SECURE COMMUNICATION PLATFORMS

Developing A Framework for Assessing the Metadata of Communication Platforms

Centre for Innovation & Free Press Unlimited

CENTRE FOR INNOVATION
Leiden University

FREE PRESS UNLIMITED

Leiden University's Centre for Innovation is a group of interdisciplinary researchers, designers and developers driven by creating impact on how people learn, work, and live. Our mission is to put people first in innovation.

Free Press Unlimited is a not-for profit, non-governmental organisation based in Amsterdam, the Netherlands. Free Press Unlimited believes that objective information can be of life-and-death importance. FPU help local journalists in war zones and conflict areas to provide their audience with trustworthy news and information.

Author:

- Joanna van der Merwe - *Privacy and Protection Lead and Data Protection Officer, Centre for Innovation Leiden University.*

Project Team:

- Nouska Du Saar - *OSINT and Journalism Fellow, LightHouse Reports. Former Publeaks Mobile Project Officer, Free Press Unlimited.*
- Thomas Baar - *Policy Officer, Department Multi-Actor Systems, TU Delft*. *Former Project Leader, Centre for Innovation Leiden University.*

Reviewers:

- Massimo Marelli - *Head of Data Protection Office, International Committee of the Red Cross; Fellow and Advisory Board Member, European Centre on Privacy and Cybersecurity, University of Maastricht.*
- Dr Lena Hohfeld - *Policy Programme Officer, United Nations World Food Programme.*
- Linet Kwamboka - *Founder and CEO, DataScience LTD; Former Mozilla Tech Policy Fellow.*
- Marcel Oomens - *Program Manager, Zamaneh Media.*
- Josje Spierings - *Project Manager, Centre for Innovation Leiden University.*

PUBLICATION: 9 April 2020

This report is a part of the Free Press Unlimited project Publeaks. For information on this project please visit:
https://www.freepressunlimited.org/en/projects/publeaks-nederland-secure-and-anonymous-whistleblowing

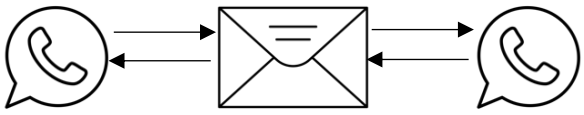
Table of Contents

1. INTRODUCTION

When assessing the security of a communication platform or channel, the focus tends to be on the privacy of the content of the messages. However, there is growing concern regarding the metadata collected by various messaging platforms - and how to ensure the security and privacy of a data subject, with regards to their metadata. A method and framework for understanding what metadata is collected and the implications for security of the messaging platform, needs to be established

In 2018 the number of people using messaging apps on their mobile phones was approximately 3.6 billion[1], a figure that will continue to increase. Apps such as Facebook Messenger, WhatsApp and WeChat are quickly becoming the primary means of communication on mobile phones. The growth of these messaging platforms presents a great opportunity for civil society, international development and humanitarians as it provides an easy communication link between these groups and their beneficiaries.

However, the use of these platforms also produces risks around data privacy and security that are yet to be fully explored[2]. Research focus into the privacy of these communication channels tends to be on the content of messages, with little attention paid to metadata surrounding the messages and use of these platforms. This report aims to shift the focus to metadata surrounding the communications, with regards to the security and privacy of a data subject. This report will provide an assessment framework and a comparative assessment of selected communication platforms, based on key aspects of the metadata collection and production.



Metadata includes the information about what channel, when, who and how of a communication but does not include the contents of the communication. Therefore, metadata can include Personal Data.

As the world becomes increasingly digitised the amount of data produced, collected, analysed and stored increases. One of the categories that is growing exponentially in the digital era is metadata. A common way to define and understand metadata, is the

[1] Vota, Wayan. 2018. *5 Guidelines for Using Mobile Messaging Applications in International Development.* September 10. Accessed January 25, 2019. https://www.ictworks.org/messaging-applications-international-development/#.XGKOO-JKjOR

[2] ICRC; The Engine Room; Block Party. January 2017. "Humanitarian Futures for Messaging Apps."

data generated about or describing other data[3]. For example, it is the data that indicates when you sent a message or an email but does not include the actual content of that message or email. Amongst other data points it includes the channel, when, who and the how of a communication[4].

In general, the ways in which metadata are generated and/or collected can be broken down into 3 main categories, namely:

1. Volunteered data
2. Behavioural data ("digital breadcrumbs")
3. Other data (e.g. device and system data)

Volunteered data includes the data subjects knowingly provide when carrying out actions such as starting an account. This includes data such as birth date, geographical location (e.g. place of residence), financial details etc. Behavioural data includes the data that is left behind as a data subject interacts with the digital world and can be generated both knowingly and unknowingly. For example, making use of travel cards for public transport systems, use of messaging apps, food delivery services etc. Other data includes data that is not about the behaviour of the data subject but rather the systems and devices which they are using. This includes data such as the operating system of a computer or phone, the version of an app being used, the other apps or programs present on the device etc.

Metadata can be used for a number of different purposes such as targeted advertising, optimised searches, data management and platform diagnostics. Additionally, collecting and storing metadata makes working with any data easier as it allows for efficient categorisation, sharing, searching and use of data.

However, in 2014, the former Director of the Central Intelligence Agency stated that they "kill people based on metadata". Whether referring to groups of people or individuals, this statement accurately highlights the way in which metadata can be used to target vulnerable populations, human rights activists, whistle-blowers etc. Furthermore, as organisations digitise the way they work, they generate increasing amounts of metadata about their beneficiaries, their employees, and their work – data that is not necessarily under the exclusive control of these organisations. The use of metadata for sinister purposes raises the question: Do we understand what qualifies as data that may put identifiable individuals or groups at risk? This has led to a gap in the tools and policies implemented to protect individuals.

"[M]etadata absolutely tells you everything about somebody's life. If you have enough metadata, you don't really need content"
Stewart Baker, NSA General Counsel, quoted by David Cole

"We kill people based on metadata"
General Michael Hayden, former Director of the CIA

[3] International Committee of the Red Cross (ICRC); Privacy International. October 2018. "The Humanitarian Metadata Problem: "Doing no harm" in the digital era."
[4] For a long list of the metadata assessed, see Appendix A

Identifying Individuals Using Metadata

'People You May Know'

Over the years Facebook has developed its People You May Know function, a feature users cannot opt out of. This feature works with data that the user has no control over and will definitely include metadata. In 2016 suggestions came out that the company was using location data to provide these suggestions. This led to concerning reports of the protected sources of journalists, clients of sex-workers, and patients who share psychiatrists all being suggested as connections through this feature. Similar features exist on other social media sites such as LinkedIn and Instagram.

Source: Hill, Kashmir. 2018. "'People You May Know': A Controversial Facebook Feature's 10-Year History." *Gizmodo.* August 9.

When metadata is used to create social graphs, representations of the interconnections between people who make up online social networks, or identify the behavioural patterns of individuals or groups it can place these people in dangerous positions. Therefore, despite not collecting or storing the contents of messages, the metadata collected and stored by messaging platforms may present challenges in protecting the privacy of an individual or group.

When combined with other data sets ("mosaic effect"[5]), metadata can provide a detailed image of a person, their habits, travel routes, personality and much more. As MIT Professor Alex "Sandy" Pentland puts it; "metadata allows for 'reality mining'"[6]. Metadata can be seen as pieces of data left behind as people move about the digital (and physical) world. It places a microscope on the behavioural habits of people. When Edward Snowden leaked documents regarding the National Security Agency (NSA) and its data collection methods, many downplayed the privacy implications by arguing that the content of phone calls was not listened to, only the length and location of phone calls was being stored (Pearson 2013)[7]. In response, a number of researchers have been studying just how close they can get to identifying individuals by studying various sets of metadata.

