Ontological Friction: ways in which libraries & archives limit access to personal data

Paul Pedley, David Bawden and Lyn Robinson

to be published in IFLA Journal 2018 (accepted 13 December 2017)

Abstract

Ontological (or informational) friction is the term used by Luciano Floridi to describe the amount of effort that would be required for some agent to obtain, filter or block information about other agents in a given environment, by decreasing, shaping or increasing the level of friction. The aim is to find the optimal level of friction in the infosphere. This article considers a range of ways in which ontological friction is used by libraries and archives in order to control the flow of personally identifiable information. Techniques include obscurity (including practical obscurity), obfuscation, use of blocking to inhibit tracking mechanisms, as well as temporal frictions such as limiting the length of time personal data is held. We base our analysis on Floridi’s information ethics, and specifically on his conception of information privacy.

1. Introduction

Personal data flows back and forth as part of our day to day activities, and this is true of library usage just as it is in other aspects of life. When a library user checks out items at a self-service issue point, borrows an ebook, or makes use of a library computer to access the internet, their personal data moves back and forth in the infosphere¹. It could potentially be available to library staff, the library user, a vendor, a third party, or even a fourth party illegal entity.

Research aims and objectives

The research aims to establish whether it is possible for librarians to protect the privacy of their users, and if so, how. It intends to identify the key privacy issues facing the library & information sector, and the ways in which it is possible to control the extent to which personally identifiable data passes back and forth. It draws on Luciano Floridi’s concept of “ontological” or “informational friction” as a way of establishing the factors that determine how easily or not data is able to flow. At this stage it is not limited to one particular type of library; nor is it limited to the digital environment.

Some background to Floridi’s concept of “ontological friction” is provided, followed by a consideration of information privacy. The main part of the paper explores three types of ontological frictions – technological frictions, obscurity, and temporal frictions, to better understand the speed and extent to which personal data flows back and forth. A number of lessons from copyright are identified which could potentially be applied in a data protection context.

The paper refers primarily to legislation and cases drawn from within the UK, US and EU jurisdictional boundaries. Data protection laws in those geographic areas include both transnational agreements and national legislation. In the United States there are federal laws as well as the laws of each of the 50 states. For member states of the Council of Europe there is The Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Within the European Union there are the laws of each nation state, the Data Protection Directive and the Privacy and

¹ “Infosphere” – We are moving from living in the biosphere to the infosphere. Information is our environment. The infosphere is a newly-created digital space built by new technology (Floridi 2014). It should not be confused with cyberspace, because it encompasses online as well as offline and analogue domains.
Electronic Communications Directive. Once the General Data Protection Regulation comes into force there will be a greater degree of harmonization across the Union.

(Sedgewick 2017 p1514) argues that the current data privacy regulatory approach is poorly matched to transborder data flows, because it is based on outdated geographically specific privacy regulations and that these could be productively reconceived in an international trade framework.

It is important to note that the article was written right at the outset of a PhD research project, and as such the scope of the anticipated research, the theoretical framework and the methodologies to be used were still in the process of being decided at the time of writing.

1.1 Literature review

An initial literature review was undertaken which looked at privacy as a concept as well as privacy in a library context. A further literature search covering ontological friction was then undertaken. This found only a handful of items, none of them from a specifically library & information science context [Barn, Primiero et al (2015); Casanovas (2014); Floridi (2006), (2013), (2014), (2017); Gutwirth, Leenes et al. (2014); Hildebrandt (2011); Martin (2011); McGeeveran (2013); Pagallo (2010); Primiero (2016); Primiero, Athiappan et al. (2017); Strikwerda (2010); Taddeo, Vaccaro (2011); and Tavani (2008)]. One recent item from Bates (2017) at Sheffield’s iSchool looks at the politics of data friction.

1.2 Theoretical framework

This article is based around Floridi’s theory of ontological friction. Over 50 privacy theories were considered, including Nissenbaum’s (2011) theory of contextual integrity; Petronio’s (2002) theory of communications privacy management, Dinev & Hart’s (2006) privacy calculus theory, Richards’ (2015) theory of intellectual privacy, Anita Allen’s (2011) theory of unpopular privacy; J D Elhai’s (2017) anxiety model; and Jens Eric Mai’s (2016) datafication theory. All of the theories mentioned give interesting perspectives on privacy. Floridi’s theory of ontological friction was chosen as the main theory to use firstly because of its emphasis on informational privacy, which Koops et al (2017) recognise as being present in every type of privacy scenario; secondly because with its strong emphasis on digital ICT’s, it is well-suited to emerging privacy challenges and dilemmas; thirdly because he recognises the revolutionary impact of digital information and communication technologies; fourthly because he acknowledges the importance of data ethics which shifts the level of abstraction from an information-centric to a data-centric one (Floridi 2013 p230) and finally because there is very little in the literature about the theory, giving huge potential to further explore, develop and work on the broad concept, using the library and information sector as a case study within which to see how the theory works in practical terms.

