

ARTICLE



Facebook is creating records — but who is managing them?

Dominique Glassman

Faculty of Information, University of Toronto, Toronto, Canada

ABSTRACT

Facebook is creating records from the data that users are pouring into their profiles. This paper examines Facebook's data practices in the context of record creation through the theoretical lens of the Records Continuum Model. Three stakeholder groups are identified in record creation: the users, the site, and third-party platform participants. This paper analyses Facebook's data sharing practices in conjunction with the Cambridge Analytica scandal, and the new General Data Protection Regulation (GDPR).

KEYWORDS

Facebook; Records
Continuum Model; records
management

Introduction

Facebook is the largest social media platform in the world. The site has over 2 billion active users, posting 510,000 comments and updating 293,000 statuses every minute.¹ This immense quantity of data and content furnished by Facebook users and processed by the site, takes on different shapes, meanings, and uses, for people on either side of the flow of information. The data and content collected by Facebook become digital records, and is conceived of in a number of ways: a user's personal profile, the profile the site creates through accumulating user data, and the records that third parties create using data and content from Facebook.

Digital records are widely proliferated and continue to inform power and decision-making, now at the global level. Their scale and importance also make digital records vulnerable to falsity and inauthenticity, particularly in peer sharing, loosely regulated environments like Facebook. As the parameters for defining a record broaden, the practical requirements for managing digital records have begun to differ significantly from managing paper records. Digital records increasingly move away from having a fixed format, and because of this, 'despite the turn toward preserving digital media, archivists and individual creators have yet to address many of the archival challenges presented by social media with regard to the multiple contexts of collection creation'.² This paper posits that the multiple contexts of information use for a single Facebook record are tied to three stakeholder groups and rather than one fixed digital record, this results in three varied representations of the same user-generated data. These representations are the conceptual record as it is understood by the user, the record created by Facebook software and user interface, and finally, the marketable personality profile created by third-party partners and customers. Identifying and discussing these representations will identify how the social expectations for Facebook are not

balanced with the commercial realities of the site's relationship with its third-party platform participants. This paper will also conduct a brief survey of relevant methodology, for the purpose of defining a digital record in the theoretical sense, and laying the groundwork for making recommendations for records management at Facebook based on widely held standards in the records management community.

Using Facebook, for the casual user, is at least partially motivated by self-representation and self-promotion, which has trickled down to from the blog genre, and has resulted in the creation of an 'ad hoc personal archive'.³ The danger of situating Facebook too snugly within the blog genre, is that it ignores the platform's commercial motivations. Although Facebook's pragmatic personal data collection practices have come to light due to recent controversies, the site's motivations have long been obscured both figuratively, by the popularity of Facebook as a tool for self-expression, and literally by opaque data policies which typically remain unread by the average Facebook user.

The lack of current infrastructure for managing digital records outside of the traditional corporate or government settings has resulted in breaches of public trust concerning personal data that is processed en masse by social media and telecommunications companies. The latter is true for Facebook especially considering the tide of public vitriol against the social media platform in the wake of the Cambridge Analytica scandal. Cambridge Analytica is a high profile data analytics company, which has 'built models that translate the data they harvest into personality profiles for every American adult'.⁴ The company creates these personality profiles through collecting data from sites like Facebook and Twitter, and the profiles are used for behavioural 'microtargeting',⁵ where content appears on Facebook users' newsfeeds to influence them as voters or consumers. Cambridge scholar and digital developer Aleksandr Kogan created an app where willing users would provide their personal information in order to take a personality test. Kogan used Facebook's networking capabilities to mine personal data from the profiles of the app-users friends, without their consent. Kogan then transferred this data to a third party (Cambridge Analytica) for commercial purposes, which is specifically against Facebook's data policies, and a breach of public trust.

Breaches of personal privacy arising from Facebook's opaque commercial motivations are not a new controversy. However, this scandal is particularly relevant in that it arose from the illicit collection of data in order to construct records containing personal information, for the financial and political gain of third parties. Furthermore, the portrayal of the Cambridge Analytica scandal in the news clearly defines the same three stakeholders that are present in this paper's consideration of Facebook's digital records: Facebook, the users whose data were mined, and the third party data broker (in this case, Aleksandr Kogan and Cambridge Analytica). Consequently, this paper will focus on Facebook's digital records in 2018, at the time when the Cambridge Analytica scandal went public and the General Data Protection Regulation was first implemented.