A study carried out by Yves-Alexandre de Montjoye, *et al.* was able to uniquely identify 95% of the mobility patterns of individuals, based on the spatial-temporal information of specific cell phones[8]. Another study conducted at Stanford University demonstrated that metadata associated with telephone calls allowed for relatively easy identification of people, and allowed for sensitive inferences such as the social relationships of data subjects[9]. The concerns raised by the generation of metadata in a humanitarian setting

[5] The "Mosaic Effect" refers to a method of gathering data in which disparate pieces of data are brought together. Although the individual pieces of data may hold limited value, they become significant when combined with other data and information.
[6] Greene, Kate. 2008. "TR10: Reality Mining." *MIT Technology Review.* February 19. Accessed August 28, 2019. http://www2.technologyreview.com/news/409598/tr10-reality-mining/.
[7] Pearson, Michael. 2013. "Obama: No one listening to your calls." *CNN Politics.* June 10. Accessed August 28, 2019. https://edition.cnn.com/2013/06/07/politics/nsa-data-mining/index.html
[8] Montjoye, Yves-Alexandre de, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. *Unique in the Crowd: The privacy bounds of human mobility.* Scientific Report, Nature.
[9] Mayer, Jonothan, Patrick Mutchler, and John C. Mitchell. 2016. "Evaluating the privacy properties of telephone metadata." *PNAS* (PNAS) 5536 - 5541.

have also been analysed from a computer science angle, to identify specific needs applicable in the humanitarian sector and possible technical solutions[10].

Despite increasing awareness of the sensitivity of metadata and its inclusion in the definition of Personal Data in documents such as the General Data Protection Regulation (GDPR), policies and frameworks governing communications and data are focussed on protecting the contents of messages - whilst paying little attention to the protection of data subjects with regards to their metadata. This is due to lack of understanding how seemingly benign metadata can lead to the identification of an individual or group in various contexts. Furthermore, there is inadequate awareness of the varying risks associated with different types of metadata, as well as lack of awareness on the side of the data subject as to when and how they are producing the metadata. This is the result of the difficulties in gaining access to the metadata collected and in scoping the amount of metadata that exists, as it is highly fragmented across different software providers, platforms etc. In general, there is a lack of transparency about how, when and why metadata is generated and collected[11].

Death by metadata – the US military and targeted drone strikes

Metadata is the basis for targeted drone strikes by the United States of America (USA) in countries such as Yemen, Somalia and Afghanistan. Using a combination of metadata sources including SIM-card tracking, handset tracking technologies and satellite imagery groups such as the Joint Special Operations Command (JSOC) high value targeting taskforce, in collaboration with the National Security Agency (NSA) and Central Intelligence Agency (CIA) select a target for remote drone strikes. This strike is often carried out without human intelligence confirming the identity of the person holding the phone. Human operatives are only used to determine the damage after the strike has occurred (Scahill and Greenwald 2014).

Source:
Scahill, Jeremy, and Glenn Greenwald, interview by Amy Goodman. 2014. *Death By Metadata: Jeremy Scahill& Glenn Greenwald Reveal NSA Role in Assassinations Overseas* (February 10).

Metadata and the Law

Since the European Union's General Data Protection Regulation (GDPR) came into effect in May 2018, it set the bar for the level of protection that should be afforded to personal data across all sectors. It influences the data protection laws emerging in other areas of the world. Despite not explicitly using the term metadata, it does cover this type of data by its definition of Personal Data. The GDPR is not new in using a definition of Personal Data that covers metadata. Council of Europe Treaty No. 108, introduced in 1981, is a binding international document governing the automatic processing of personal data for both members and non-members. The Council of Europe decided to update this document in 2011, with the updated adoption occurring

[10] Blond, Stevens Le, Alejandro Cuevas, Philipp Jovanovic Juan Ramón Toncoso-Pastoriza, Bryan Ford, and Jean-Pierre Hubauz. 2018. "On Enforcing the Digital Immunity of a Large Humanitarian Organisation." *2018 IEEE Symposium on Security and Privacy* 424 - 440.
[11] Montjoye, Yves-Alexandre de, Erez Shmueli, Samuel S. Wang, and Alex Sandy Pentland. 2014. "openPDS: Protecting the Privacy of Metadata through SafeAnswers." *PLoS ONE.*

in 2018. In both of these documents, personal data is defined as "any information relating to an identified or identifiable individual ("data subject")"[12].

In 1995 the European Union introduced the Directive on the protection of individuals, with regard to the processing of personal data and the free movement of such data. This Directive provided a more comprehensive and precise definition of personal data. It defined this data as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number of to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity"[13]. The General Data Protection Regulation (GDPR), which replaced the 1995 Directive, expanded this definition and now explicitly includes data that fall under the definition of metadata.

The GDPR definition is "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 reference to an identifier such as name, and identification number, location data, and online identifier or to one or more factors specific to physical, physiological, genetic, mental, economic, cultural or social identity of that natural person"[14]. The evolution of the definition of personal data provided in these documents, has come to more and more explicitly include metadata. Under the GDPR, metadata must be processed in line with the requirements of the regulation.

Personally Identifiable Information (PII) vs. Personal Data

PII refers to data that can be linked back to an individual. This includes data such as identity numbers, data of birth, email address, biometric records etc. PII also includes data that can be linked to an individual and easily trace back such as educational, financial or medical information.
Personal Data refers to a much broader category of data that, according to legal documents such as the GDPR, includes data that can directly or indirectly lead to the identification of an individual including the cultural, economic or social identifiers.

It has been argued that for enabling organisations and companies to meet their obligations under the GDPR, they need to collect metadata. To enable the rights afforded by the law, such as the right to be forgotten and to request one's data, organisations will need to collect metadata. As long as data subjects have requests regarding their data, metadata will be necessary in order to organise it. Thus, the law itself necessitates collecting metadata.

An additional aspect of legal development that currently does not apply to these messaging platforms, is the e-Privacy Directive of the European Union - as these platforms are considered to be "over-the-top" services. The European Data Protection

[12] Council of Europe. n.d. "Modernisation of Convention 108." *Council of Europe Data Protection.* Accessed September 4, 2019. https://www.coe.int/en/web/dataprotection/convention108/modernised.
[13] Direcive 95/46/EC of the European Parliament and of the Council of 24 OCtober 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of data. European Union.
[14] Article 4 – General Data Protection Regulation (GDPR)

Board (EDPB) has called for the swift adoption of an e-Privacy Regulation, a document that is currently in negotiation stages. The EDPB has acknowledged that the extensive electronic communications usage make them likely to either contain or reveal personal data, not only explicitly but also because "of mere accumulation and combination of electronic communications content or metadata, which can allow very precise conclusions concerning the private lives of the people"[15]. Therefore, although not in force at the time of writing this report, there are indications that future legal structures will explicitly address the metadata risks around communication platforms.

Frameworks Covering Metadata

Various actors have produced frameworks and policy guidelines to address some of the risks associated with metadata generation and collection. As discussed above, one of the main risks surrounding metadata is the "mosaic effect" and the ability to identify individuals and possibly their social group through the collation of metadata. A number of policy documents produced by USAID[16], The Engine Room[17], Harvard Humanitarian Initiative (HHI) Signal Program[18], the United Nations (UN)[19] and the International Coalition of the Red Cross (ICRC)[20] establish the risk assessment need to include 're-identification potential' that data holds. The ICRC is leading the work on metadata within the humanitarian sector, having published their comprehensive report "The Humanitarian Metadata Problem: Doing no Harm in the Digital Era" in 2018. This report aims to build on ICRC research by mapping "who has what kind of access to which kind of metadata, and for how long", and providing a comparative assessment of key communication channels.

This report establishes an assessment framework for metadata associated with communication platforms. It aims to provide methodology for use by policy-makers and decision-makers to carry out risk assessments regarding the use of different communication platforms, on the basis of how they collect, process and store metadata.

A major challenge faced by users trying to understand the data implications of using a communication platform, is the complexity of documents such as the terms of service and privacy policy. Recent studies show that reading a privacy policy requires at least a college level education (and sometimes much higher levels of professional experience or specialised skills). This framework provides a method that can be used to 'de-mystify' the privacy policies, terms of services and related communications from and about these platforms. The framework can be filled in by individuals or teams with the skills needed, and then distributed to others. This facilitates easier communication and decision-making about the risks of using these platforms.

