Framework for Assessing Privacy of Internet Applications

James PH Coleman
Department of Computing
Edge Hill University,
Ormskirk, Lancashire, United Kingdom

***Abstract*—This paper presents a new framework for assessing and documenting the privacy risks associated with developing and managing internet applications. The Framework for Assessing Privacy of Internet Applications (FAPIA) provides a tool to aid the analysis of privacy risks and a structured means of analyzing the risks and documenting a control systems to ensure compliance with data protection and privacy legislation in a range of different countries.**

Keywords—privacy; privacy compliance; risk; data protection; privacy impact assessments; internet applications

I. INTRODUCTION

The "Internet of Things" [1], "Ubiquitous Computing" [2] and "Private area networks" [3] are similar concepts and relate to computing and network devices that are being created and installed to improve life and lifestyle. There are CCTV cameras in railway stations and along busy shopping streets, there are RFID tags to protect expensive clothing in high street shops as well as wearable health technology, eg devices to monitor heart beat with the ability to "send" this data to other devices and people.

Care and attention is being paid by developers to building user friendly interfaces, and secure encryption of data over networks, what is lacking is that far more subtle of concept – that of privacy. Internet users are concerned about the privacy of their personal information when it is online for it can lead to invasions of privacy [4], [5]. A study in 2004 [6] shows that users are particularly concerned about 3 aspects:

- the collection of personal information,
- the control of personal information and
- the awareness of users about actual privacy practices being used by system developers.

Given the wide range of applications that are involved, including: home and private networks, smart metering, monitoring for health and social care needs, observation by CCTV and other tools of public spaces, airports and train stations as well as the increasing use of the public Internet for phone calls and emails etc. with tools able to "*provide tailored services*" based on the content of emails etc. It is increasingly important that users and developers are aware of the privacy implications of their existing and any new systems or services they offer and experience tells us that a powerful tool for understanding this is the privacy Impact Assessment.

As the International Association For Impact Assessment (IAIA) [7] state - an "*... Impact assessment ... is the process of identifying the future consequences of a current or proposed action*".

In this article, a framework is proposed that will help developers with the understanding of the potential privacy implications of their system. It aims to aide developers and system managers to:

- ensure compliance with relevant privacy and data protection laws,
- manage risks to the perception of the system's owner and consumer confidence in the system (specifically related to privacy and data protection), and
- develop confidence in these new technologies

As privacy and data protection is applicable to a wide range of different applications and systems, it will provide a set of tools that developers and managers can use.

A Privacy Impact Assessment (PIA) is a document which "…seeks to… [explicitly].., in as much detail as is necessary…, [specify]…the essential components of any personal information system or any system that contains significant amounts of personal information…[under the headings]… description; data collection; disclosure and use of data; privacy standards and security measures…" [8] This will help developers to fully understand the privacy and data protection implications of the system they are developing, and to then be able to provide the necessary and relevant information that users, customers or the public want and need about the systems that will be used or are proposed. The PIA is a very generic concept that needs to be adapted to differing situations and scenarios, not just those relating to networked computerised environments. This article will focus on its applicability to networked systems, (where the Internet is involved) and will provide a framework for developers to frame their PIA analysis and develop better networked systems.

A. Definitions of a PIA

There are numerous definitions of a PIA and while the definitions are different, they agree that the PIA is an "assessment of any actual or potential effects that the activity or proposal may have on individual privacy and the ways in which any adverse effects may be mitigated" [9]

"PIA is an analysis of how information in identifiable form is collected, stored, protected, shared and managed. [to] ensure that system owners and developers have consciously incorporated privacy protection throughout the entire life cycle of a system" [10]

''a process whereby the potential impacts and implications of proposals that involve potential privacy-invasiveness are surfaced and examined'' [11]

B. Framework for assessing privacy

The purpose of this Framework is to assess the privacy risks involved in a computerised system, and having determined the risks to privacy, to then enable the system developers/managers to ensure those risks are properly managed. The precise systems or controls a developer uses to manage those risks will depend upon the legal framework in which they operate however this Framework will enable them to understand what is happening (from a privacy perspective) in their system.

By working through the Framework the developer/manager will be able to develop controls (either technological or procedural) to manage privacy within the system.

The Framework for Assessing Privacy for Internet applications (FAPIA) is a tool that is designed to simplify the PIA process for internet application developers. It will enable them to assess the privacy risks, assess the likelihood and provide a structured way of addressing these risks and providing a documented outcome, particularly if a legislative framework exists, in which the system will operate. In the words of [8] "*... a basic function of a ... privacy impact assessment is to ask probing, detailed questions of the proponents, builders and designers in order to promote comprehension. The role is in effect that of a devil's advocate*."