Standards and models

The multiple contexts of use for Facebook records developed in this paper are informed by the Records Continuum Model (RCM). The RCM was developed by Frank Upward and Sue McKemmish, at Monash University in the 1990s. Although the records management field has produced several important theoretical models, the RCM is widely regarded for its multi-dimensionality, which allows records to be analysed and continually re-

contextualised within their creation, maintenance, use, and social plurality.⁶ The RCM has four dimensions: 'Create', 'Capture', 'Organise', and 'Pluralise', which correlate to different persons, actions, and purposes organised by 'Evidentiality', 'Transactionality', 'Recordkeeping Containers', and 'Identity'. This model provides the necessary vocabulary for describing how one record can serve so many purposes for different actors and institutions while existing in several different contextual dimensions.⁷

The RCM is very theoretical and has diverse applications in the recordkeeping world, including acting as a framework for the International Organisation for Standardisation's ISO 15489-1. ISO 15489-1 is an international standard that outlines the concepts and principles to support and inform the practice of records management. The standard defines a record as 'information created, received and maintained as evidence and as an asset by an organization or person, in pursuit of legal obligations or in the transaction of business'.⁸ Although relatively succinct, the ISO definition of a record is robust for the purpose of this paper. Facebook, after all, is a business, and Facebook retains and monetises user data as a commercial asset.

The InterPARES project defines a record as 'a document – that is, recorded information – made or received in the course of practical activity as an instrument or by-product of that activity and set aside for further action or reference'.⁹ This definition is more conducive to understanding the Facebook record as something created by the Facebook user for the purpose of commemorating social activity. The InterPARES Project is led by digital diplomatic scholar Luciana Duranti, based at the School of Library, Archival and Information Studies at The University of British Columbia, and aims to develop a framework for verifying the authenticity of digital records based in part on Diplomats.¹⁰ The latter is a method of analysis based on an examination of the extrinsic and intrinsic elements of a document. Although diplomats is a controversial methodology for records analysis, extrinsic elements of form are particularly significant when analysing user expectations for the Facebook platform based on specific elements of the user interface.

In her article 'Diplomats of born digital documents – considering documentary form in a digital environment', Corinne Rogers identifies three levels of abstraction for understanding the record as a digital object. The various abstractions are useful in defining how the Facebook record can actually exist in the various use-contexts explored in this paper. The first is the conceptual record as we see it onscreen. This is the level of abstraction for most casual Facebook users. The second is the physical object, 'information, consisting of data and the instructions for its manipulation and representation'.¹¹ This is the record that is reproduced and shared between Facebook and its platform partners. Lastly, there is the record as a logical object, the hardware and software components used to create the conceptual record.

Facebook records created by the user

Isolating the conceptual abstraction, or the Facebook record as the user perceives it, is the simplest way to understand Facebook as a platform for creating an archive of personal records. Facebook and many other popular social media sites have become a way for users to curate and commemorate their 'lived reality'.¹² Drawing a parallel to the RCM, this record is captured evidence of individual memory and activities.

The context for this kind of record creation is the user's personal experiences and desire for self-representation. Facebook's ancestral genre is the web blog, and according to Miller and Shepherd in 'Blogging as Social Action: A Genre Analysis of the Weblog', 'When bloggers talk about blogging, two themes relevant to these questions are ubiquitous: self-expression and community development.'¹³ Although Miller and Shepherd do not reference the RCM in their article, the idea of community development through record creation is a form of creating collective memory through pluralising your information. When a record is pluralised, it is 'brought into a framework to provide a collective social, historical and cultural memory of institutionalised purposes'.¹⁴ The theme of community development is evident, particularly in the earlier days of Facebook when it was marketed as a platform to connect users with one another based on school networks. Although the networking feature of Facebook has far exceeded its original intentions, people often use Facebook to connect with people or add people to their personal, curated network. Returning to the Cambridge Analytica scandal, Aleksandr Kogan abused the social action of community development among Facebook friends in order to access and profit from the Facebook records of over 50 million unsuspecting users.¹⁵ Kogan did all of this despite the site's feeble entreaty to app creators, 'Only use friend data (including friends list) in the person's experience in your app'.¹⁶