---

[15] European Data Protection Board. 2018. "Statement of the EDPB on the revision of the ePrivacy Regulation and its impact on the protection of individuals with regard to the privacy and confidentiality of their communications." May 25.
[16] USAID, fhi360, and mSTAR. 2019. *Considerations for Using Data Responsibly at USAID.* USAID.
[17] The Engine Room. 2016. *The Handbook of the Modern Development Specialist.* The Engine Room.
[18] Raymond, Nathaniel A., Faine Greenwood, Caitlin Howarth, Danielle Escudero Poole, and Daniel P. Scarnecchia. 2017. *The Signal Code: A Human Rights Approach to Information During Crisis.* Harvard Humanitarian Initiative Signal Progam on Human Security and Technology.
[19] OCHA. 2019. *Data Responsibility Guidelines: Working Draft.* The Hague: UNOCHA Centre for Humanitarian Data.
[20] International Committee of the Red Cross (ICRC); Privacy International. October 2018. "The Humanitarian Metadata Problem: "Doing no harm" in the digital era."

2. ASSESSMENT FRAMEWORK

The aim of this research is to conduct a comparative assessment of various messaging platforms based on the amount of metadata produced and collected through their use. In turn, it assesses the level of privacy afforded to those who make use of the platform. In order to do this the following assessment framework, made up of the overarching categories and sub-categories has been established:

Metadata

This category is focussed on mapping the metadata collected by a messaging platform. It establishes whether the data is collected or not, the amount and type of metadata, whether the data is actively stored by the company running the platform, and how ownership of this data is determined.

Category	Sub-category	Main question	Assessment
Metadata	Collected	*Does the platform collect metadata?*	*Yes/no/unknown*
	Types	*What types of metadata are collected by the platform?*	*List*
	Storage	*Is metadata stored by the platform?*	*Yes/no/unknown*
	Ownership	*Who 'owns' the metadata?*	*Open*

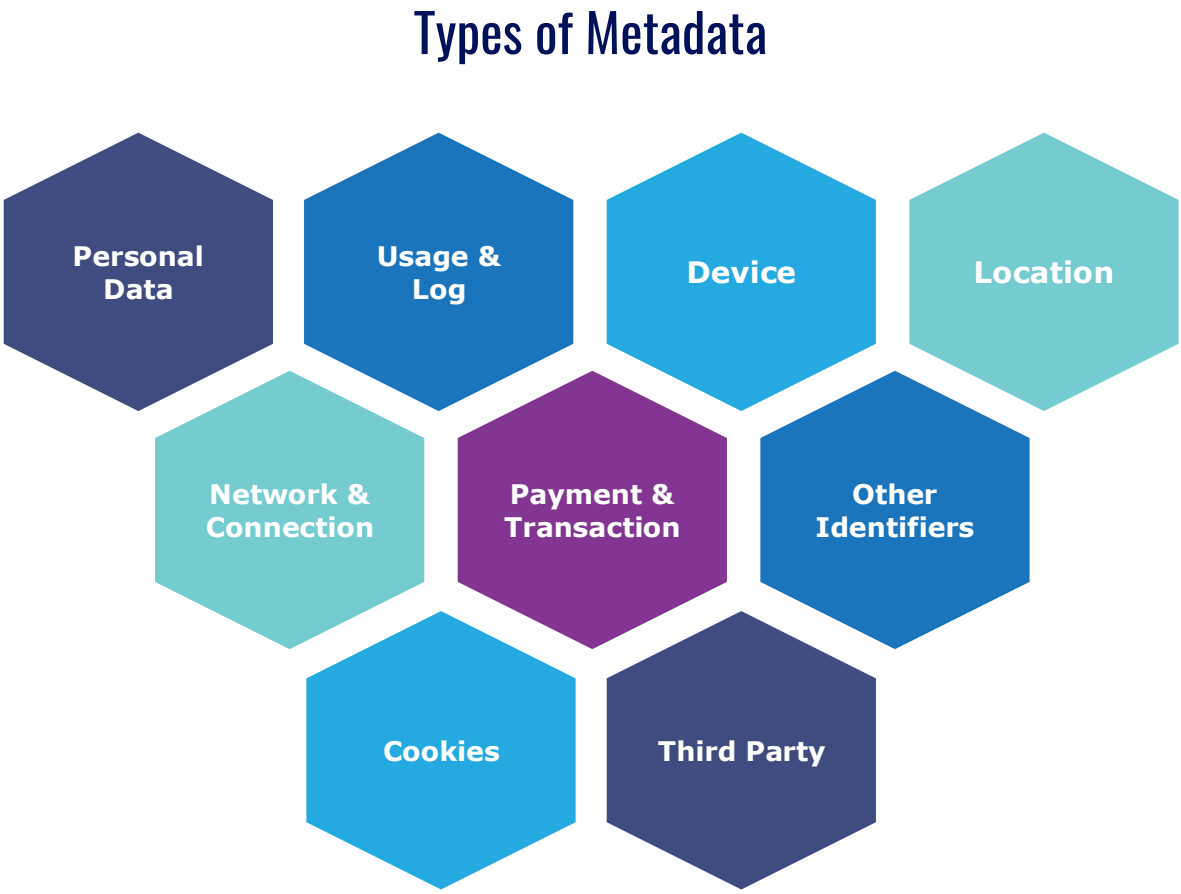


Figure 1: 9 categories of metadata collected by communication platforms

For the purposes of this framework 9 overarching types of metadata collected by messaging applications were identified.

- *Personal Data*: Including data such as profile name, address book, profile picture etc.
- *Usage & Log Data:* Including performance logs, features used, last use of service etc.
- *Device Data:* Including hardware model, operating system, app version etc.
- *Location Data:* Including cell tower location, IP address geographical location, nearby devices etc.
- *Network & Connection Data:* Including mobile network, Internet Service Provider (ISP) etc.
- *Payment & Transaction Data:* Including payment receipts, credit card number, cardholder name etc.
- *Other Identifiers:* Including other ID's (OpenID, UnionID), Facebook Unique identifiers etc.
- *Cookies*: Including collecting data from other cookies present on the device.
- *Third Party Data:* Including data from other users about the data subject, purchasing from other third-party providers etc.

Storage

Once the mapping of the metadata is complete, this category assesses the storage of said data. It establishes whether the platform has a data policy that addresses the metadata of its users, the storage location of the data, how long the data is stored for and whether the platform encrypts the data they store. This is important for understanding of risks that can arise from data breaches or general lack of security around stored metadata.

Category	Sub-category	Main question	Assessment
Storage	Policy	*Is there a metadata storage policy for the platform? Does the policy address the storage of metadata?*	*Yes/no/unknown*
	Server Location	*In which country is the data stored? Is there a choice for storage location?*	*Open*
	Retention	*What is the retention policy regarding data/metadata?*	*Limited/Indefinite*[21]
	Encryption	*Does the platform encrypt stored data (including metadata)? Does it offer 'at rest' encryption?*	*Yes/no/partially/ unknown*
	Key (encryption)	*Who controls/holds the encryption key?*	*Open*
	Protocol (encryption)	*What encryption protocol is used?*	*Open*

[21] *Limited:* Means that the data is either deleted upon request when you actively delete your account with the platform provider or that a specific retention period was stated; *Indefinite:* If information regarding retention periods was incomplete or did not clearly indicate that all data would be deleted within a certain period or it was stated that they keep the data for as long as they need it beyond providing Services to the user. (Note: Policies may state limited for specific data points - but not indicate the policy for other data points.)

Identity Management

In order to make use of a communication platform users must have a digital identity, which may (or not) be directly linked to their natural 'real-world' identity. This category establishes how the platform manages these identities, and whether it requires linking to a natural identity. It is important to establish and understand the identity management of a communications platform, in order to assess the risk of metadata being linked to a person - this linkage may allow for the targeting of an individual or group.