FAPIA has been developed in the UK and therefore has been heavily influenced by the needs of the European Convention on Human Rights [12], in particular Article 8 which states:

Article 8 – Right to respect for private and family life

1) Everyone has the right to respect for his private and family life, his home and his correspondence. [12 Art 8]

This is a limited right as it goes on to say in section 2 - that *"…there shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society…"* and goes on to define certain conditions when this interference can occur.

Related to this is the European Union (EU) Directive on Data Protection [13] which provides a set of 8 principles on the collection and processing of personal data. While this may seem to be European-centric, it must be remembered that many nations have privacy laws or statutory requirements surrounding the processing of data and privacy including the so-called "safe harbour" schemes of the US Department of Commerce [9] which provides 7 principles for data processing and storage that must be used in certain circumstances.

While traditionally privacy relates to proper persons – that is, actual identifiable people. As the Internet grows however, and as the number of institutions uses digital systems to manage their data and environment, it is important that privacy be considered in a wider context. Under traditional terminology, privacy relates to a person, in the modern world we need to consider privacy in relation to institutions and organisations. Businesses use digital and internet systems to store their data and they expect this information (perhaps about business costs, incomes etc) to be properly managed. Managed in terms of data loss but also in the form of being kept confidential. Whole industries are reliant on being able to control access to business data (which may not fall under the protection offered by privacy legislation). Therefore while legal privacy protection may not exist, privacy as a concept applies equally to business data as it does to personal data given contractual obligations that form part of many business contracts.

As the author is EU based, then FAPIA will ensure that ECHR/DFD [12, 13] can be implemented, it does mean that FAPIA is equally usable and valid in other legislative areas and the Author believes is valid for all existing jurisdictions. This article will explain how FAPIA is structured, and how it can be applied to different domains.

II. Framework for Assessing Privacy for Internet Applications

The first question that needs to be asked by a developer is whether there are any privacy implications. This relates to whether issues of privacy as a result of legal or contractual obligations exist. Figure 1 for a decision tree which is the first step in the FAPIA process.

An assessment completed using FAPIA will consider each of the elements of the Internet application (called the Key Elements in FAPIA) and will analyse these in relationship with the privacy targets (called Privacy Targets in FAPIA) and will identify what FAPIA calls the Key Outcomes.

FAPIA is a system that aims to help developers understand their own (ideally proposed) system, which will then allow the developers to build a system that has minimal privacy impacts, or at least where the impacts can be managed properly and easily. Often what happens is that Developers find that they have built a system that is difficult to control, or where expensive "*procedural*" systems or add-ons are needed to maintain the level and type of privacy and data protection that is needed and expected from the customers.

Having applied the FAPIA, Developers should then have an understanding of the implications of what they are proposing, and are able to build the relevant controls into the system. It is unfortunate that many developers do not consider the privacy impacts until the end of the design phase (and sometimes the end of the development phase) of a system when modifications can be both expensive and difficult.

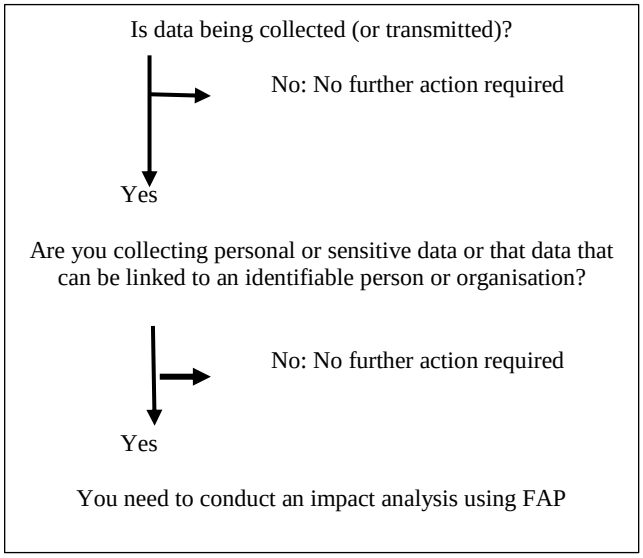

Fig. 1. Decision Tree for deciding if an assessment is required

A. Key Outcomes

The *Key Outcomes* of FAPIA are the following pieces of information about how the IA. The

- **Business need** being addressed by the proposed system?
- The **data being collected**, and the reason why that data is being collected?
- Description of the **Mechanisms** used in the IA which processes or stores (even temporarily) data?
- What **privacy impacts** may arise as a result of?
    - Algorithms used to process the data
    - Data storage and is that storage necessary for the operation of the system?
    - What is the justification for negative privacy impacts?
- **Procedures** and protocols used to manage access to data
    - What methods are proposed whereby those negative privacy impacts will be ameliorated?

In order to identify the Key Outcomes of an IA, it is first necessary to identify main actors/agents that are involved. These are called the Key Elements and it is the analysis of what is generally pieces of software and/or hardware that will result in improved understanding of the privacy impact of an IA.