1.3 Research methodologies

At the time of writing, the methodologies for the research hadn’t been finalised. The research problem in this case doesn’t lend itself to precise analytical techniques, and therefore it is anticipated that a Delphi study will be used. Delphi studies can’t produce a right or a wrong answer, only expert opinion. The research will consider the issues from the perspective of the individual, the group, and society as a whole. A second methodology might be a discourse analysis using three pieces of data protection legislation dating from 1981, 1995, and 2016. Discourse analysis is useful as a technique for examining social and cultural issues, and how these issues affect society as a whole, including the portrayal of social injustice, and how social groups might be misrepresented in discourse. Use of the three legislative texts from 1981-2016 could also potentially identify how things have changed over time.

Interviews and questionnaires may also be used at a later stage, as it is anticipated that the initial research methodologies will help identify what are the issues that need to be examined in more depth.
2. Background to Floridi’s concept of ontological friction.

Floridi (2014: 102) writes “The ethical problem of privacy has become one of the defining issues of our hyperhistorical time”, and uses the terms “informational friction” and “ontological friction” to refer to the forces that oppose the flow of information within a region of the infosphere.

Privacy is both a function of the informational friction in the infosphere and of the personal information disclosed and its nature. Given some amount of personal information available in a region of the infosphere, any increase or decrease in the level of informational friction will affect privacy: the lower the level of informational friction, the higher the accessibility to personal information about the agents will be and vice versa.

Floridi (2014: 105) says that ICTs have made privacy one of the most pressing issues in our society because they unquestionably affect informational friction. “ICT’s are more redrawing rather than erasing the boundaries of informational privacy” (Floridi 2013 p230)

The information revolution has had a profound impact upon libraries and the ways in which they deliver their services. Libraries rely heavily on technology: integrated library management systems, ebook platforms, RFID technology, self-service issue systems, online databases, and discovery services. Many libraries utilize cloud computing

Another dimension is the way in which users access services remotely on their own devices rather than doing so solely on equipment provided by and located within a bricks and mortar library. The increasing use of mobile technology raises issues and challenges for librarians around protecting the personally identifiable information of their users. But it should also be recognized that libraries face a number of privacy challenges that do not relate exclusively to technology. Libraries need to think about the privacy implications of any new services that they introduce, or where they make any changes to existing services. They should undertake privacy audits and evaluate their library privacy procedures on an ongoing basis. Principle 12 of the NISO privacy principles recognises this:

“Libraries, content, and software providers should continuously assess and strive to improve user privacy as threats, technology, legal frameworks, business practices and user expectations of privacy evolve” NISO (2015 n.p.).

In an interview Professor Floridi said “Quite obviously, our world has been transformed radically and deeply and continues to be transformed radically and deeply by the digital revolution” Green (2017: 16). He considers individuals as agents owning their information, not in terms of a possession such as a car, but in the sense of it being part of the individual’s very essence, as a constitutive part of their identity: one may still argue that an agent “owns” his or her information in the precise sense in which an agent is her or his information. “My” in “my information” is not the same “my” as in “my car” but rather the same “my” as in “my body” or “my feelings”: it expresses a sense of constitutive and intimate belonging, not of external and detachable ownership, a sense in which my body, my feelings and my information are part of me but are not my (legal) possessions, Floridi (2006: 39-40).

Floridi (2006: 39) says that interpreting the revolutionary nature of digital ICTs in an ontological way provides a fruitful approach to develop a robust theory of informational privacy. In the same way as the digital revolution is best understood as a fundamental re-ontologization of the infosphere, informational privacy requires an equally radical re-interpretation, one that takes account of the informational nature of human beings and their operations as social agents. Such re-interpretation is achieved by considering each individual as constituted by his or her information, and hence by

---

2 “Re-ontologizing” - a neologism that Floridi uses to refer to a radical form of re-engineering which not only designs, constructs or structures a system, but which also fundamentally transforms it intrinsic nature, its ontology or its essence Floridi (2010)
understanding a breach of one’s informational privacy as a form of aggression towards one’s personal identity.

Floridi’s analogy is a vivid one. It gets across the importance of one's personal information, and the potential harm that can arise when someone breaches another person’s information privacy. In terms of “my body” as personal information, Kulynych (2017) points out that “Unlike a medical record number or credit card number, genome sequences, unique and permanent, can’t be replaced when compromised, and sequence data are a wellspring of information about health risks, ancestry, and sometimes, unexpected parenthood”.

Personal data is non-rivalrous, the use (or processing) of your personal data by one individual or organization does not prevent it from being used by someone else. So, to take Floridi’s analogy of a car, if I “borrow” a neighbour’s car, then it is no longer available for the neighbour to use whilst I have it because there is only one car. Whereas with personal information, the concept of borrowing doesn’t make sense. How can I “borrow” someone’s personally identifiable information such as their date of birth, their mother’s maiden name, the name of their first pet, and the year their first child was born. The data subject still knows all of this information. Floridi therefore talks about the cloning or breeding someone else’s personal identity (2006: 39).

A good is considered non-rivalrous if, for any level of production, the cost of providing it to a marginal (additional) individual is zero. “Data have been described as a ‘non-rivalrous’ good in that they can be used for multiple purposes on multiple occasions without reducing their value for other users”, European Data Protection Supervisor (2014: 2).