Category	Sub-category	Main question	Assessment
Identity Management	Registration (identification)	*What type of information is necessary for registration?*	*List*
	Registration (authentication)	*How is the registered account authenticated and/or identity verified?*	*Open*
	Account log-in (authentication)	*How is the user's identity authenticated?*	*Open*
	2-Factor Authentication	*Does the platform offer 2-factor authentication upon log-in?*	*Default/Optional/No*
	Data grouping[22]	*Does the platform undertake data grouping with the data it collects?*	*Yes/no/unknown*
	Social graph[23]	*Does the platform build social graphs (itself or with third parties)?*	*Yes/no/unknown*
	Additional Identifiers[24]	*Does the platform assign other identifiers?*	*List*

[22] Data grouping refers to the bringing together of data collected about the user from other sources (these sources may be other companies from under the same umbrella organisation or purchased from third-party services).
[23] A social graph is a graph representing the social relationships between different entities.
[24] These include identifiers such as advertising identifiers, essentially adding an extra level of identification to a user and often allowing for profiling of a user. This does not include the additional identifier used to verify the user upon log-in.

User Rights

This category aims to assess the ability of a data subject to control and mitigate risks regarding their metadata. Firstly, it indicates whether the communication platform has a transparent policy in place regarding the collection and storage of metadata. Secondly, it indicates the right of the user in relation to the platform, for requesting access and deletion of its metadata. Finally, an assessment of the policies and rights with regard to their active implementation is provided.

Category	Sub-category	Main question	Assessment
User rights	Policy	*Does the platform have a published data policy covering user rights?*	*Yes/no/unknown*
	Jurisdiction(s)	*Under which jurisdiction(s) does the platform and/or the data fall?*	*Open*
	Government access	*Does the platform grant government access under any circumstance?*	*Yes/no/unknown*
	Data sharing	*Does the platform share data with third parties[25]?*	*Yes/no/unknown*
	Right to request data	*Does the user have the right to request a copy of all data (including metadata) associated with their identity and/or account?*	*Yes/no/unknown*
	Right to be forgotten	*Does the user have the right to request all data (including metadata) associated to their identity and/or account to be deleted[26]?*	*Yes/no/unknown*
	Transparency	*Has the platform released or been assessed in a transparency report?*	*Open[27]*

[25] Including sharing within a wider group of companies under one umbrella company.
[26] If the privacy policy indicates that all data will be deleted when the account is deleted then this is considered granting the right to be forgotten.
[27] Important to note who carried out the transparency report as well as what elements of the platform were assessed and the methodology.

Content

Although this framework is aimed at assessing the metadata generated and stored by communication platforms, three aspects of the content have been included as they still play a vital role in assessing the platform. For example, the metadata is used to target an individual or group and the content is used to actively build a case for prosecution. Therefore, this still needs to be acknowledged in the framework but is not the main focus.

Category	Sub-category	Main question	Assessment
Content	Storage	*Does the platform store the content of communications?*	*Yes/Limited[28]/no/unknown*
	Encryption	*Does the platform offer end-to-end encryption?*	*Default/optional/no*
	Protocol (encryption)	*What encryption protocol does the platform use?*	*open*

[28] Limited refers the storage mechanisms whereby a platform stores the contents of a message for a limited period of time until the message is delivered to the intended recipient.

3. COMPARATIVE ASSESSMENT

Methodology

In order to complete the framework content for the communication platforms included in this comparative assessment, the information was gained from official documents of each platform[29]. The following documents were used:

- Terms of Service
- Privacy Policy (and Data Policy for Facebook products)
- Cookie Policy

If the necessary information was not available in these documents but found in other documents, these appear in the citations.

All platforms are assessed according to their default settings and data collection. Should an option exist that dramatically changes the data collection of the platforms, it is indicated in the assessment.

Platform Selection

The choice of the platforms included in the comparative assessment attempts to be geographically representative of the most commonly used communication platforms, as well as the more commonly known 'secure' apps such as Telegram and Signal[30]. The communication platforms that have been included in this study are:

- WhatsApp
- Telegram
- Signal
- WeChat
- Facebook Messenger

WhatsApp

WhatsApp was acquired by the Facebook Companies in February 2014. WhatsApp is the most commonly used messaging platform when looking at worldwide numbers, with statistics reporting 1,5 billion active monthly users as of December 2017[31]. This platform allows users to share text, image, video messages and their status as well as allowing both voice and video calls with contacts. Despite being the most popular

[29] It is important to note that a number of these platforms have multiple Terms of Service, Privacy Policies and other official documents. These tend to be applicable based on the habitual place of residence of the user and is usually related to the different applicable laws for different users.
[30] Appendix B provides an overview of the all of the existing platforms identified by the author.
[31] Statista. n.d. *Number of monthly active WhatsApp users worldwide from April 2013 to December 2017 (in millions).* Accessed September 09, 2019. https://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/

messaging platform worldwide, it faces strong competition from other applications such as Facebook Messenger in the US and WeChat in China/Asia[32].

Telegram

Telegram is currently based in Dubai, but is said to move between countries based on the changing laws of countries and security of its team members. As of March 2018, the platform had 200 million monthly active users[33] and is the leading communication platform used in Iran, Uzbekistan and Ethiopia[34]. Alongside Signal, Telegram tends to be a popular choice for users looking for secure, encrypted chat platforms.

Signal

Although Signal is not as popular as WhatsApp or Facebook Messenger, its usage is increasing especially in countries with corruption problems or where surveillance is common [35]. Signal is a privacy-oriented application offering features such as disappearing messages, end-to-end encryption and generally reduced data collection from users[36]. The platform offers similar services to WhatsApp, allowing users to share messages, image, video as well as make encrypted voice and video calls[37].

WeChat

WeChat began as a messaging platform but has evolved into platform that provides services such as shopping, payment, official accounts that can be followed, GPS functions and finding other users in your vicinities ("Make fast friends with Friend Radar")[38]. This platform is dominant in the Chinese market and has more than one billion users with reports showing the platform had 1097.6 million accounts in the fourth quarter of 2018[39].

Facebook Messenger

Any person with a Facebook profile automatically has a Facebook Messenger account. However, on mobile devices the apps for Facebook and Facebook Messenger are separate. The total number of Facebook users worldwide is over 2 billion. The number of people making use of Facebook, or at least one of its products, is ever increasing and becoming a key element of a person's digital identify. In emerging cellular markets

---

[32] Statista. n.d. *Number of monthly active WhatsApp users worldwide from April 2013 to December 2017 (in millions).* Accessed September 09, 2019. https://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/

[33] Statista. n.d. *Number of monthly active Telegram users worldwide from March 2014 to March 2018 (in millions).* Accessed Spetember 09, 2019. https://www.statista.com/statistics/234038/telegram-messenger-mau-users/.

[34] Kim, Larry. 2018. "The Top 7 Messenger Apps in the World." *Inc.* 20 September. Accessed September 09, 2019. https://www.inc.com/larry-kim/the-top-7-messenger-apps-in-world.html.

[35] Hughes, Matthew. 2018. "Signal and Telegram are growing rapidly in countries with corruption problems." *The Next Web.* 23 January. Accessed September 09, 2019. https://thenextweb.com/apps/2018/01/23/signal-and-telegram-are-growing-rapidly-in-countries-with-corruption-problems/.

[36] Newman, Lily Hay. 2018. "Encrypted Messaging Isn't Magic." *WIRED.* 14 June. Accessed September 09, 2019. https://www.wired.com/story/encrypted-messaging-isnt-magic/.

[37] McMahon, Jordan. 2017. "Ditch All Those Other Messaging Apps: Here's Why You Should Use Signal." *WIRED.* 05 November. Accessed September 06, 2019. https://www.wired.com/story/ditch-all-those-other-messaging-apps-heres-why-you-should-use-signal/.

[38] Some features are only available in certain regions.

[39] Kim, Larry. 2018. "The Top 7 Messenger Apps in the World." *Inc.* 20 September. Accessed September 09, 2019. https://www.inc.com/larry-kim/the-top-7-messenger-apps-in-world.html.

such as Myanmar the state-owned newspaper highlights the pervasiveness of Facebook in statements such as “a person without a Facebook identity is like a person without a home address” (Mclughlin 2018)[40]

[40] Mclughlin, Timothy. 2018. “How Facebook's Rise Fueled Chaos and Confusion in Myanmar.” *WIRED.* 06 July. Accessed September 06, 2019. https://www.wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar/.

