Mobile Privacy and Big Data Analytics

About the GSMA

The GSMA represents the interests of mobile operators worldwide, uniting nearly 800 operators with almost 300 companies in the broader mobile ecosystem, including handset and device makers, software companies, equipment providers and internet companies, as well as organisations in adjacent industry sectors. The GSMA also produces industry-leading events such as Mobile World Congress, Mobile World Congress Shanghai and the Mobile 360 Series conferences.

For more information, please visit the GSMA corporate website at www.gsma.com

Follow the GSMA on Twitter: @GSMA

Introduction

Big data analytics and data driven services play a critical role in digital life and will continue to do so in the future. By using big data analytics to enhance connected cars, smarter homes, smarter cities and smarter health systems, we can have a positive impact on societal aims such as the UN Sustainable Development Goals and deliver more effective health outcomes, better environmental management, increased opportunities for learning and improved goods and services for consumers. In short, big data analytics changes the way we live – for the better.

The considerations in this document comprise safeguards to think about when engaging in big data analytics activities involving personal data. They should be considered alongside existing GSMA material on data privacy topics[1].

A key driver of big data analytics is the Internet of Things (IoT) which is giving rise to a connected world, where an ever-increasing number of devices with sensors that collect and communicate data are being deployed across a wide spectrum of use cases.

Big data analytics and IoT depend both on the availability of data and on consumer trust. The mobile industry is determined to help realise the economic and societal benefits of big data analytics through good digital responsibility practices, so that society can unlock the huge potential of big data analytics in a way that respects well established privacy principles and fosters an environment of trust.

These considerations are intended to encourage good data privacy practices in the context of big data analytics.

'Big data analytics'

In this document we have used the term 'big data analytics' to mean the use by any organisation of big data analytics techniques or the services provided by them across a range of scenarios including:

- Mobile network operators (MNOs) using their in-house analytics services to provide users with enhanced services;
- Making selected data available to an analytics service provider for them to conduct analytics and return insights;
- Making selected insights available to an application provider for them to deliver an enriched service to subscribers opting into the enhanced service;
- Making de-identified or pseudonymised data available via an API (application programming interface) or a data feed to third parties for them to conduct their own analytics using their own service providers;
- Several parties making their respective data available under a common framework to allow third parties and their service providers to access the data via an API for them to conduct their own analytics;
- Several parties transmitting their data to an analytics service provider to conduct analytics and return insights on all the data; and
- Several parties transmitting their data to a common hosting service provider that provides access via an API to third parties for them to conduct their own analytics using their own service providers.

[1] See The GSMA Mobile Privacy Principles; The GSMA Mobile Policy Handbook position on Privacy and Big Data; GSMA guidelines on the protection of privacy in the use of mobile phone data for responding to the Ebola outbreak; The GSMA Privacy Design Guidelines for Mobile Application Development

THE FOLLOWING EXAMPLES ARE INTENDED TO ILLUSTRATE SOME OF THE DATA PRIVACY CONSIDERATIONS THAT MAY BE TAKEN INTO ACCOUNT:

SCENARIO 1

A local transport authority is seeking to understand city inhabitants' travel patterns in order to respond more effectively to customer demand. For example, it would like to understand the likely increase in public transport usage during inclement weather. A big data service to address this need is being provided by an MNO in partnership with a third party data analytics firm. The data being utilised is device location data, originating with the MNO, and weather data from IoT sensors deployed by a meteorological organisation that capture temperature, air pressure and precipitation. The MNO ingests the meteorological data and then sends both the de-identified location data and the weather data to the data analytics partner who undertakes the analytics and provides the results to the transport authority.

SCENARIO 2

In the aftermath of a natural disaster, there is a fear that disease will break out. The government and relevant NGOs wish to gain access to de-identified device location data held by MNOs in order to combine it with existing government records, live health data from medical professionals, historical data from previous disasters and other context data, such as weather, to understand the risks. The government would also like to send alerts to specific citizens about potential risks and suggested actions, if required. The initiative is being run by the World Health Organisation (WHO) which is collecting the data, undertaking the analysis and working with the government to organise the response. The WHO accesses the de-identified device data and combines it with the other data to run the analytics. As a result of this analysis, certain groups of individuals whose identity is not known to the WHO are determined to be at risk of a life-threatening disease based on their location. This data is relayed back to the respective MNOs who then send them a group alert and action message.