Self-expression is another social action common to both Facebook and blogging. Much like community development, self-expression has become a Facebook feature that is being capitalised upon by the platform and third-party partners. Although self-expression denotes autonomy and creativity, in the context of Facebook, it is dictated heavily by the format of the site itself. Because identical formats confine the content generated by all Facebook users, tenor and articulation become the only way to manifest freedom of personal expression. Facebook users are giving up any actual ownership of the content they are creating when they agree to Facebook's 'Data Policy', which gives the site full control over user information. Because very few casual users actually parse the site's policies to understand how Facebook processes their data, the biggest problem for users is the illusory ownership and authorship they believe they possess over their content. This belief is informed in large part by the history of the web blog, and the structural similarities Facebook shares to the blog that cause users to believe they are creating records only for themselves and their accepted list of friends. The format of the blog is far more proprietary for the user than that of Facebook, and the gap between public and private, which are the intrinsic and extrinsic motivations identified by Miller and Shepherd within the blogging discourse, has been replaced with the gap between a controlled public and an invisible third party with extrinsic motivations unknown to the user. Users know that their content is public to people in their network, but they are less aware that their personal disclosures are being collected and processed by Facebook and its platform partners, often for financial gain. The self-disclosure associated with Facebook's ancestral genres lulls users into a false sense of security when divulging personal information both to and on the platform.

Part of why the commoditisation of Facebook users' social actions, is stirring public ire, is because users expected Facebook to act more like a blog, and less like a platform for advertising and marketing. Users depend on social memory to understand Facebook because the forms of social interactions that occur on the site loosely mimic the structure of actual conversation. Furthermore, the interface shares some similarities

with the traditional web blog specifically, dated entries, space for commentary, and content displayed in reverse chronological order.¹⁷ Facebook even borrowed the term ‘blog post’, which has faded into obscurity in comparison to the ubiquitous ‘Facebook post’. Facebook operates within a commercial structure which was not present in blogs or earlier ancestral genres, so dependence on widely understood social structure is obscuring Facebook’s relationship with their platform partners. Because of the similarities between genres, the blog is a memory that Facebook users draw upon to infer a legitimisation, and inform their social practices on the site.

Facebook records created by Facebook

Within the socio-economic context of contemporary social media, it is easy to paint Facebook as the villain responsible for appropriating and abusing user data. However, the way that data and content is formatted by the website is what allows the information to become a record in the first place. According to digital diplomatics scholar Luciana Duranti, ‘the assessment of the value of specific data sets is conducive to the definition of the form of the records that should contain them and of the digital presentation that will allow for their long-term use, accessibility and preservation’.¹⁸ The Facebook system and format are what allows for the creation, standardisation, and preservation of personal content on the platform. The value placed on data contained by Facebook’s uniform ‘form’ (or interface) is a way to encourage online community building, and to democratise how information appears on the site. Although users are free to post whatever they want (within reasonable limits), all Facebook profiles are identically formatted.

Facebook formats the data it collects from users in several different ways. ‘Your categories’, which is part of ‘your ad preferences’ is a profile that Facebook has created based on user online activities. This record is meant to show what information advertisers are using to target individual users. Although some of the information is innocuous, there are categories like ‘Close friends of expats’, which reveals the degree to which Facebook collects and classifies data across user networks. Facebook users can also access an html. file which is a downloadable record created by the site that contains every action by or about the user since they joined the platform. This record is particularly significant because Facebook is formatted to show the most recent content first, and the older information seems to disappear.

In 2010 Zuckerberg told reporters, ‘I always read these articles that are like, “OK, you guys must be doing this because it’s going to make you more money”, And honestly, for people inside the company, that could not ring less true’.¹⁹ Zuckerberg has changed since 2010, which means that Facebook has changed too. According to the site’s ‘Terms of Service’:

You give us permission to use your name, profile picture, content, and information in connection with commercial, sponsored, or related content (such as a brand you like) served or enhanced by us. This means, for example, that you permit a business or other entity to pay us to display your name and/or profile picture with your content or information, without any compensation to you.²⁰

In addition to furnishing advertisers with user information, Facebook's privacy settings make user information available even to people off the site. The 'Terms of Service' reflect how Facebook profiles help bring in money to the site, while the only cost to users is their information. Until very recently, Facebook's 'Terms of Service' were actually called the 'Statement of Rights and Responsibilities'. The latter is discursively misleading because it is not about users rights; it is about Facebook's rights over users' data, and was it not for the pressure put on by the General Data Protection Regulation (GDPR), this wording would have no doubt remained. Because the creation and format of a user record are contingent upon the Facebook user interface, in addition to the high level of control Facebook can exercise over user data, this results in the creation of a split between the ownership of record at the moment of creation and the dimension of record capture.

