assessing information quality. This is relatively well-developed for academic, bibliographic citations. Could this approach be applied to more general news items on web sites or on news channels? This would be a very interesting metadata initiative.

Role of information professionals

So what role can the library and information profession play when it comes to information ethics? CILIP and the Carnegie UK Trust have already done a great deal to raise the issue (Carnegie UK Trust 2018). Groups such as ISKO UK (2018) and the Society of Chief Librarians (2018) have made it a feature of their campaigning and meetings and it looks as though these issues are here to stay. Information and media literacy play an important role. This equips people to evaluate information reliability for themselves. Currently there is a great deal of focus on 'fake news', but further work is needed on assessing information quality as well as educating people about effective use of metadata resources.

References

ALA, 2017. Issues and Advocacy: Privacy. Available at: http://www.ala.org/advocacy/privacy [Accessed June 7, 2018].

Angwin, J., 2014. *Dragnet Nation: a quest for privacy, security, and freedom in a world of relentless surveillance*, New York: Times Books, Henry Holt and Company.

Anon, Privacy International. Available at: http://www.privacyinternational.org/.

Apple Inc., 2013. *Report on Government Information Requests*, Cupertino, CA. Available at: http://www.apple.com/pr/pdf/131105reportongovernmentinforequests2.pdf [Accessed June 11, 2013].

Carnegie UK Trust, 2018. Online Data Privacy. Available at: https://www.carnegieuktrust.org.uk/project/online-data-privacy/ [Accessed June 7, 2018].

CILIP, 2017. Facts Matter. Available at: https://archive.cilip.org.uk/advocacy/facts-matter [Accessed March 16, 2018].

Clarke, D., 2018. Post-Truth Forum. *Synaptica*. Available at: https://www.posttruthforum.org/ [Accessed April 16, 2018].

Democracy Now, 2013. Court: Gov't Can Secretly Obtain Email, Twitter Info from Ex-WikiLeaks Volunteer Jacob Appelbaum. Available at: https://www.democracynow.org/2013/2/5/court_govt_can_secretly_obtain_email [Accessed March 21, 2017].

EFF, 2018. Electronic Frontier Foundation. Available at: https://www.eff.org/ [Accessed June 7, 2018].

Gentleman, A., 2018. The Children of Windrush: 'I'm here legally, but they're asking me to prove I'm British'. *The Guardian*. Available at: https://www.theguardian.com/uk-news/2018/apr/15/why-the-children-of-windrush-demand-an-immigration-amnesty [Accessed June 7, 2018].

Gervais, N., 2014. Governmental Internet Information Collection: Cookies Placing Personal Privacy at Risk. *Bulletin of the Association for Information Science & Technology*, 40(2), pp.27–31.

Greenwald, G., 2013. NSA Collecting Phone Records of Millions of Verizon Customers Daily. *The Guardian*. Available at: http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order [Accessed July 7, 2014].

Haynes, D., 2018. *Metadata for Information Management and Retrieval: understanding metadata and its use* 2nd ed., London: Facet Publishing.

ISKO UK, 2018. ISKO UK. Available at: www.iskouk.org [Accessed June 7, 2018].

Marques, R.M. & Pinheiro, M.M., 2014. Informação e Poder na Arena da Internet (in Portuguese). *Informação e Sociedade: Estudos*, 24(1), pp.47–60.

ORG, 2018. Open Rights Group. Available at: https://www.openrightsgroup.org/ [Accessed June 7, 2018].

Pariser, E., 2011. *The Filter Bubble: what the Internet is hiding from you*, London: Viking.

Pierson, J. & Heyman, R., 2011. Social Media and Cookies: challenges for online privacy. *info*, 13(6), pp.30–42.

Privacy International, 2017. Privacy 101. Metadata. Available at: https://www.privacyinternational.org/node/53 [Accessed March 23, 2017].

SCL, 2018. Society of Chief Librarians. Available at: http://goscl.com/ [Accessed June 7, 2018].

Sipior, J.C., Ward, B.T. & Mendoza, R.A., 2011. Online Privacy Concerns Associated with Cookies, Flash Cookies, and Web Beacons. *Journal of Internet Commerce*, 10(1), pp.1–16.

United States, *U.S. Constitution Amendment IV*, United States.

Waller, V., 2011. Not just information: Who searches for what on the search engine Google? *Journal of the American Society for Information Science & Technology*, 62(4), pp.761–775.

Zeng, M.L. & Qin, J., 2015. *Metadata* 2nd ed., Chicago, IL: ALA Neal-Schuman.

Biography

David Haynes is a postdoctoral research fellow in the Department of Library and Information Science at City, University of London. His current research is funded by the Royal Academy of Engineering and the Government Office of Science intelligence community initiative. He is investigating online behaviour and what it reveals about the decisions people make when they disclose personal data online. His research blog is available at: https://blogs.city.ac.uk/privacycalculus/

You can follow David on Twitter @JDavidHaynes