SCENARIO 3

A family lives in a smart home that contains systems and devices such as a smart central heating system that can be controlled remotely from their smartphones, as well as smart televisions that can track what the family watch and that listen for voice commands. Other smart products including light bulbs, movement sensors, doors and appliances report data about their status to a central platform when they are in use. This smart home service is provided by an MNO who is operating the central platform for the service.

3A: The family is also signed up to a home security service offered by a third party service provider, and have requested for their smart home data to be shared with the home security service provider so that the service can be delivered.

3B: An external company wishes to utilise data from the MNO's smart home service, in order to understand behavioural patterns in different regions of the country and formulate marketing insights.

SCENARIO 4

A MNO uses call detail records, sales data and network performance monitoring in order to optimise the network, keep the network secure, detect fraud, improve customer services and conduct targeted marketing.

In many instances de-identified or aggregated data is sufficient, for example to understand traffic volumes at certain locations. However, in other instances, for example, where the MNO wishes to understand how the network performs for a specific customer in order to offer them future discounts or credits or in order to identify those customers who might be on the wrong tariff plan, identifiable information is needed.

Privacy considerations

In order to realise the potential societal and economic benefits of big data analytics in a way that is compatible with recognised data privacy principles, the following considerations may be taken into account:

Personal data

Much of the data used in IoT and big data services is not personal data[2]. Readings from weather sensors, for example, would not constitute personal data.

Big data analytics services should take into account that such non-personal data can become personal data if it is associated with a particular individual, for example, if the location of a connected car detected by a traffic management system is subsequently combined with the vehicle registration number and the vehicle ownership records.

Big data analytics services can consider guarding against the possibility of re-identification of individuals when the data is merged with other data sets.

Where personal data is collected, for example, when a mobile phone user's location is recorded, this can be de-identified through the removal of data fields that enable identification and through reporting the analytic insights only in aggregate or approximated form, as in Scenario 1.

Example

Of course, there are situations where identifiable information is needed. For example, in Scenario 4, the MNO may wish to conduct analytics on call detail records and network performance data, but then identify individual customers who regularly experience poor network issues or appear to be on the wrong tariff so that they can offer them a credit or different tariff based on the data that is specific to them.

Transparency, control and purpose

In the context of big data analytics services, providing fair notice *before* collection can be challenging. Big data analytics services are often designed to analyse large amounts of data to derive new insights about individuals' behaviours that, in turn, result in new uses of that data or new decisions in relation to the individual. Some big data analytics services will pull data from machine-to-machine or IoT systems, which may have no practical means of communicating information to individuals.

Big data analytics services can make sure that any consumer-facing notices reference how and by whom the data may be used for analytics in such a way that individuals are able to understand easily.

Internal procedures can be implemented to review proposals to subject data to new analytics in order to understand whether it would go beyond the uses initially communicated to the individuals.

Dashboards can help individuals manage their personal data and make choices about how that personal data is processed.

Big data analytics services, together with others, can engage in longer-term education campaigns to foster a more meaningful understanding of the value exchange.

Example

The family whose smart home data is analysed in Scenarios 3A and 3B could be informed of exactly which data will be disclosed to the home security service. It may be disproportionate to notify the family each time a third party requests access to non-identifiable, aggregated data as in Scenario 3B. However, they could be informed when they start the smart home service that their data may be shared in this way with certain third parties, or they could be directed to a website that tells them about all the organisations that have requested access to the non-identifiable, aggregated data.

Privacy impact assessment, privacy-by-design

Through identifying new correlations across data sets, many big data analytics services hope to provide actionable insights that have a positive impact on society or individuals.

[2] Definitions vary, but generally speaking personal data is considered to be information relating to a living individual or from which an individual may be identified either from the information itself or when combined with other data that is likely to come into the possession of the organisation.

A key tool for recognising privacy impact on individuals is the data privacy impact assessment, which helps organisations to identify and mitigate privacy risks[3].

Developers of big data analytics services can consider adopting a 'privacy-by-design'[4] ethos or methodology by which privacy and security safeguards are considered and designed into products, services, processes or projects at each stage of the lifecycle from cradle to grave.

Example

In Scenario 2, assuming the system is designed and built in the anticipation of future emergencies, a data privacy impact assessment could reveal the sort of impact that data disclosures may have. Rather than disclose identifiable data to the WHO, the system can be designed to only send aggregated insights to the WHO. If the WHO needs to communicate directly with individuals whose life is at risk, the MNO can send a message to the affected population.

Accountability

As mentioned above, there are many different scenarios in which big data analytics services may be delivered.