There is also the idea in economics of a public good, which means a good that is both non-excludable and non-rivalrous in that individuals cannot be effectively excluded from use and where use by one individual does not reduce availability to others. Some might go so far as to argue the case that personal data is a public good, particularly in the case of health information.

3. Informational privacy

To put informational friction into context it is first necessary to consider the concept of “informational privacy”. Floridi (2014: 102) distinguishes four types of privacy: physical privacy, mental privacy, decisional privacy, and informational privacy. Freedom from informational interference or intrusion, namely “informational privacy” can potentially be increased by new or digital ICTs according to Floridi (2014).

Koops, Newell et al. (2017: 484) examine the main dimensions along which privacy can be classified. They identify eight categories (see schedule 1), which they then split into two headings: firstly, those with an emphasis on freedom from (being left alone); and secondly, those with an emphasis on freedom to (people having freedom for self-development).

In the category of “freedom from” they list bodily privacy, spatial privacy, communicational privacy and proprietary privacy. And in the category of “freedom to” they list intellectual privacy, decisional privacy, associational privacy, and behavioral privacy. The one obvious omission is “informational privacy”, and that is because they treat it as an “overlay” which runs across all of the other dimensions. It overlaps, but doesn’t coincide with the eight basic types.

---

3 Richards (2008), (2015) proposes “Intellectual privacy”. Allen (2011) says that intellectual privacy is a complex hybrid between associational and informational privacy: it encompasses what people read, think, plan, and discuss with their personal or business associates.
Koops, Newell et al. (2017: 555) distinguish between informational privacy relating to personal data and "physical privacy" understood in a broad sense. Under "physical privacy" they categorise bodily privacy, spatial privacy, communicational privacy, proprietary privacy, behavioral privacy, decisional privacy, and associational privacy. By "physical" they mean that the object of privacy can be watched. Informational privacy overlays all of these because Koops, Newell et al. (2017: 555) have in mind the information about each of them, saying "Often the protection of such information is also a precondition to protecting the underlying physical privacy type".

Informational privacy can't be reduced purely to information, because there is always an additional element CPDP (2017 n.p.). Koops, Newell et al. (2017: 487) push back against the trend to focus predominantly on informational privacy and data protection because that approach neglects the other types of privacy that nevertheless remain worthy of protection even in a digitized world.

Utilising the typology of privacy by Koops, Newell et al. (2017) we have provided examples for each of the privacy categories which are relevant to the work of the library & information sector (see Schedule 1).

4. Ontological frictions

This article explores technological frictions, obscurity, and temporal frictions. Other types of friction identified, but which aren't covered, include spatial frictions, sensory frictions, training & awareness, information behaviour, contextual issues, and regulatory frictions.

4.1 Technological frictions

Technological frictions include privacy enhancing technologies, privacy invasive technologies, decisional privacy, and privacy by design. Specific frictions can include, amongst other things, the use of encryption, technologies to limit disclosure of personal data, and the use of secure networks. This paper explores two themes identified by Floridi (2014: 105, 112): empowerment through digital ICTs, and anonymisation; and then looks briefly at the use of blocking to inhibit tracking mechanisms.

4.1.1 Empowerment

Floridi (2014: 114) asserts that in the infosphere, we as inforgs are increasingly empowered by new ICTs not only to gather and process personal data, but also to control and protect them (We are not becoming sci-fi cyborgs but re-appropriating ourselves as connected, informational organisms, inforgs (Floridi (2014:94-96)). He sees empowerment in terms of inclusion in the decision-making processes, as well as in the "more opportunities" sense, in terms of improvement in the quantity and quality of available choices, Floridi (2014: 113-114).

One example Floridi (2014: 114) gives of the way in which new ICTs enable people to control and protect personal data is how the reputation management companies are mushrooming. Michael Fertik is well known for pioneering the online reputation industry, Fertik, Thompson (2015). Citing reputation management companies as a form of empowerment begs the question, who is empowered by the availability of reputation management services. A reputation management company might successfully get embarrassing content to appear lower down the search engine rankings, to the point where it has to all intents and purposes disappeared. The problem is that this isn’t a one-off exercise - what happens when those search engines tweak their algorithm, and the story/stories appear more prominently again. Paying a company to continuously manage your online reputation can be expensive, and out of many people’s reach. It raises the issue of fairness, and whether the services of reputation management companies are only available to those who are well-off.
How is this relevant to libraries? Librarians have an important role to play in digital privacy literacy training for their users, to help them have a better understanding of how their digital footprint is created, what the implications are, and what can be done to minimize the data that is gathered; or to manage it more effectively to present a positive image of oneself online. If libraries do not do this, who else will? Brantley (2015 n.p.) believes that “public libraries are among the last protectors of privacy in contemporary society”, while Mattlage (2015: 76) says “Having special obligations to protect information rights means that information professionals must first of all take information rights seriously by defending them against countervailing pressures for more expedient public policies. It is the unique role of information professionals to be last to abandon the defense of these rights”.