4. ASSESSMENT

Category	Sub-Category	WhatsApp	Telegram	Signal	WeChat	Facebook Messenger
Metadata	Collected	Yes	Yes	Yes	Yes	Yes
	Types	Personal Data; Usage and log information; Device information; Location information; Network and connection information; Payment and transactional information; Other identifiers; Cookie information; Third-party information	Personal data; Usage and log information; Device information; Cookie information	Personal Data; Device Information; "Technical information"[41]	Personal Data; Usage and Log information; Device information; Location information; Network and connection information; Payment and transactional information; Other identifiers; Cookie Information; "Shared information – Profile data"; "Shared information – Profile media"	Personal data; Usage and log information; Device information; Network and connection information; Location information; Payment and transactional information; Other identifiers; Cookie information; Third-Party Information
	Storage	Yes	Yes	Yes	Yes	Yes
	Ownership	User/*Unknown*[42]	*Unknown*	User	User[43]	User/Facebook

[41] The exact specifications of this technical information is not specified in the Privacy Policy dated May 25, 2018, of Signal.
[42] Privacy Policy only refers to information that has been submitted by the user and does not explicitly address metadata generated through the use of the platform. Therefore, it may imply that only part of the metadata around a communication is owned by the user.
[43] Although the platform documents state that ownership remains with the user, it also states that agreeing to the terms of service of the platform grants a license to the company to make use of user data.

Storage	Policy (metadata)	Yes	Yes	Yes	Yes	Yes
	Server Location	Ireland (EU customer) United States of America (non-EU customer)	Netherlands (UK/EEA User)	*Unknown*	Canada & Hong Kong	*Unknown*
	Retention	Limited[44]	Limited	*Unknown*	Limited	Limited
	Encryption	*Unknown*	Yes[45]	Partially	Partially	*Unknown*
	Key (encryption)	*Unknown*	Multiple locations. Different location to where data is stored.	*Unknown*	*Unknown*	*Unknown*
	Protocol (encryption)	*Unknown*	*Unknown*	Signal Protocol	"SSL"	*Unknown*
Identity Management	Registration (identification)	Phone number; Username	Phone Number[46]	Phone Number	Phone Number; Facebook Account; Google Account	Real name[47]
	Registration (authentication)	Phone number	Phone Number[48]	Phone Number	Phone Number	Email address/phone number[49] Drivers license/passport/birth certificate[50]

[44] WhatsApp only provides a time limitation for the storage of messages. It is unclear how long other data is stored for.
[45] Only state that they encrypt personal data and not all metadata. Therefore, only part of the metadata related to the communication.
[46] Telegram Privacy Policy (August 14, 2018);
Telgram. n.d. *User Authorisation.* Accessed September 06, 2019. https://core.telegram.org/api/auth.
[47] facebook. n.d. *What names are allowed on Facebook?* Accessed September 06, 2019. https://www.facebook.com/help/112146705538576
[48] Telgram. n.d. *User Authorisation.* Accessed September 06, 2019. https://core.telegram.org/api/auth.
[49] facebook. n.d. *Create an Account.* Accessed September 06, 2019. https://www.facebook.com/help/basics.
[50] "Verifying your identity could mean sending Facebook documents like a driver's licence, passport, or birth certificate" facebook. n.d. *What types of ID does Facebook accept?* Accessed September 06, 2019. https://www.facebook.com/help/159096464162185.

	Account log-in (authentication)	*Unknown*	Phone Number[51]	Phone Number	*Unknown*	Email address; phone number; username; password[52]
	2-Factor Authentication	Optional[53]	Yes	Optional[54]	Optional	Optional[55]
	Data grouping	Yes	*Unknown*	No	No	Yes
	Social Graph	-	Yes[56]	No	Yes	Yes
	Additional Identifiers	Device identifiers; Unique Facebook Company Product identifiers associated to device or account; Cookie information	Cookie information	No	QQID Facebook Connect Token Cookie information	Unique identifiers; Device IDs "other identifiers, such as from games, apps, or accounts you use"; Family Device IDs; Facebook Company Product Unique Identifiers
User Rights	Policy	Yes	Yes	Yes	Yes	Yes
	Jurisdiction	Republic of Ireland (WhatsApp Ireland Limited) United States of America (WhatsApp Inc.)	British Virgin Islands (Telegram Group Inc); Dubai (Telegram FZ-LLC)	USA	Netherlands (EU users) Singapore (non-EU users)	Republic of Ireland United States of America
	Government Access	Yes	Yes	Yes	Yes	Yes
	Data Sharing	Yes	Yes	Yes	Yes	Yes

[51] Telegram. n.d. *FAQ: Troubleshooting.* Accessed September 06, 2019. https://telegram.org/faq#login-and-sms.
[52] facebook. n.d. *Log Into Your Account.* Accessed September 06, 2019. https://www.facebook.com/help/105803362095509/?helpref=hc_fnav.
[53] WhatsApp. n.d. "Using two-step verification." *WhatsApp.* Accessed September 06, 2019. https://faq.whatsapp.com/en/general/26000021/?category=5245245.
[54] Tested in app.
[55] facebook. n.d. *What is two-factor authenticaion and how does it work?* Accessed September 06, 2019
[56] "Retain your social graph" (Telegram Privacy Policy August 14, 2018)

	Right to request data	Yes	Yes	Yes	Yes	Yes
	Right to be forgotten	Yes	Yes	Yes[57]	Yes	Yes[58]
	Transparency	Yes[59]	*Unknown*[60]	No official reports released but blog does provide insight into government requests etc.[61]	No	Self-published: Only covers elements of the Data Policy[62]
Content	Storage	Limited	Yes	No	No	Yes
	End-to- End	Default	Optional	Default	*Unknown*	Optional[63]
	Protocol (encryption)	Signal Protocol[64]	MTProto 2.0 (server-client); 256-bit symmetric AES [65] encryption (client-client); 2048-bit RSA encryption (client-client)	Signal Protocol	*Unknown*	*Unknown*

[57] According to Signal two alternatives are provided "permanently disabling your phone number from being registered as a Signal User" or "Delete account" through the app. These would appear to have differing degrees of permanence.

[58] Only the items that posted by the user – items posted by other users posted about the user concerned are not included.

[59] Cardozo, Nate, Andrew Crocker, Jennifer Lynch, Kurt Opsahl, and Rainey Reitman. 2017. *Who Has Your Back?* Electronic Frontier Foundation. facebook Transparency. 2019. *Facebook Transparency Report.* Accessed September 06, 2019. https://transparency.facebook.com/.

[60] Transparency report is publicised by Telegram but is only available through the Telegram app and cannot be viewed without a Telegram account.

[61] https://signal.org/blog/

[62] facebook Transparency. 2019. *Facebook Transparency Report.* Accessed September 06, 2019. https://transparency.facebook.com/

[63] https://www.facebook.com/help/messenger-app/1084673321594605/

[64] WhatsApp. 2017. *WhatsApp Encryption Overview.* Technical White Paper, WhatsApp.

[65] Telegram. n.d. *FAQ: Security.* Accessed September 06, 2019. https://telegram.org/faq#q-can-i-run-telegram-using-my-own-server.

5. DISCUSSION

ANALYSIS: The Platforms

Metadata

It is not surprising that all the platforms collect metadata, as this is necessary for the functioning of a communication platform. However, the difference in amount of data collected by each of them provides an interesting point for discussion. When looking at the categories of metadata provided in Section 5, WeChat collects the most metadata. It collects data from all nine categories, but also two extra categories "Shared information – Profile Data" and "Shared information – Profile media". These two categories were recorded separately, as they are provided as distinct data in WeChat's Privacy Policy. The Signal app collects the least amount of metadata, which is in line with the strong stance they have taken towards data privacy and data minimisation.

The amount of metadata collected by WeChat and Facebook Messenger could be influenced by both of these platforms being social media platforms that include a 'chat' functionality rather than just purely a communication platform. When looking at the platforms that are purely used for communication, WhatsApp collects all 9 categories of metadata whilst, Telegram and Signal only collect 4 and 3 categories respectively.

Mapping out this first category as accurately as possible is vital in order to enable a sufficient risk model, associated with different platforms. If less metadata is collected, it for example reduces the risk associated with not storing the data encrypted, data coupling etc. Therefore, just based on this category alone it appears that Signal and Telegram are the more secure of the five with regards to amount of metadata collected.