Where one organisation makes their data or insights available to third parties, that organisation is in a position to set out clearly what they will be responsible for and what they expect of the third parties. For example, the organisation may want to keep a log of the API access events or ask third parties to report on their use of the data and/or insights.

Where several organisations are collaborating under a common framework or where they are pooling their data[5], a responsible approach might be to set out clearly in an agreement or protocol the ground rules for access, for example:

- What data will be made available to whom and in what form?
- Will the data be pseudonymised?
- Will there be access to any raw data?
- Who has access to the findings?
- Who can determine the design/purpose of the algorithms?
- How and to whom will insights from the analytics be disseminated?
- Who will control actions taken based on the insights?

In addition, the organisation sharing the data or insights can consider maintaining an audit log detailing requests for data and insights in order to be able to identify the third party making the request. The method of the requests could also be recorded, for example, whether the access was via an application or a user accessing a file store/ website.

It is also advisable that the organisation sharing their data or insights should have a process to deal with any suspect activity or independently reported misuse by any third party with which it is working.

Example

Particular care should be taken when accessing or making available data from internet-enabled devices where many parties may be involved in the value chain, as it may not be clear who is responsible. For example, in Scenario 3A, the manufacturer of the smart light bulb, the MNO that provides the smart home service and the home security service provider all need a clear understanding of who is responsible for limiting access to the data feeds.

Security access

As with other personal data processing activities, security is a key safeguard to protect people's privacy. The particular considerations for big data analytics services are that they often involve third party access to data or several parties operating together under one framework.

Big data analytics services can therefore improve security by limiting access to data or insights to authorised users only and by securing the data or insights in transit, in rest during the analytics phase, and in the release of reports or insights.

When setting time limits for data retention, big data analytics services can consider the sensitivity of the data and whether it is possible to continue beyond a certain time with pseudonymised/aggregated data only.

Example

An ongoing live service, for example to send intruder alerts to the home owner in Scenario 3A, may require identifiable data for only a very limited time, but thereafter may need to keep de-identified data for a number of years to understand trends and correlations. Other big data analytics services where several parties operate under one framework or as one platform may want to establish themselves as a longer-term resource, in which case other security measures such as pseudonymisation, access control or encryption will take on a greater significance.

[3] See ICO Code of Practice on conducting privacy impact assessments; Centre for Information Policy Leadership - A Risk-based Approach to Privacy: Improving Effectiveness in Practice

[4] See Office of the Information and Privacy Commissioner of Ontario – Privacy by Design;
Office of the Information and Privacy Commissioner of Ontario - Privacy Considerations at Each Stage of the Big Data Lifecycle;
Office of the Information and Privacy Commissioner of Ontario - Big Privacy: Bridging Big Data and the Personal Data Ecosystem through Privacy By Design

[5] See ICO Data Sharing Code of Practice; ICO data sharing checklists

Cross-border transfers of data

Cross-border transfers of personal data are currently regulated by a number of international, regional and national instruments and laws intended to protect individuals' privacy, the local economy or national security.

Transmitting personal data across national boundaries can sometimes trigger additional duties. Big data analytics services can comply with cross-border transfer requirements and enable their services if they implement contractual privacy safeguards or if they embrace accountability mechanisms such as the APEC Cross-Border Privacy Rules or the EU's Binding Corporate Rules which allow organisations to transfer personal data generally under certain conditions.[6]

Example

The arrangements in Scenario 2 are likely to involve some transfers of data across national boundaries, particularly if the WHO, major NGOs and MNOs all agree to run the system on a common platform in anticipation of future emergencies. The parties may want to consider entering into contractual terms that will protect the interests of future individuals whose data will be collected and use these terms to comply with data transfer restrictions.

Ethics

In addition to considering legal requirements, big data analytics services may also consider the overall fairness and the ethical dimension of what they, or the third parties accessing the data, are proposing to do.

Organisations can incorporate ethical decision-making models into their business processes to help build better services and foster an environment of trust.

Example

If the disclosure of aggregated data under Scenario 2 were to show the movement of particular ethnic groups, but not individuals, in the aftermath of an emergency, the MNO might consider the proposed big data analytics services and the circumstances in which such information may be released from an ethical rather than just a legal point of view.

[6] See The GSMA Mobile Policy Handbook position on Cross-Border Data Transfers

www.gsma.com/mobileprivacy

GSMA HEAD OFFICE
Floor 2
The Walbrook Building
25 Walbrook
London, EC4N 8AF
United Kingdom
Tel: +44 (0)20 7356 0600
Fax: +44 (0)20 7356 0601