It is questionable whether the means of empowerment available through digital ICTs are commensurate to the privacy threats posed by those same ICTs. It feels as though everyone needs to ask “How much effort am I willing to put into keeping my information private?”. One might go so far as to say that it requires a disproportionate amount of effort. Consider the invasions of privacy that are part of one’s everyday activity on the web:

- Personal data is made public by default
- Most people tend not to change the default settings
- To read the privacy policies for all the websites someone visits would take 76 days per year, Madrigal (2012: n.p.)
- The (US) national opportunity cost to read privacy policies is around $781 billion, McDonald, Cranor (2008: 541).
- Cloud computing makes it easy for governments to get at our personal data
- Governments do not want the information provider to tell users their data has been accessed, or how often
- An individual’s digital footprint is being continually added to all the time
- There is no way to guarantee absolute privacy. The best we can hope for is to minimize privacy risks

4.1.2 Anonymisation

Another way in which digital ICTs can increase informational privacy according to Floridi (2014: 105) is through anonymisation. However anonymisation is hard to achieve when there are correlation attacks and when in amongst millions of items of data someone having access to four random pieces of information can deanonymise over 90% of those records, Singer (2015 n.p.).

Green, Cunningham et al. (2017: 4) believe that because so much data is now available from a wide variety of sources, and because databases can be manipulated and combined in complex and unpredictable ways, information that might not be deemed personally identifiable information (PII) can lead to the identification of a specific individual and enable inferences to be made about that individual.

Indeed, perfect anonymisation is a myth. There is also a tension between the level of usefulness of the data and the risk of privacy being compromised: the more granular the data the more interesting and useful it is for businesses, for policymakers, for researchers and for the public; but the more detailed the information the greater the risk that personally identifiable and potentially highly sensitive information can be revealed.

Risks include:
- Re-identification
- False re-identification (When data is partially anonymous, individuals are at risk of having sensitive facts incorrectly connected to them through flawed re-identification techniques.)
- Jigsaw identification (The ability to identify someone by using two or more different pieces of information from two or more sources—especially when the person’s identity is meant to be secret for legal reasons)
- The Mosaic Theory of privacy (which suggests that having the whole picture reveals far more than the individual components of which it comprises)

Risk mitigation measures include removing low numbers, aggregating data sets, removing fields, removing records, generalizing data, adding noise, and creating anonymous identifiers.

It isn’t simply a question of whether the information that is made available contains anything that immediately identifies a particular living individual, because data protection legislation requires that you also take into account whether that information could potentially be combined with something else which together identifies the person. Article 4 (Definitions) of the GDPR says that “‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person”, European Union (2016).

For example, the New York City Taxi and Limousine Commission released a dataset containing the details about every taxi ride (yellow cabs) in New York in 2013, including the pickup and drop off times, locations, fare and tip amounts, as well as anonymized (hashed) versions of the taxi’s license and medallion numbers. From this Tockar (2014: n.p.) was able to identify the home addresses of frequent visitors to a strip club in the city.

The identification risk may increase over time given the increasingly low-cost availability of technical means to identify individuals in datasets and the increasing public availability of other datasets (such as through open data).

4.1.3 Use of blocking to inhibit tracking mechanisms

Another way in which libraries achieve informational friction is to install browser addons on computer terminals available for public use such as adblockers, or addons designed to protect privacy such as Privacy Badger, Ghostery, or Disconnect.

It is quite enlightening to install the Firefox add-on Lightbeam (https://www.mozilla.org/en-GB/lightbeam/) to see an interactive visualisation of the first and third party websites you interact with on the web. If, for example, you were to use a library discovery service to search for information, it is interesting to see what sites your information was shared with without your direct knowledge.

4.2 Obscurity, practical obscurity, and obfuscation

Obscurity as a means of ontological friction covers a lot of ground. Online obscurity, for example, covers areas such as search visibility, unprotected access, identification, clarity and the right to be forgotten. Practical obscurity covers issues around the difficulty of collecting the data, the difficulty of being able to aggregate the data, the degradation of data over time, interoperability, and how burdensome or unrealistic it is to obtain the information. Meanwhile, obfuscation is the deliberate use of ambiguous, confusing or misleading information to interfere with surveillance and data collection projects.

4.2.1 Obscurity
Hartzog, Stutzman (2013: 26) say “Empirical research demonstrates that Internet users rely on obscurity perhaps more than anything else to protect their privacy. Yet, online obscurity has been largely ignored by courts and lawmakers”. They believe that obscurity could serve as a compromise protective remedy: “instead of forcing websites to remove sensitive information, courts could mandate some form of obscurity”.

It isn’t simply a question of whether online information is “public” or “private”, but rather of how obscure it is. For Hartzog & Stutzman (2013: 32) “information is obscure online if it lacks one or more key factors that are essential to discovery or comprehension”. They identified four factors whose presence diminishes obscurity, and whose absence enhances it.

1. Search visibility is one of the most significant factors in online obscurity. If online information isn’t visible to search engines, if it cannot be discovered through search, then individual searchers have to rely instead on knowing and manually entering a URL, or being directed to the information through websites or email messages

2. Unprotected access - in determining whether information is private, courts will look at whether technology has been used to limit or restrict access. This could, for example, be in the form of a password, or through the use of privacy settings.