Storage

The first sub-category of storage looks at whether the platforms have a policy addressing the metadata collected. Although all the platforms have policies that address metadata, the extent to which it is covered differs. Because this framework includes Personal Data as an element of metadata generated on these platforms, all the policies cover this. However, when for example looking into the details of how network and connection metadata is handled, the information provided varies greatly. This directly effects the ability to answer the question of retention policies for each of the platforms. For example, despite each of the platforms stating that they store the data for a limited period of time, it is not clear whether this applies to all data connected to a data subjects' profile or only the Personal Data that can directly lead to the identification of an individual

Another important element of storage is whether it is encrypted at rest, i.e. when it is not being sent between devices. Often the discussion about encryption and communication platforms focusses on the protection afforded messages as they travel

from one device to another. However, the metadata stored on the platform's servers may not be encrypted. The metadata connected to the data subjects who make use of these platforms will be stored at rest on servers. If un-encrypted, it presents a greater risk than if it were encrypted. Should these servers be accessed by malicious actors, all the data will be instantly 'readable'.

Another important element related to storage and encryption is the control of the encryption key, as this key is used to 'read' the data when it is stored in an encrypted format. Storing this key in the same location as the data presents a greater risk than if it is stored separately. Telegram is the one communication platform analysed in this report that has taken this into account in its privacy policy, stating that they have not only encrypted all data at rest, but also store their encryption keys in multiple locations separate from the data.

When assessing the importance of the encryption at rest, it must be viewed in relation to the amount of data collected. WhatsApp collects a significant amount of data that could lead to direct or indirect identification of the data subject(s), and at the same time does not indicate whether this data is encrypted on their servers. Signal on the other hand states that only part of the data stored (profile information) is encrypted. However, because they collect very little data the risk is significantly lower: Even if the unencrypted data is accessed, the smaller amount reduces the likelihood that it could lead to direct or indirect identification of the data subject(s).

Identity Management

The way in which a communication platform manages the identities of data subjects is important, as it indicates the ease with which data subjects' real identity can be found. In this category there are three stages at which the identity of a data subject is managed: The identity required to register with the platform (registration), the method used to verify that identity (authentication), and the identification method used every time the account is used (account log-in).

All of the platforms except Facebook and WeChat, use phone numbers as the main identifier of a data subject/user. This means that when opening the account this is the main piece of data to which everything else is associated. Facebook requires that the main identity is the real name of the account holder, whilst WeChat allows for the main identifier to be a phone number, a Facebook account or a Google account. This is the first layer of identity management, which is followed by verification of this main registration identity.

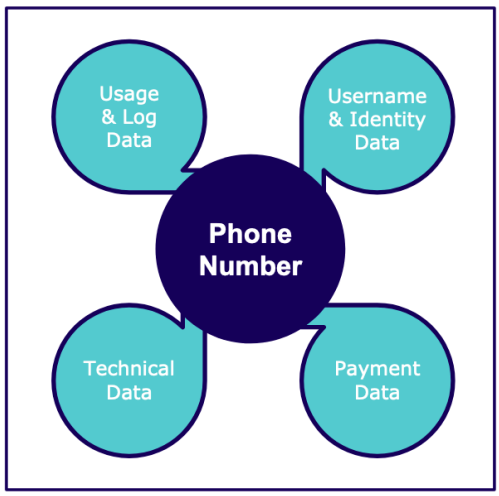


Figure 2: *How different data points relate to the main identification datum of the user.*

All of the platforms verify the identity of the account using the phone number provided at the account registration. However, as Facebook's policy requires that the user account identity be the real name of the user, they state that they can request direct Personal Data in order to verify this name. On top of using a phone number to verify

the real name, Facebook state they can use an email address, drivers' licence, passport or birth certificate. Therefore, Facebook Messenger carries the highest risk of identification through metadata attached to the account, as the account already carries the real-world identity of the data subject.

User Rights

It is important to evaluate the rights that a user has, regarding all of the data that may directly or indirectly lead to their identification. Platforms that grant these rights allow users to maintain a level of control over their own safety and security, based on their own evaluation of their situation. Although all of the platforms grant access to governments entities[66], the amount of data collected by the platform becomes an important factor in evaluating the level of risk associated with this sharing. In 2016 the Signal messenger app published a number of documents related to a grand jury subpoena for user data in the Eastern District of Virginia, USA. In the related blog they reiterated the number of data points that they do not have, and therefore cannot share when complying with such a request[67]. This demonstrated the importance of the first category, assessing amount of metadata collected.

The sub-categories of implementation and transparency are key elements in determining the implementation of the privacy policy, and the extent to which the platform adheres to its stated obligations. Through the assessment of these five platforms, it becomes clear that these two elements can either be self-assessed (Facebook publishes its own transparency report) or assessed by an independent organisation such as the Electronic Frontier Foundation. Of the five platforms studied in this report, WhatsApp was the only one assessed by an independent organisation whilst WeChat did not appear in any transparency report. Facebook Messenger was self-assessed and only elements of the data policy were covered, not the entirety of the data involved in the use of the platform. Although Signal does not have its own collated report, it does have a regularly updated blog in which it provides a high level of transparency, releasing information such as the aforementioned subpoena from a US court.

Content

Although this report is focussed on the metadata of messages, it is important to always note whether the content of the messages is collected. The collection of message content inherently increases the risk associated to the metadata.

Application: Building a Risk Model

This framework aims to assist organisations in understanding the metadata surrounding messaging platforms and associated risks. A thorough understanding is necessary, in order to build a comprehensive risk model that can be used by those involved in decision-making processes. In order to build a risk model and understand the risk to users, the results of the framework must be analysed in light of the project

[66] This includes regulators, law enforcement and other entities that may be able to legally request access to the data in certain circumstances.
[67] Signal. 2016. *Grand jury subpoena for Signal user data, Eastern District Virgina.* 04 October. Accessed September 01, 2019

context and the aim of the communication. For example, the use of a phone number for identification may be considered 'low-risk' in a context where these phone numbers are not necessarily attached to a 'real-name' identification, as it is possible to anonymously buy sim cards. However, in other contexts every purchase of a sim card must be linked to a national identity number as well as physical address. This has a significant effect on how one analyses the risk associated with that single element of the framework. Each element requires similar analysis of the context.

Throughout the framework, there are points of information that were not available. These unknowns should be assessed in light of the amount of metadata that is collected by the platform. The larger the amount of data collected the more significant impact the unknown element should have on a risk model. Therefore, the 'unknowns' for Signal are less risky than those for example associated to Facebook Messenger.

When using such a framework, regular reassessments need to be established for the duration of the project or use of the communication method. Ensuring that the assessment is based upon the latest platform documents as well as latest contextual information, ensures that the risk model presents the latest level of risk. Should the updates result in significant changes to the risk model, this should be considered. Incorporating this into a project plan from the start will provide protection for all data subjects involved, as well as the project itself. This may also play an important role in choosing which platform is used for communication.

The level of education needed for understanding the policies of many of the commercial companies is very high. Having the information synthesized by the above framework provides a mechanism for builders of threat models in their contexts, to easily understand the metadata collection of different platforms.

6. CONCLUSION

All of the platforms analysed in this report collect amounts and types of metadata that could lead to the identification of individuals and possibly their group affiliations, without needing access to the content of the messages. Additionally, there is a general lack of transparency around metadata that is not considered to be directly identifiable personal data. Therefore, the manner in which the data is managed by the platforms is hard to determine.

An organisation using any of these and any other platform will need to pay close attention to the policies, documents, updates and emerging investigations of them. Even the smallest change can have significant effect on the level of security of the data subject(s) with whom they are communicating, and the overall work of the organisation. Although this framework incorporates user rights in order to evaluate the level of security that the data subject can control, the organisation should not rely on the data subject's judgement. The level of analysis necessary to fully grasp the implications of the platform operations, is too high for relying on the expectation that the average data subject will understand the implications. This is clearly demonstrated in the level of education needed to understand the privacy policies of platforms such as Facebook[68]. Therefore, any project that requires communication through a third-party platform must include reassessing the policies using this framework, and necessary adjustments to the threat model based on changes throughout the projects lifetime.