Much in the same way that blogs have informed the way that people use Facebook, it is the discourse attached to the word 'platform', that shapes how Facebook treats both its users and dependent third-party developers. In his article "The Politics of "Platforms"" Tarleton Gillespie explains that platforms are part of political, architectural, computational, and figurative discourses. Because these discourses are connected by shared vernacular, Facebook, along with other tech and media giants like Google and YouTube also style themselves as platforms, which allows them to appear simultaneously as a platform for self-expression and online participation, as well as a financial platform for third-party developers.²¹ Tamsin McMahon of the *Globe and Mail* recognises the discursive fluidity of the word platform, and writes, 'to a growing array of critics, Facebook is a media company and should be regulated as such. For its part, Facebook wants to be viewed differently – as a platform – free of the regulations and responsibilities of traditional outlets such as newspapers and broadcasters'.²² Facebook actually has a specific 'Platform Policy' to assist third party developers in launching their apps on the Facebook platform. The policy prohibits developers from selling or purchasing data retrieved from Facebook, although it is clearly a common practice to do so. This, among others, is the rule that Kogan ignored when he collected data from users without consent and sold it to Cambridge Analytica.

According to Gillespie, the word platform, 'like other structural metaphors... depends on a semantic richness that, though it may go unnoticed by the casual listener or even the speaker, gives the term discursive resonance'.²³ The architectural discourse of the platform as a flat, even surface holds semantic significance for the relationship between corporate Facebook and third-party platform developers. Developers have access to unprecedented amounts of user data, but the 'Platform Policy' allows Facebook to profit equally from third party involvement. According to Facebook's 'Platform Policy, 'we can analyze your app, website, content, and data for any purpose, including commercial'.²⁴ Despite associations with the structural metaphor, the Facebook platform is not an even plane because Facebook can monitor and collect data from any app on their platform, but their policy also warns that access to the platform might not always be free for advertisers.²⁵ The exchange of information between Facebook and third-party customers depends on the freedoms afforded by both the literal and figurative platform.

The idea of the platform denotes openness and visibility; archives, however, evoke deep, closed, regulated systems for records. The discourse associated with archives or

digital repositories is about long-term controlled retention and preservation. However, by styling itself as a platform, Facebook circumvents the rules for proper records retention and disposition that are required for public records. It is these measures that help protect the record subject's privacy in conjunction with organisational accountability. Facebook has to tell users how long the site will keep their data, but their 'Data Policy' relies once again on semantic vagueness in order to maximise their power for data retention.

We store data for as long as it is necessary to provide products and services to you and others, including those described above. Information associated with your account will be kept until your account is deleted, unless we no longer need the data to provide products and services'.²⁶

'As long as necessary' gives Facebook unregulated power over information from user profiles or gleaned through cookies tracking user activities on other apps. Even deleting your Facebook account is no guarantee that your records will be destroyed because 'Information that others have shared about you isn't part of your account and won't be deleted'.²⁷ Retention should be decided at the moment of record creation, and the irony, in this case, is that it actually is. The 'as long as necessary' clause is written into the policy that users agree to when they sign up for Facebook in the first place. This lack of platform regulation is indicative of the struggle for recordkeeping practices to adapt to the current digital climate, especially when companies like Facebook are working to avoid proper recordkeeping practices that might prevent them from storing and disseminating user data long term.

Facebook users are benefiting from Facebook in that it offers access to online community building as well as a platform for self-expression. Unfortunately, the casual Facebook user would not understand how styling itself as a platform has allowed Facebook to share user data with third-party developers, to keep their data indefinitely, and generally avoid the constraints and information governance models faced by traditional forms of media.

Facebook records created by third party platform participants

The record created by the user and the record created by Facebook share the same binary content, despite the fact that the digital representation of the record is different for the user and the company. However, Facebook also furnishes third-party participants with this binary content. 'Between 2007 and 2014, [Facebook] gave developers access to its social graph – the map of users' networks of friends, interests, and likes'.²⁸ It is within their purview to do this because users agree to third-party platform participation when they agree to Facebook's 'Data Policy'. Unfortunately, Facebook has no way of controlling what third parties do with records like the social graph.

Although Facebook and its platform partners share a profit-based exigence, they each use content and data from Facebook differently in order to realise this exigence. Facebook uses its records to make money from third-party platform participants, while the latter uses records from Facebook in order to target users as potential customers and consumers. Because there are two different actions based on the same data, two distinct records are created. Furthermore, as evidenced by the Cambridge

Analytica scandal, data analytics firms, as well as marketing and advertising agencies collect data from multiple sources, not just Facebook, in order to create personality profiles they can sell to advertisers or even political parties. The record created by third-party platform users is connected to Facebook records but also exists separately.