3. Identification – if it isn’t possible to link information to a particular individual then there is a reduced threat to that person’s privacy. The worry with big data is that it is easy to identify people even if they aren’t named.

4. Clarity – even if online information is easily discoverable, can it be easily understood. It may be incomprehensible as a result of the information being intentionally vague or incomplete.

Hartzog, Stutzman (2013: 2) outline a number of techniques internet users regularly use in order to make online information obscure:

- routinely hiding information by making it invisible to search engines (using the robots.txt file, although the Internet Archive announced in April 2017 that they were no longer intending to respect robots.txt commands, Graham (2017)
- using pseudonyms
- multiple profiles
- making full use of privacy settings
- use of confidentiality agreements
- relational knowledge
- true anonymity (encryption)

Libraries have as their core remit providing access to information, and their codes of ethics require them to do so whilst respecting the privacy and confidentiality of their users.

Section 3 of IFLA’s code of ethics, IFLA (2012) says that “Librarians and other information workers respect personal privacy, and the protection of personal data, necessarily shared between individuals and institutions.

The relationship between the library and the user is one of confidentiality and librarians and other information workers will take appropriate measures to ensure that user data is not shared beyond the original transaction”.

Page 8 of 19
CILIP’s Code of Professional Practice, CILIP (2012) says that one of a librarian’s responsibilities to information and its users is to “Protect the confidentiality of all matters relating to information users, including their enquiries, any services to be provided, and any aspects of the users’ personal circumstances or business”; while one of their responsibilities to society is to “strive to achieve an appropriate balance within the law between demands from information users, the need to respect confidentiality, the terms of their employment, the public good and the responsibilities outlined in this Code”.

In the case of archives, “Archivists promote and provide the widest possible accessibility of materials, consistent with any mandatory access restrictions, such as public statute, donor contract, business/institutional privacy, or personal privacy” and archivists “establish procedures and policies to protect the interests of the donors, individuals, groups, and institutions whose public and private lives and activities are recorded in their holdings”, Society of American Archivists (2016: n.p.).

In 2010 a campus information technology security audit was carried out at Duke University which looked at born-digital records. Information technology auditors were surprised by the volume of potentially sensitive digital materials that the Duke Archives staff had collected and stored on its servers. When archives staff pointed out that they had plenty of experience of managing personally identifiable information (PII), because their paper based records also contained sensitive data, the auditor’s response was to point out that those records relied on "security through obscurity" and that a thief seeking personal data would have to manually sift through the files page by page to access the data, whereas digital PII data was vastly different because scripts could be written to run looking across the data for credit card sequences and similar sensitive content, Goldman, Pyatt (2013: 38).

Mayer-Schönberger (2011: 126) describe three key characteristics of digital remembering:

- accessibility (shareability of data with others out of its original context)
- durability and
- comprehensiveness (the extent of one’s data now captured digitally and thereby "memorable")

To these three key characteristics, Brake (2014) adds a further two - namely how digital remembering is:

- unequal and
- evolving.

4.2.1.1 Library digitization projects & privacy

Library digitization projects can, potentially, impact the privacy of individuals significantly. They could decrease the "informational friction", and therefore the level of privacy, by increasing the discoverability of information, the ease of access, and ease with which a particular individual becomes identifiable. Consider the digitization of a collection of fanzines published well before the invention of the world wide web. It is true that they were “published”, but it would be unhelpful to equate that decision with deciding whether to make one’s information public or private, as though that were a binary decision with no gradations as to the level of privacy one might expect once an item has been published.
Zines are usually devoted to specialized and often unconventional subject matter. They are often a vehicle for radical voices. They could be a political zine, a feminist zine, an LGBT zine and so on. They are ephemeral in nature, and often have very small print-runs.

Fans may well have been comfortable having their formal legal name rather than a pseudonym appear in print on a fanzine cover. It is worth thinking of the decision in terms of Dinev & Hart’s (2006) privacy calculus theory, whereby an individual decides whether or not to disclose personal information based on a risk-benefit analysis. They would have compared the perceived risks and anticipated benefits. But therein lies the problem – how could they have made a truly informed choice in the period prior to the world wide web, one that would be valid for all time. They could not possibly have anticipated that their fanzine would one day be digitized, that the world wide web would be invented, and that search engines would index the contents of the publication so that their real name could quickly be found on the cover, in the table of contents, and in the ads circulated to market the zine. In other words, they allowed their real name to appear in print, fully in the expectation of a high level of privacy.

There is a need for searching and using fanzines with a degree of privacy and untraceability “rather than give the government fodder to harass them” Hedtke (2008: 41). By digitizing the fanzine, by making information about the zine maker readily available, there is a risk that it could lead to the authorities flagging them up as someone with an alternative viewpoint. The zine creator may still be living and digitization could have a negative impact on their personal and/or private life if their deepest secrets were archived in a research library, and subsequently digitized.

The curator of the Bingham Center Zine Collections at Duke University Libraries tells how “we have been contacted to remove a last name from our database that was associated with a zine title that the author felt damaged her reputation in her current career—at age 16, she had no idea that the flippan title would ever be available online” Wootten (2009).