In highly sensitive contexts, relying on these commercial platforms may be problematic for these reasons. Also, as the context changes it may require changes to which platform is used – thus causing a lack of continuity in the methods of communication with target audiences or beneficiaries. In these circumstances, an organisation should use a communication platform over which they can maintain oversight and control of data. In a generic setting (lowest risk context), it appears that Signal is the most secure of the assessed platforms.

Critical Reflection

Whilst drafting this report there were a number of challenges that affected the collection and assessment of information, concerning each of the communication platforms included. In order to ensure veracity of the information used for assessment, it is recommended that official documents or reports that include references to where the information was found are used. However, not all the information necessary to fill out the assessment framework is available from such documents. Another method for obtaining such information is the secure testing of the platforms themselves.

Another challenge regarding the official documents such as the privacy policy is the difficulty of reading and understanding the implications of statements included in

[68] Litman-Navarro, Kevin. n.d. "We Read 150 Privacy Policies. They Were an Incomprehensible Disaster." *The New York Times.* Accessed January 24, 2020. https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html.

these policies. Recent studies have shown that reading a privacy policy requires at least a college level education, and in some cases much higher levels of professional experience or specialised skills. Even with specialised knowledge in the area of data and related risks, certain assumptions needed to be made about the implications of statements due to language used, opacity of the documents, vagueness of information etc.

Due to these challenges of collecting relevant information, this framework provides a method that can be used to 'de-mystify' the privacy policies, terms of services and related implications of these platforms. The framework can be filled in by individuals or teams with the skills needed and then distributed to others. This facilitates easier communication about the risks of using these platforms.

When information is readily available, a number of assumptions may need to be made in order to fill out the information. Firstly, some of the platforms offer a number of different options for privacy and security. For example, Signal allows for 'disappearing messages' that only remain visible in a chat for the time chosen by the users. However, this is not a default setting. Many platforms have such optional settings that would reduce the amount of data collected. For the assessment carried out above, only the default settings of the platforms were assessed. This allows for an indirect assessment of the platforms level of privacy and security by design and default.

Another related challenge is the links between certain commonly used messaging platforms and social media profiles (e.g. Facebook Messenger and WeChat). It is impossible to separate the data that is collected between the social media side, in which options such as purchasing games may be included, versus what data is collected when the profile is used for purely messaging purposes. Therefore, it is important to include all data generated by the platform and not just the use of the associated 'messenger'.

Bibliography

(EDPB), European Data Protection Board. 2018. “Statement of the EDPB on the revision of the ePrivacy Regulation and its impact on the protectiuon of individuals with regard to the privacy and confidentiality of their communications.” 25 May.

Blond, Stevens Le, Alejandro Cuevas, Philipp Jovanovic Juan Ramón Toncoso-Pastoriza, Bryan Ford, and Jean-Pierre Hubauz. 2018. “On Enforcing the Digital Immunity of a Large Humanitarian Organisation.” *2018 IEEE Symposium on Security and Privacy* 424 - 440.

Cardozo, Nate, Andrew Crocker, Jennifer Lynch, Kurt Opsahl, and Rainey Reitman. 2017. *Who Has Your Back?* Electronic Frontier Foundation.

Council of Europe. n.d. “Modernisation of Convention 108.” *Council of Europ Data Protection*. Accessed September 4, 2019. https://www.coe.int/en/web/data-protection/convention108/modernised.

n.d. “Direcive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individauls with regard to the processing of personal data and on the free movement of data.” European Parliament and The Council.

facebook. n.d. *Create an Account*. Accessed September 06, 2019. https://www.facebook.com/help/basics.

—. 2018. *Data Policy*. 19 April. Accessed September 06, 2019. https://www.facebook.com/policy.php.

—. n.d. *Log Into Your Account*. Accessed September 06, 2019. https://www.facebook.com/help/1058033620955509/?helpref=hc_fnav.

facebook Transparency. 2019. *Facebook Transparency Report*. Accessed September 06, 2019. https://transparency.facebook.com/.

facebook. n.d. *What is two-factor authenticaion and how does it work?* Accessed September 06, 2019. https://www.facebook.com/help/148233965247823?helpref=faq_content.

—. n.d. *What names are allowed on Facebook?* Accessed September 06, 2019. https://www.facebook.com/help/112146705538576.

—. n.d. *What types of ID does Facebook accept?* Accessed September 06, 2019. https://www.facebook.com/help/159096464162185.

Greene, Kate. 2008. “TR10: Reality Mining.” *MIT Technology Review*. 19 February. Accessed August 28, 2019. http://www2.technologyreview.com/news/409598/tr10-reality-mining/.

Hill, Kashmir. 2018. “'People You May Know': A Controversial Facebook Feature's 10-Year History.” *Gizmodo*. 9 August. Accessed September 4, 2019. https://www.gizmodo.com.au/2018/08/people-you-may-know-a-controversial-facebook-features-10-year-history/.

Hughes, Matthew. 2018. “Signal and Telegram are growing rapidly in countries with corruption problems.” *The Next Web*. 23 January. Accessed September 09, 2019. https://thenextweb.com/apps/2018/01/23/signal-and-telegram-are-growing-rapidly-in-countries-with-corruption-problems/.

ICRC; The Engine Room; Block Party. January 2017. “Humanitarian Futures for Messaging Apps.”

International Committee of the Red Cross (ICRC); Privacy International. October 2018. “The Humanitarian Metadata Problem: "Doing no harm" in the digital era.”
Kim, Larry. 2018. “The Top 7 Messenger Apps in the World.” *Inc.* 20 September. Accessed September 09, 2019. https://www.inc.com/larry-kim/the-top-7-messenger-apps-in-world.html.
Litman-Navarro, Kevin. n.d. “We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.” *The New York Times*. Accessed January 24, 2020. https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html.
Mayer, Jonathan, Patrick Mutchler, and John C. Mitchell. 2016. “Evaluating the privacy properties of telephone metadata.” *PNAS* (PNAS) 5536 - 5541.
Mclughlin, Timothy. 2018. “How Facebook's Rise Fueled Chaos and Confusion in Myanmar.” *WIRED*. 06 July. Accessed September 06, 2019. https://www.wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar/.
McMahon, Jordan. 2017. “Ditch All Those Other Messaging Apps: Here's Why You Should Use Signal.” *WIRED*. 05 November. Accessed September 06, 2019. https://www.wired.com/story/ditch-all-those-other-messaging-apps-heres-why-you-should-use-signal/.
Montjoye, Yves-Alexandre de, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. *Unique in the Crowd: The privacy bounds of human mobility*. Scientific Report, Nature.
Montjoye, Yves-Alexandre de, Erez Shmueli, Samuel S. Wang, and Alex Sandy Pentland. 2014. “openPDS: Protecting the Privacy of Metadata through SafeAnswers.” *PLoS ONE*.
Newman, Lily Hay. 2018. “Encrypted Messaging Isn't Magic.” *WIRED*. 14 June. Accessed September 09, 2019. https://www.wired.com/story/encrypted-messaging-isnt-magic/.
OCHA. 2019. *Data Responsibility Guidelines: Working Draft*. The Hague: UNOCHA Centre for Humanitarian Data.
Pearson, Michael. 2013. “Obama: No one listening to your calls.” *CNN Politics*. 10 June. Accessed August 28, 2019. https://edition.cnn.com/2013/06/07/politics/nsa-data-mining/index.html.
Raymond, Nathaniel A., Faine Greenwood, Caitlin Howarth, Danielle Escudero Poole, and Daniel P. Scarnecchia. 2017. *The Signal Code: A Human Rights Approach to Information During Crisis*. Harvard Humanitarian Initiative Signal Progam on Human Security and Technology.
n.d. “Regulation (EU) 2016/679 of the European Parliamnet and of the Council .” European Union.
Scahill, Jeremy, and Glenn Greenwald, interview by Amy Goodman. 2014. *Death By Metadata: Jeremy Scahill& Glenn Greenwald Reveal NSA Role in Assassinations Overseas* (10 February).
Signal. 2016. *Grand jury subpoena for Signal user data, Eastern District Virgina*. 04 October. Accessed September 01, 2019. https://signal.org/bigbrother/eastern-virginia-grand-jury/.
—. 2018. *Privacy Policy*. 25 May. Accessed September 06, 2019. https://signal.org/legal/#privacy-policy.
Statista. n.d. *Number of monthly active Telegram users worldwide from March 2014 to March 2018 (in millions)*. Accessed September 09, 2019.