Corinne Rogers writes, 'Digital objects can be infinitely reproduced, and their meaning and determination of their reliability and authenticity depend on knowledge of their context and provenance'.²⁹ In her article 'Contemporary Archival Diplomats as a Method of Inquiry: Lessons Learned from Two Research Projects', Heather MacNeil argues that determining the authenticity of a record is difficult because 'electronic records are deeply embedded within multiple and overlapping contexts that are complex and subject to rapid change'.³⁰ The provenance and context of the profiles created by Cambridge Analytica are especially murky. Cambridge Analytica was purporting to represent certain relevant information about Facebook users to paying parties, but the subjects of the records were not directly involved in the creation of the records detailing among other things, their potential voting preferences. David Carroll characterised his own Cambridge Analytica profile as 'incomplete but accurate',³¹ contrarily, Kogan admitted, 'In practice my best guess is that we were six times more likely to get everything wrong about a person as we were to get everything right about a person. I personally don't think micro-targeting is an effective way to use such datasets'.³² Kogan's admissions clarify that the context of the records he was creating was in order to micro-target people, but because he couldn't just poll millions of willing participants, he used a workaround to collect data and create records that were not even accurate. Returning to the RCM, it is evident why the third party platform participation becomes a problem. The information that is provided to third parties is still an active record if the subject is still a Facebook user. Third party participants are not adhering to access controls, and Cambridge Analytica is pluralising user records without their consent. Although the Cambridge Analytica scandal is represented as breaching privacy rather than an archival mismanagement, when examining the provenance of the content and data, the fact that users' Facebook content was taken and changed without their permission indicates that the records lack reliability and authenticity. This is compounded by the fact that Cambridge Analytica was building profiles using data from a number of different sources, and the full context of record creation is still unclear because investigations against their unlawful data collection are still ongoing.

What are Facebook's motivations for managing their records?

Facebook's motivations to support data privacy and digital records management have changed and intensified in the wake of Cambridge Analytica. In their article 'Motivations for records management', authors Sari Mäkinen and Pekka Henttonen, examine the motivations for records management in organisations from both the private and public sector. These motivations are gauged against a qualitative analysis of Records and Information Management (RIM) motivations from ISO 15489-1 and distilled into three categories: internal motivations, external motivations, and cultural societal goals. The authors hypothesise that an organisation with natural motivations for records management (the example cited in the article is a defense industry enterprise) will be more likely to enforce stricter records management. Henttonen and Mäkinen write:

An organisation with a high natural need for records management can be in a danger zone when records management is neglected. Conversely, understanding the relationship between organisational context and records management may help to identify “societal risk” organisations which lack internal motivation for investing in records management, but in which good records management is desirable for broader societal reasons.³³

Applying Henttonen and Mäkinen’s hypothesis and research framework to Facebook is relevant because Facebook is an organisation with a high natural need for records management. The company deals with a massive volume of user data and they hold all the personal information that is required to open a Facebook account for billions of users around the world.

Historically, Facebook has not exhibited much proof of internal motivations to properly manage digital records. Although they have ‘Terms of Service’ and a ‘Data Policy’, which outline acceptable data collection practices for third-party platform participants and the site itself, there is no motivation to enforce the rules. This lack of motivation stems from past precedent, where Facebook would settle mass consumer lawsuits regarding privacy violations by donating a few million dollars to nonprofit organisations.³⁴ Facebook was aware of Cambridge Analytica’s illegal data collection for profit as early as 2015 but was not motivated to confront the issue until the story became public.³⁵ It is this ability to pay away lawsuits and bad press, that helped to relax any internal motivations for better management of user records, and explains Kogan’s admission during his interview with *BBC4*, that Cambridge Analytica assured him that in regard to his data mining from the profiles of non-consenting Facebook users, ‘thousands and maybe tens of thousands of apps were doing the exact same thing’.³⁶ Facebook was aware of the potential ‘societal risk’ resulting from third parties collecting and selling user data but was not internally motivated until the risk became a public scandal.

Facebook has been relatively unscathed in terms of external motivations for records management, in part because the greater public has only recently become more aware of the extent to which the site collects and distributes users’ personal data. In their article, Henttonen and Mäkinen identify an important external motivation as ‘meeting legislative requirements’.³⁷ This kind of motivation has traditionally been more important to government and public sector organisations that are accountable to the public and governed by law and legislation. As of 25 May 2018 however, a new external motivation has been enacted that is already having an effect on Facebook’s data management practices. All companies no matter where they are located, that collect or process data from EU citizens, are subjects to the General Data Protection Regulation (GDPR). Broadly, this regulation ensures that companies build data protection into their electronic systems at the outset, rather than adding privacy measures as a response to a problem. The GDPR will also permit EU citizens to file claims for the erasure of data no longer deemed necessary for a specific processing operation. The GDPR’s most powerful compliance motivator is the possible penalties for data privacy violations, which can go up to 4% of global annual revenues.³⁸