The Zine librarians/archivists’ code of ethics says that they should strive to make zines as discoverable as possible while also respecting the safety and privacy of their creators, Zine Libraries Interest Group (2015: 16). An example of how zine libraries have increased the amount of “informational friction” is where they have set up separate public and private catalogues in order to keep certain information such as zine makers’ names more private.

Brake (2014: 107) gives an example of information published on the internet in 1986. He says: “in 1986 people who posted messages on the USENET newsgroup “net.motss” about gay-related issues generally did so under their own names (it was then both technologically difficult and even outside social norms to have anonymous email addresses). They might have been aware that the newsgroup could be archived and searchable but they surely could not have anticipated that, thanks to the world wide web and vastly increased storage and processing power, everything they posted has been digitally “remembered” from 1986 to the present day at https://groups.google.com/forum/#!forum/net.motss”

4.2.2 Practical obscurity

Practical obscurity is a concept which pre-dates the online world. It refers to impediments to data retrieval. The concept was articulated by the Supreme Court in U.S. Department of Justice v. Reporters Committee for Freedom of the Press 489 US 749 (1989). In evaluating the privacy of a “rap sheet” containing aggregated public records, the Supreme Court found a privacy interest in
information that was technically available to the public, but could only be found by spending a burdensome and unrealistic amount of time and effort in obtaining it. The information was considered practically obscure because of the extremely high cost and low likelihood of the information being compiled by the public, Hartzog, Stutzman (2013: 21).

In the Scottish case of Her Majesty’s Advocate v. William Frederick Ian Beggs (High Court of Justiciary 2001) (No2) (2002 S.L.T. 139) the judge ruled that information held on the internet archives of newspapers was published anew each time someone accessed it. However, he didn’t take the same view of the paper archives held by public libraries⁴. This distinction takes into account the ease with which material on the internet can be accessed as opposed to having to visit a physical library. In an appeal case involving the same parties [2010] HCJAC 27 the judgment makes a different point about online obscurity, saying “It appears also to have been accepted by both sides that those materials were archived material which had originally been published before the criminal proceedings became active on 21 December 1999 and that the action of entering the appellant's name into a standard search engine would not lead the searcher to these materials. Instead the searcher would have to go to the website of a particular newspaper or broadcaster and then search its archived material" (William Frederick Ian Beggs v Her Majesty’s Advocate [2010] HCJAC 27. 2010).

Brock (2016: 12) refers to practical obscurity in connection with the right to be forgotten. “The internet is massively accumulative, searchable in a second, and, despite the extensive decay of hyperlinks, can preserve information indefinitely. That changes the nature of debates about rectification, something legal scholars call ‘practical obscurity’ (referring to a piece of information not erased but hard to find), and forgetting”.

Ghezzi, Pereira et al. (2014: 3) acknowledge that oblivion has been incorporated in legal norms for some time now, but the tradition of the "right to oblivion" has been grounded on a "clean slate" which is linked, for example, to dignity and aspects of personality formation.

The right to be forgotten is a misnomer. It gives the impression of someone having a right for material to be removed. It isn’t a question of all traces of the information being removed from the internet, but rather a case of the information becoming harder to find as there is no longer a readily accessible link to it on the search engine. However, the level of obscurity depends on a number of factors:

- The geographic extent of the delinking (e.g. Was the link removed for the search engine’s UK version, across the European Union, or was it removed globally)
- Which search engines was it removed from (not all search engines currently operate a right to be forgotten takedown service; and it will depend whether it is removed from the ones which receive the most traffic, and that will vary from one country to another)

It is not the data in and of itself that it is important, what is more important is the ability to readily find the data by association with a personal name (see also (Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González. 2014)); because if it doesn’t appear in the first few pages of the search engine results, it is to all intents and purposes invisible to all but the most determined of internet users. The offending story may still be accessible on the original website, but to have been de-linked by the search engine. Floridi (2016) expresses a view which can be summed up as “the map determines the territory”. In other words, what matters

---

⁴ This case predates the Defamation Act 2013 where rules akin to the single publication rule apply to links to internet articles.
is the means to reaching the information and how discoverable it is, rather than the information itself.

Librarians need to think carefully before making private papers generally available. They have to consider ethical issues around privacy. In many cases the existence of copyright in unpublished manuscripts will ameliorate the problem, and if the individual concerned is the donor of the papers and has not placed restrictions on their use, there may be little to worry about. However, occasionally libraries are given material which has been illicitly obtained or come into the donor’s hands by chance, Fielding (1978: 188)

4.2.3 Obfuscation

Chen & Rea (2004: 87-88) identify three primary types of techniques used by online participants to manage identity disclosures:

1. the falsification of information shared online, which involves techniques such as using multiple email accounts, deleting cookies, and lying to websites.
2. passive reaction, which involves the use (or destruction) of technology that would connect a person to his or her online footprints.
3. identity modification, which involves the creation of gender-neutral avatar names, and the use of online identities that are disassociated from the personal identity.