https://www.statista.com/statistics/234038/telegram-messenger-mau-users/.
—. n.d. *Number of monthly active WhatsApp users worldwide from April 2013 to December 2017 (in millions)*. Accessed September 09, 2019. https://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/.
Telegram. n.d. *FAQ: Security*. Accessed September 06, 2019. https://telegram.org/faq#q-can-i-run-telegram-using-my-own-server.
—. n.d. *FAQ: Troubleshooting*. Accessed September 06, 2019. https://telegram.org/faq#login-and-sms.
—. 2018. "Telegram Privacy Policy." *Telgram*. 14 August. Accessed September 06, 2019. https://telegram.org/privacy.
Telegram. n.d. *User Authorisation*. Accessed September 06, 2019. https://core.telegram.org/api/auth.
The Engine Room. 2016. *The Handbook of the Modern Development Specialist*. The Engine Room.
USAID, fhi360, and mSTAR. 2019. *Considerations for Using Data Responsibly at USAID*. USAID.
Vota, Wayan. 2018. *5 Guidelines for Using Mobile Messaging Applications in International Development*. 10 September. Accessed January 25, 2019. https://www.ictworks.org/messaging-applications-international-development/#.XGKOO-JKjOR.
WeChat. 2018. *Privacy Policy*. 10 May. Accessed September 06, 2019. https://www.wechat.com/en/privacy_policy.html.
WhatsApp. n.d. "Using two-step verification." *WhatsApp*. Accessed September 06, 2019. https://faq.whatsapp.com/en/general/26000021/?category=5245245.
WhatsApp. 2017. *WhatsApp Encryption Overview*. Technical White Paper, WhatsApp.
—. 2018. "WhatsApp Privacy Policy." 24 April. Accessed September 06, 2019. https://www.whatsapp.com/legal/?eea=1#privacy-policy.
—. 2016. "WhatsApp Privacy Policy." 25 August. Accessed September 06, 2019. https://www.whatsapp.com/legal?eea=0#privacy-policy.

Appendix A – Longlists of data collected

- Personal Data
- Usage and log information
- Device information
- Location information
- Network and connection information
- Payment and transactional information
- Other Identifiers
- Cookie information
- Third party information

WhatsApp Inc. and WhatsApp Ireland Limited

Personal Data

- Phone number
- Email address (optional)
- Profile Name
- Address book
- Date of registration
- Status
    - Online
    - Last seen
    - Last status message update

Usage and log information

- Log files
- Diagnostics
- Crash information
- Websites
- Performance logs and reports
- Features used
    - Messaging
    - Calling
    - Status
    - Group features
    - Profile photo
    - 'About' information
- Whether you are online
- When you last used the Service
- When you last updated your information

Device information

- Hardware Model
- Operating System
- Battery level
- Signal strength
- App version

- Browser information
- Device operation information

Location information (optional *-if you use the location features)*

- IP
- Bluetooth Signals
- Nearby Wifi access points, beacons and cell towers

Network and connection information

- Mobile network
- Connection information
    - Phone number
    - Mobile operator or Internet Service Provider (ISP)
    - Language
    - Timezone
    - IP

Payment and transactional information

- Payment receipt (from app stores or third parties processing payments)

Other Identifiers

- Device Identifiers
    - Including identifiers unique to Facebook Company Products associated with the same device or account

Cookie information

Third party information

- From other users about you
- Third-party providers
- Third-party service providers

Telegram

Personal Data

- Phone number and contacts (Of others)
    - Mobile Number
    - First Name
    - Last Name
- Account data (Personal Data)
    - Mobile phone number
- Account data
    - Profile name
    - Profile picture
    - About information
    - Username (optional)
    - Email address (optional)
    - IP address

Usage and log information
- Past Telegram apps
- History of username changes

Device Information
- Device data

Cookie Information
- Cookies

[CONTENT] Cloud Chats
- Messages
- Photos
- Videos
- Documents

Signal

Personal Data
- Phone Number
- Profile Name (Optional)
- Profile Picture (Optional)
- Contacts (Optional)

Device Information
- Authentication tokens
- Keys
- Push tokens
- “Other material”

Technical Information
- The exact specifications of what this technical information includes is not specified in the Privacy Policy from May 25, 2019 of Signal.

WeChat

Personal Data
- Name
- User alias (nickname)
- Mobile phone number
- Password
- Gender
- IP address
- Profile ID
- Photo
- Voiceprint (optional)
- Emergency Contacts
- Email address

- Facebook Connect OpenID
- QQ ID
- Contact list
- Pseudoanonymised and aggregated personal information
- Verified Account Data (in available jurisdictions)
    - Nationality
    - Date of birth
    - Residential address
    - Copy of residential address proof
    - Copy of personal identification document
    - Personal identification document number
    - Occupation
    - Source of funds

Usage and log information

- Record of search inquiries
- Mobile carrier-related information
- Configuration information (made available by your web browser or other programs you use to access WeChat)

Device information

- Managed Devices
- Device version number
- Device identification number
- Media stored on your device

Location information

- GPS
- WiFi
- Compass
- Accelerometer
- Location information
- IP Address (derived)
- Device (derived)
- Internet Service Provider (derived)

Network and connection information

- Mobile carrier-related information
- WiFi
- Internet Service Provider
-

Payment information (in available jurisdictions)

- Credit card number
- Expiry
- CVC
- Cardholder name

Other Identifiers

- Facebook Connect Token
- OpenID
- UnionID
- QQ ID

Cookie information

"Shared Information – Profile Data"

- "Any information that you include in your publicly-visible profile, which may include your profile ID, name and photo"

"Shared Information – Profile Media"

- "This comprises all of the information you make available to other users via WeChat, comprising WeChat Moments posts and responding to other users' WeChat Moments; and information made available by another user about you via their use of WeChat – for example, any Shared Information that others using WeChat make available about you via WeChat Moments and communication they make to you and other using WeChat"

Facebook Messenger

Personal Data

- People connected to
- Pages connected to
- Accounts connected to
- Hashtags connected to
- Groups connected to
  - Interaction with all of the above
- Contact information (optional)
  - Address book
  - Call log
  - SMS log history

Usage and log information

- Content viewed/engaged
- Features used
- Actions taken
- Interaction with people/accounts
- Time of activities
- Frequency of activities
- Duration of activities
  - Usage log
    - Times of use
    - Last time of use

Device Information ("we collect information from and about the computers, phones, connected TVs and other web-connected devices you use that integrate with our Products, and we combine this information across different devices you use)

- Device attributes
  - Operating system

    - Hardware versions
    - Software versions
    - Battery level
    - Signal strength
    - Available storage space
    - Browser type
    - App name
    - File Name
    - App types
    - File types
    - Plugins
- Device operations
    - Foregrounding of window
    - Backgrounding of app
    - Mouse movements

Network and connection information

Device signals

- Bluetooth signals
- Information about nearby Wi-Fi access points
- Beacons
- Cell towers

Network providers

- Mobile operator
- Internet Service Provider (ISP)
- Language
- Mobile phone number
- IP address
- Connection speed
- Information about other device nearby or connected to the same network

Data from device settings

- Camera (optional)
- Photos (optional)

Location information

- Information and content you provide "can include information in or about the content you provide (like metadata)", location of a photo, date a file was created.
- GPS location (optional)
- Time zone

Payment and transactional information

    - Payment information
        - Card numbers
    - Account information
    - Authentication information
    - Billing details
    - Shipping details
    - Contact details

Other Identifiers

- Unique identifiers
- Device IDs
- Other identifiers (from other companies)
    - Games
    - Apps
    - Accounts
    - Family Device IDs
- Facebook Company Products unique identifiers

Cookie data

- Cookie IDs
- Cookie settings

CENTRE FOR
INNOVATION
Leiden University