The GDPR forced Facebook to revise and rename their ‘Terms of Service’, and millions of users received emails about the site’s updated ‘Data Policy’. However, as soon as the GDPR came into effect, Facebook was hit with a lawsuit from Austrian privacy activist Max Schrems. Schrems claims that Facebook ‘blackmails’ its users by forcing them to either accept the updated ‘Terms of Service’, or deactivate their accounts. The lawsuit was raised

on the basis that Facebook's new GDPR-fuelled policies violate the terms of the regulation by causing users to click a single box to agree to a myriad of different terms and conditions.³⁹ This is coupled with the lawsuit from the UK Information Commissioner's office that arose from Cambridge Analytica. Because Cambridge Analytica occurred before the GDPR came into effect, Facebook was fined less than it would have been if the scandal erupted today (the estimated post-GDPR fine is \$1.9 bn).⁴⁰

The GDPR is a very powerful and enforceable regulation. However, the GDPR was developed to protect the record as a physical abstraction rather than the conceptual record. The GDPR conceptualises personal information as data because this is a concrete way to control the flow of information. Because records exist in a continuum, the document or digital object is just one component of the record's full worth and meaning. The GDPR isolates the specific accountabilities of the data controller, processor, and data subject. While the GDPR contains a wide range of rules that are very beneficial to the rights of the data-subject, it lacks specific clauses for how data should be managed when it becomes part of a record.

This paper has examined three different abstractions of the Facebook record, in order to identify the expectations each group has when using the site. It is fair to say that the average Facebook user did not expect Facebook to give and receive information about them from third-party data providers while they are both on and offline. Facebook and their third-party platform participants have shared a mutually beneficial relationship for a long time, which has allowed the site to accrue enough money to pay off lawsuits until the implementation of the GDPR. Facebook's new data policy spells it out in plain English: 'We collect the content, communications and other information you provide when you use our Products, including when you sign up for an account, create or share content, and message or communicate with others'.⁴¹ This new clarity is mandated by the GDPR, which states that 'any information addressed to the public or to the data subject should be easily accessible and easy to understand, and that clear plain language is used' (GDPR).⁴² While Facebook's data collection practices have not changed, the site's data policy is now significantly more transparent, which means that users should be better informed, and may have to shift their expectations of the site.

If Facebook developed enforceable internal policies (rather than public data policies simple for the sake of the users) specifically to govern the flow of information between the site and the platform partners, instead of turning a blind eye to the obvious data mining of user information, the site would have had documentation to prove that there were internal protections in place for user data.

It is difficult to tell what internally motivates Facebook to make changes. From an external point of view, significant system and policy changes at Facebook only occurred after pressure from the GDPR. If Facebook continually updated their policies to reflect changes in their technology, relationships with other business entities, and how this might impact 'changing perceptions of risk or priorities',⁴³ they would be much less susceptible to lawsuits and poor public opinion.

As Henttonen and Mäkinen suggest,

decisions about implementing identified records requirements should be based on an assessment of risks balanced against resource implications. Requirements with more significant risks should be addressed with a greater investment of resources and additional monitoring and evaluation measures.⁴⁴

The Cambridge Analytica scandal is an example of a lack of investment and poor risk calculation, and the implementation of the GDPR is an enforced risk management strategy. Facebook should be allotting more resources to manage the risks associated with processing so much data and personal information.

The final element of Henttonen and Mäkinen's analytical framework for motivating records management, is 'cultural and societal goals'. This motivation is usually about records preservation for social and historical posterity in organisations with an archival mandate. Since it is a social networking site, Facebook's supposed cultural and societal goals are to connect people with communities, family, and friends. Their mission statement, updated in 2017, is to 'Give people the power to build community and bring the world closer together'.⁴⁵ Allowing companies like Cambridge Analytica unfettered access to millions of user records is certainly not in line with trust and community building, but being aware of this data breach and not doing anything about it makes Facebook's social goals and mission statement ring especially false. 'This was clearly a mistake', Zuckerberg told CNN, 'We have a basic responsibility to protect peoples' data, and if we can't do that then we don't deserve to have the opportunity to serve people'.⁴⁶ This statement was born out of a lack of internal and external motivation to come forward and address the societal risk posed by the Cambridge Analytica data breaches. Although Facebook profits from advertisers and marketing firms, maintaining cultural and societal goals is important because it is still a social media company, which thrives on user engagement. Examining Facebook in this cultural moment could represent a shift in the company's motivations for proper records management, influenced by the GDPR, and the public backlash from the Cambridge Analytica scandal.