Brunton, Nissenbaum et al. (2015: 1183) define obfuscation as “the deliberate addition of ambiguous, confusing, or misleading information to interfere with surveillance and data collection”. They put forward a manifesto for the digitally weak and powerless to protect their digital lives from monitoring, and argue that obfuscation is necessary to counteract information power imbalances when data about us is collected and used in ways we may not fully understand. Richards, Hartzog (2017: 6) believe, however, that while obfuscation is motivated by distrust, it creates further distrust by hiding from the surveillance economy and even intentionally feeding it with bad data. Richards & Hartzog believe that polluting the information economy may be a useful short-term solution, but that it is costly and even unsustainable in the long run. They believe that a better, more sustainable strategy, one they think of as a “first-best” privacy protection, is to promote trustworthy information relationships: “In this direction lies a better digital future for all of us – a digital society in which privacy rules promote trust and make life better for the humans who inhabit it”.

A couple of examples of “obfuscation” are:

- **Trackmenot** ([http://cs.nyu.edu/trackmenot/](http://cs.nyu.edu/trackmenot/)), a browser addin which protects users against search data profiling by issuing randomized queries to popular search engines. Using a strategy of noise and obfuscation, actual web searches performed using TrackMeNot are lost in a cloud of false leads, and are essentially hidden in plain view.
- **FaceCloak**, proposed by Luo, Xie et al. (2009) is architecture which protects user privacy on social networking sites by shielding a user’s personal information from the site and from other users that weren’t explicitly authorized by the user. FaceCloak provides fake information to the social networking site and stores sensitive information in encrypted form on a separate server.

4.3 Temporal frictions

Temporal frictions are frictions which are related to or limited by time. Examples include:

- Paper based records for computer bookings destroyed at the end of each week
• Transactional logs generated by access control software and network authentication anonymized/destroyed when no longer needed
• How long is a library user’s reading history retained by default?
• Does the user have a choice as to whether the data is retained or not?
• How long are details of previous searches kept for on each of the online services the library subscribes to?

4.3.1 Collection and retention of data

In order to protect the privacy of library users, librarians need to ask themselves whether the information they collect is strictly necessary for the purpose for which it is being collected; and where the information is required, to ask themselves how long it needs to be kept.

How long do you retain borrowing history data; or interlibrary loan requests data?

• Is it forever
• Is it for the default period used by your library management software provider, or
• Is it never (i.e. as soon as an item is returned, the record is erased)

There is a trade-off between privacy and convenience. Some people may like to be able to look back at their borrowing history as a useful reminder of the titles they have consulted, especially if they want to take another look at a particular item. The point is: Do your users get a choice as to whether, and for how long, their reading history is retained?

Some libraries make use of CCTV cameras. The British Library, for example, has 400 cameras, Professional Security (2014). One justification libraries might use for camera deployment is to prevent criminal activity. Where surveillance cameras are deployed, it is important to have a clear policy about who could potentially access the footage, under what circumstances it would be viewed, and for how long the footage is kept.

In January 2017 it was reported that Iowa Public library were using security cameras in their library bathrooms, Collier (2017); although by April 2017 the Iowa Senate had approved a Bill which bans cameras in restrooms and locker rooms at government buildings. If libraries feel that having CCTV cameras in their library bathrooms is absolutely necessary and represents a proportionate response to the threat/risk that it is intended to address, they should have visible and clear signage so that library users are fully aware of the existence of the cameras.

5. Lessons from copyright law

We believe that copyright law has a number of lessons relevant to privacy law:

5.1 Communication to the public / who your personal data is available to

The first point relates to how copyright law applies to hyperlinking. Communicating a work to the public by electronic means is an exclusive right of the rightsholder, so one of the most crucial questions is whether a hyperlink constitutes a communication to the public. In Nils Svensson & Others v Retriever Sverige AB (C-466/12) the ECJ came up with a novel answer to that question, namely that hyperlinks were a communication to the public but would only infringe if the communication was directed at a “new public”. In other words, to a public that wasn’t taken into account by the copyright holders at the time the initial communication was authorised.

The body of ECJ case law that has developed indicates that there are a number of factors that need to be taken into account:
- whether the content was published with the rightsholder’s consent
- whether there was a profit making intention
- whether there was actual knowledge that the content linked to was unlawful
- whether it constituted a communication to a new public

These factors (consent, profit-making, knowledge of unlawfulness, and communication to a new public) are all ones which could easily be applied to use of personal data. We are thinking here of the question of who has access to someone’s personal data, whether the individual ever intended that those people would have access to their personal data, whether they gave informed consent for it to be processed, and whether (even if they had given their consent) they could reasonably have anticipated the use(s) to which their data is now being put.

5.2 Contract override

The second point from copyright law which we think could usefully be applied in privacy law relates to contract override. We believe that individuals should have a series of unambiguous rights which are meaningful, have real teeth, and cannot be overridden by contract law. Data protection laws work on the basis of consent. Institutions use terms and conditions to legitimise their data processing; and to govern precisely what they can and can’t do. Mayer-Schönberger, Cukier (2013) recognise that privacy has become much more difficult to protect, especially with old strategies such as individual notice and consent. The use of terms and conditions disempowers the data subject, because:

- The terms and conditions are non-negotiable
- They can be changed at any time
- The changes can take place retrospectively

Hoback’s (2013) documentary “Terms and conditions may apply” draws attention to the use of T&C’s. They are typically very long, written in legalese, appear in a small font, and the text often uses capital letters. As soon as you use a service, you are deemed to have agreed to the terms and conditions. You wouldn’t be expected to agree to a set of T&C’s before being able to watch television, or before being able to read a book; whereas that is precisely what is expected of you if you read an ebook on a Kindle device, or watch a video on YouTube.