B. Key Elements

The Key Elements identifies the components that Developers and Maintainers need to identify in their system. By doing so they are then able to develop and maintain the systems and controls that are needed to maintain privacy for not all risks can be addressed by technology alone.

The Key Elements are:

Data Subjects – The Data Subject is the person (or group of people) who is the subject of the collection of data. Note that this may not be the same as the intended data subject.

The Data Subject is the individual whose particular data is provided to the system and the person can be identified or distinguished from others in the system. (eg a person has brought a product that uses Radio-frequency identification (RFID) tags to identify products in a store, and the same RFID readers are used in a number of shops, so by linking the original sales transaction to the RFID ID then the person can be identified).

In identifying the data subjects, it is important to consider those who are not the intended subjects. So for example, in a CCTV system for a railway station, all people who move through the station may be data subjects if the images are recorded (for 14 days) in a crime-detection programme. Often the data subject is not the person about which data is being generated but is the person who receives the data. So for example, in an email system the data subjects are both the recipients and senders of the emails.

Data Elements – This is the data that is being collected and which may be processed in response to instructions and/or is recorded with the intention that it should be processed at a later date. The data is collected (or being sent) by the Network Device. In determining the data elements involved in the IA, it is important to remember that often other data is being collected as well as that which is intended.

WHAT IS PERSONAL DATA?

The data collected, processed and/or stored – It is important to realise that privacy is primarily concerned about data that is related to a person or a group of people or which can be linked to a real person or organisation. This means that if the system is collecting data on activities or events for which the data is not linkable to an identifiable person/organisation, then privacy considerations are generally not applicable. So for example, a RFID system does not need to consider privacy considerations when the tags are in the warehouse, and there is no link to a person. Once the link to a person can be made because of a sale, then privacy considerations now apply.

Network Sensor – A network sensors is a device to monitor conditions (eg CCTV, temperature, sound, pressure, etc). and to pass their data through the network to a main location – an Application. As networks may be bi-directional, also enabling control of sensor activity This means that the network sensor may be a movement sensor, biometric sensor or a keyboard or tablet computer.

Network Node – The network node includes devices that pass on data received from a sensor or pass on data over a network (which may include the public Internet but is quite often a private network or a combination). There may be multiple network nodes involved in data collection (or transmission).

Application – An application is a process or group of processes that involve the processing of the data collected about a data subject which includes potentially the linking of this data with other data sources. This can include the organisation, retrieval, disclosure of data or the linking of collected data with other data sets.

Back-end – The back-end is the area used for the storage of data about or involving a data subject. This includes the ability to store what is initially un-linked data but through later analysis linkages are created.

Data Controller – the Data Controller is a person or an organisation or a group who process or use the data about a data subject. This includes the situation where an organisation is processing (including storing) data on behalf of another organisation in the form of an out-sourcing arrangement.

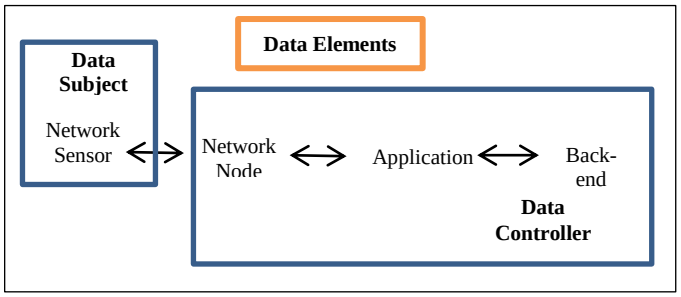

Fig. 2. FAPIA Key Elements

C. Privacy Targets

In order to develop a system that is compliant with privacy legislation (as implemented in different regions of the world), it is necessary that the developers fully understand their own system, who is involved in the system, and how it works. Once the Key Elements have been identified, FAPIA then tests these elements against privacy and data protection targets, called the *Privacy Targets*. These are characteristics that any system that manipulates personal or sensitive information needs to address. It should be borne in mind that different countries will have precise legal definitions for some of these characteristics but this Framework is structured in such a way that the information will be available for most legislative systems and that the developers will have the information necessary to provide reassurance for commercial companies and organisations about the privacy services that will be provided.

The Privacy targets identified as part of this Framework are based on the pargets identified as part of the EU Privacy Directive. [13].

These targets are grouped according to the 3 issues identified by [6] as being the main issues that network users are concerned about when considering privacy. The targets represent the concerns of users or customers and as such should represent a major driver for businesses who use data for whatever reason.

These targets are:

1) Collection of Private Information

Safeguarding quality of private data

Data avoidance and minimisation, purpose specification and limitation, quality of data and transparency are the key targets that need to be ensured.

Legitimacy of processing data

Legitimacy of processing personal and private data must be ensured either by basing data processing on consent, contract, legal obligation, etc.