Conclusion

Facebook has the name, photo, gender, as well as stated education, work, relationship status and much more, of close to 2 billion people. Facebook is one the biggest collectors of data for digital records in the world, and there is shockingly little in the way of rules and guidelines to help the site properly manage the content and information provided by its legions of users. Facebook and its acquisitions have often faced legal action and public scrutiny for issues around data privacy. However, because users only have two options: to either agree to all the terms and conditions or disagree and forgo the ability to use the site at all, Facebook has had little motivation for controlling the access and dissemination of the data and content their site collects. The one-two punch of the Cambridge Analytica scandal and the strict implementation of the GDPR have helped to bring the importance of data privacy and information management for social media platforms to the fore. The Cambridge Analytica scandal, in particular, helped to identify the three primary groups that are involved in data processing and record creation within Facebook. Facebook would have no records were it not for the data and content that users are generating both on the site directly, and as a result of Facebook's cookies tracking their movement online. Facebook, in turn, collects this data to learn about their users, often in such a way as to extract financially beneficial meaning from their records for the third party platform participants. The site's 'Data Policy' and 'Platform Policy' reveals how Facebook records are connected to third-party advertisers and data analytics firms. The casual user is often

unaware of who is able to view their information after they agree to Facebook's 'Terms of Service'. A single click in the 'I agree' box underrepresents the full scale of permissions users are giving to the site.

Analysing Facebook records through the theoretical lens of the Records Continuum Model is a way to illustrate how the information collected by Facebook can exist contemporaneously as different iterations for different parties involved in working with the information. Outside of the theoretical realm, identifying the web blog as a Facebook precursor, as well as the data collection practices that Facebook shares with other digital platforms, establishes that although Facebook is monolithic, it is just one element of the intertextual system that connects the world on the platform of the Internet. The records created by Facebook cannot be analysed without taking into consideration that they are part of a greater system, and perhaps now we are in the process of creating new values and attitudes as part of the institutional memory that will inform future digital practices.

Notes

1. Bernard Marr, 'How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read', *Forbes*, 21 May 2018, available at <<https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#531775a60ba9>>, accessed 9 August 2018.
2. Amelia Acker and Jed R. Brubaker, 'Death, Memorialization, and Social Media: a Platform Perspective for Personal Archives', *Archivaria*, vol. 77, Spring 2015, p. 3.
3. *ibid.*, p. 4.
4. Sean Illing, 'Cambridge Analytica, the Shady Data Firm that Might be a Key Trump-Russia Link, Explained', *Vox Magazine*, 4 April 2018, available at <<https://www.vox.com/policy-and-politics/2017/10/16/15657512/cambridge-analytica-facebook-alexander-nix-christopher-wylie>>, accessed 20 June 2018.
5. *ibid.*
6. Belinda Battley, 'Finding Aids in Context: Using Records Continuum and Diffusion of Innovations Models to Interpret Descriptive Choices', *Archives and Manuscripts*, vol. 41, no. 2, 2013, pp. 129–45.
7. *ibid.*
8. International Standard for Organization, 'Information and documentation – Records management – Part 1: Concepts and principles', 2001.
9. InterPARES 3 Project, 'The InterPARES Project Terminology Database', available at <http://www.interpares.org/ip3/ip3_terminology_db.cfm>.
10. *ibid.*
11. Corinne Rogers, 'Diplomatics of Born Digital Documents – Considering Documentary form in a Digital Environment', *Records Management Journal*, vol. 25, no. 1, 2015, p. 11.
12. Acker and Brubaker, p. 4.
13. Carolyn R. Miller and Dawn Shepherd, 'Blogging as Social Action: A Genre Analysis of the Weblog', *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*, 2004, available at <<https://conservancy.umn.edu/handle/11299/172818>>, accessed 15 April 2018, p. 9.
14. Battley, p. 134.
15. James Vincent, 'Academic Who Collected 50 million Facebook Profiles: We thought we were doing Something Normal', *The Verge*, 21 March 2018, available at <<https://www.theverge.com/2018/3/21/17146342/facebook-data-scandal-cambridge-analytica-aleksandr-kogan-scapegoat>>, accessed 20 July 2018.
16. Facebook, 'Facebook Platform Policy', 2018, available at <<https://developers.facebook.com/policy>>, accessed 1 August 2018.