UK copyright law does not have a generic “no contractual override” provision applicable in all circumstances. However, a number of the exceptions contain wording which prevents contract override. For example, the text and data mining exception says that “To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this paragraph, would not infringe any right conferred by this Chapter, that term is unenforceable” (CDPA 1988 s29A(5), and similar wording can be found in a number of the copyright exceptions.

Recital 32 of the GDPR says that “Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject’s agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement”. In an era of “big data”, where organisations create digital doppelgangers for each individual, where data is gathered through observation, is derived or inferred, where algorithms are used without meaningful transparency and accountability to determine whether individuals are entitled to get a particular product or service, and at what cost, what would meaningful consent really look like?
5.3 Reasonable expectation of privacy

In Katz v. United States 389 U.S. 347 (1967) the supreme court applied a broad test for the fourth amendment protection it would apply whenever the government violated a person’s “reasonable expectation of privacy”. According to the US Supreme Court, if your information is in the hands of a third party, then you have no reasonable expectation of privacy and as a result no fourth amendment protection.

This concept is known as the third-party doctrine, where the presence of a third party to a transaction or activity eliminates any expectation of legal privacy. It is used to determine whether Fourth Amendment protections apply to the situation and, therefore, whether law enforcement must obtain a search warrant. According to the US third-party doctrine an individual does not have a reasonable expectation of privacy to any communication he or she voluntary discloses to a third person, Bedi (2013).

In the UK the courts have attempted to clarify the concept of a “reasonable expectation of privacy”. In Murray v Express Newspapers [2008] EWCA Civ 446 (paragraph 36) it says the “question of whether there is a reasonable expectation of privacy is a broad one, which takes account of all the circumstances of the case. They include:

- The attributes of the claimant;
- The nature of the activity in which the claimant was engaged;
- The place in which it was happening;
- The nature and purpose of the intrusion;
- The absence of consent; and
- Whether it was known or could be inferred;
- The effect on the claimant; and
- The circumstances in which and the purposes for which the information came into the hands of the [person using it].”

In US v. d’Andrea, 497 F. Supp.2d 117 (D. Mass. 2007) the U.S. District Court for the District of Massachusetts presumed that the password protection of a website was sufficient to afford a reasonable expectation of privacy.

6. Conclusion

Libraries have an important role to play in increasing optimising the informational friction in the infosphere. A key role relates to digital privacy literacy, training users how to value and protect their personal data. Libraries around the world organise training on privacy issues, as well as arranging cryptoparties. Some libraries create dedicated areas about privacy on their website (such as http://sjpl.org/privacy). Other methods for achieving informational friction include the use of https:// encryption on library websites, catalogues, and discovery services; and network penetration testing to reduce the likelihood of data breaches occurring.

Floridi’s ontological friction concept is a more suitable approach for analysing privacy issues than other theories noted above, because of its ability to encapsulate and structure many of the issues of this complex area, and because it is helpfully set within Floridi’s wider Philosophy of Information, allowing a formal consideration of many relevant aspects of information ethics. The aim of the project will be inter alia to identify the sorts of ontological frictions available to librarians and their
users to help them manage the extent to which personal data can flow back and forth. This article has concentrated on three types of ontological friction:

1. Technological frictions
2. Obscurity, practical obscurity and obfuscation
3. Temporal frictions

Other types of friction which have been identified, include spatial frictions, sensory frictions, training & awareness, information behaviour, contextual issues, and regulatory frictions.

These can be expressed in a simple and provisional two-stage conceptual model for privacy and associated frictions, derived from the initial stages of this work. This model will serve as the basis for more detailed study of the topics, as noted above. The model itself will be refined in the process.

BIBLIOGRAPHY


CPDP, 2017. CPDP 2017: types of privacy (Chair: Ivan Škorvánek, TILT, Tilburg University (NL)) (Video).


Floridi, L., 2016. Think piece episode 2 https://www.youtube.com/watch?v=WTtz3qi73iM.


Green, R., 2017. Looking at philosophical problems in the "infosphere" (Rob Green talks to Professor Luciano Floridi). CILIP Update, (April), 16-17.


Hedtke, L.P., 2008. Cereal boxes and milk crates zine Libraries and infoshops are... Now.

Hoback, C., 2013. Terms and conditions may apply.


McGeveran, W., 2013. The law of friction.


Society of American Archivists, 2016. SAA core values statement and code of ethics.


LEGAL CASES

William Frederick Ian Beggs v Her Majesty's Advocate [2010] HCJAC 27.

Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González. 2014.

Her Majesty’s Advocate v. William Frederick Ian Beggs (High Court of Justiciary 2001) (No2) (2002 S.L.T. 139)


Murray v Express Newspapers [2008] EWCA Civ 446

Nils Svensson & Others v Retriever Sverige AB (C-466/12)