17. Miller and Shepherd, p. 6.
18. Luciana Duranti, 'Structural and Formal Analysis: The Contribution of Diplomatics to Archival Appraisal in the Digital Environment', in Jennie Hill (ed.), *The Future of Archives and Recordkeeping: A Reader*, Facet Publishing, London, 2011, p. 10.
19. Tamsin McMahon, 'The long road to fixing Facebook', *The Globe and Mail*, 12 March 2018, available at <<https://www.theglobeandmail.com/technology/the-long-road-to-fixingfacebook/article38268627/>>, accessed 20 June 2018.
20. Facebook, 'Terms of Service', 2018, available at <<https://www.facebook.com/terms.php>>, accessed 8 August 2018.
21. Tarleton Gillespie, 'The politics of "platforms"', 2010, available at <<http://journals.sagepub.com/doi/abs/10.1177/1461444809342738?journalCode=nmsa>>, accessed 15 April 2018.
22. McMahon.
23. Gillespie, p. 349.
24. Facebook, 'Platform Policy'.
25. *ibid.*
26. Facebook, 'Data Policy', available at <<https://www.facebook.com/policy.php>>, accessed 12 August 2018.
27. *ibid.*
28. Vincent.
29. Rogers, p. 8.
30. Heather MacNeil, 'Contemporary Archival Diplomatics as a Method of Inquiry: Lessons Learned from Two Research Projects', *Archival Science*, vol. 4, December 2004, p. 201.
31. Brent Bambury, 'Data Mining Firm Behind Trump Election Built Psychological Profiles of Nearly Every American Voter', *CBC News*, 20 March 2018, available at <<http://www.cbc.ca/radio/day6/episode-359-harvey-weinstein-a-stock-market-for-sneakers-trump-s-data-mining-the-curious-incident-more-1.4348278/data-mining-firm-behind-trump-election-built-psychological-profiles-of-nearly-every-american-voter-1.4348283>>, accessed 19 April 2018.
32. Chris Aldridge (Host), 'BBC4 Daily Broadcast', 21 March 2018, available at <<http://www.bbc.co.uk/programmes/b09vyvxx#play>>, accessed 15 April 2018.
33. Sari Mäkinen and Pekka Henttonen, 'Motivations for Records Management in Mobile Work', *Records Management Journal*, vol. 21, no. 3, Winter 2011, p. 189.
34. Jeff John Roberts, 'Facebook Has Been Hit By Dozens of Data Lawsuits. And This Could be Just the Beginning', *Fortune*, 30 April 2018, available at <<http://fortune.com/2018/04/30/facebook-data-lawsuits/>>, accessed 10 August 2018.
35. Deepa Seetharaman and Katherine Bindley, 'Facebook Controversy: What to Know About Cambridge Analytica and Your Data', *The Wall Street Journal*, 23 March 2018, available at <<https://www.wsj.com/articles/facebook-scandal-what-to-know-about-cambridge-analytica-and-your-data-1521806400>>, accessed 17 August 2018.
36. Chris Aldridge (Host), 'BBC4 Daily Broadcast', 21 March 2018, available at <<http://www.bbc.co.uk/programmes/b09vyvxx#play>>, accessed 15 April 2018.
37. Mäkinen and Henttonen, p. 192.
38. European Commission, '2018 EU reform of data protection rules', 2018, available at <https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en>, accessed 25 June 2018.
39. Derek Scally, 'Max Schrems Files First Cases under GDPR Against Facebook and Google', *The Irish Times*, 25 May 2018, available at <<https://www.irishtimes.com/business/technology/max-schrems-files-first-cases-under-gdpr-against-facebook-and-google-1.3508177>>, accessed 10 August 2018.
40. Alex Hern and David Pegg, 'Facebook fined for Data Breaches in Cambridge Analytica Scandal', *The Guardian*, 11 July 2018, available at <<https://www.theguardian.com/technology/2018/jul/11/facebook-fined-for-data-breaches-in-cambridge-analytica-scandal>>, accessed 15 August 2018.
41. Facebook, 'Data Policy'.
42. European Commission.

43. ISO, 2001, p. 11.
44. Mäkinen and Henttonen.
45. Mark Zuckerberg, 'Bringing the World Closer Together', *Facebook*, 22 June 2017, available at <<https://www.facebook.com/notes/mark-zuckerberg/bringing-the-world-closer-together/10154944663901634/>>, accessed 18 August 2018.
46. Elie Meixler, 'We Have a Basic Responsibility to Protect Peoples' Data. Mark Zuckerberg Addresses Data Crisis', *CNN*, 22 March 2018, available at <<http://time.com/5210191/mark-zuckerberg-cnn-facebook-cambridge-analytica-data/>>, accessed 22 June 2018.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributor

Dominique Glassman has successfully completed her Master of Information at the University of Toronto's iSchool. Her chosen concentration was Archives and Records Management. She is interested in issues around access and privacy for digital records and information systems.