2) Control of personal information

Compliance with the data subject's right to be informed

It must be ensured that the data subject is informed about the collection of his data in a timely manner.

Compliance with the data subject's right to object

Where there is a legal right to object (or *right-to-be forgotten*) to having data processed, then this is ensured. Transparency of automated decisions vis-à-vis individuals must be ensured especially. Even if there is no *right to object*, the Developers need to provide a mechanism for handling Data Subject's expression of views and how this is to be handled by the system. All existing privacy frameworks include some mechanism for appeals

Safeguarding confidentiality and security of processing

Preventing unauthorised access, logging of data processing, network and transport security and preventing accidental loss of data are the key targets that need to be ensured.

Allowing the data subject right of access to data, and to correct and erase data

It must be ensured that the data subject's wish to access, correct, erase and block his data is fulfilled in a timely manner. All existing privacy frameworks include some mechanism for appeals

3) Awareness of users about actual privacy practices

Compliance with notification requirements

Notification about data processing, prior compliance checking and documentation are the key targets that need to be ensured.

Compliance with data retention requirements

Retention of data should be for the minimum period of time consistent with the purpose of the retention or other legal requirements.

III. Outcomes and Future Work

As FAPIA is a process which generates a series of outcomes (see section II. A. above) this results normally in the production of a report. If the Framework was applied prior to the development of the system itself, then this report provides input into the design process for the Internet application.

Where other tools need to be addressed, then FAPIA provides a comprehensive data set that can be used. FAPIA however produces a set of Key Outcomes:

Overview specification of the Operation of the system – the business needs being addressed.

Specification of actual data collected – a comprehensive set of data elements being collected explaining how the data is collected and the purpose being addressed by the data.

Specification of actual data stored – a comprehensive set of data elements being stored explaining how the data is stored and the purpose being addressed by the data storage. This also needs to include how the data is being protected given the level of the risk involved. This also needs to include temporary storage of data (eg hours or days).

Algorithms used to process the data – this is important to know how the system is processing the data. As a consequence of data processing, some data which was collected anonymously may become identifiable, and hence become a privacy risk that needs to be managed.

Privacy Risks – what are the privacy-risks associated with the collection, processing and storage of the data. Where there are privacy risks then the justification for having negative privacy risks need to be provided, depending on the level of the risk-score involved.

Privacy Systems - Amongst the numerous different types of privacy-risks, privacy systems/processes will be used to manage the privacy requirements of all systems. There are a number of *common procedures and protocols* which all systems need to address:

- Procedures and protocols used to manage access to data and other Key Elements
- Procedures and protocols for the collection of data
- Procedures and protocols for informing the data subjects
- Procedures and protocols for managing data transfers
- Procedures and protocols for managing data loss
- Procedures and protocols for objections, appeals
- Procedures and protocols for providing data subject to access, correct and erase data (where needed or necessary)

Figure 3 illustrates a form of FAPIA Report for a sample CCTV system which consists of a video camera with images being stored on a PC. The FAPIA Report

Having completed FAPIA and produced these outputs, the system developers and managers have a very comprehensive understanding of how their system needs to be developed. The system managements tasks that need to be undertaken, and often can be built into the project are clearly identifiable within the FAPIA key Outputs.

FAPIA Report – CCTV Monitoring System

Key Elements:
Data Controller: Sample Organization Owner
Data Subjects:

- Customers entering Sample Organization's shops.
- Members of staff of Sample Organization who work in areas covered by CCTV system.
- Members of public who walk in front of sample Organization's shop.

Network Sensor: CCTV camera in shop
Network Node: PC where *CCTVSoftware* is located
Application: *CCTVSoftware* application
Back-end: *CCTVSoftware* and DVD storage

Key Outputs:

Overview specification of the Operation of the system The CCTVSoftware system consists of a Camera and related software which records all images the camera captures and stores these images for a 7 day rotational system. The camera captures images of activity within the shop-area including through the window and on the street front.

Specification of actual data collected
Images of members of public, customers and staff in the shop or outside front shop window.

Specification of actual data stored
Images of members of public, customers and staff in the shop or outside front window stored for 7 days on rotational cycle unless retained for a longer period for crime detection/prevention purposes

Algorithms used to process the data
The only algorithms used are in the capture and storage of data. Data used for crime detection/prevention purposes may be enhanced to provide better images for forensic purposes.

Privacy Risks

- Use of material for non-approved purposes
- Loss of data – possible because of:
  - Technical problems (eg data storage shortage)
  - Procedural problems (eg staff deleting files)

Privacy Systems - Procedures and protocols used to:
Manage access to data and other Key Elements

- Procedure for examining images in the event of suspected crimes

Procedures and protocols for informing the data subjects

- CCTV Notice – Notice informing the public of the use of CCTV for crime detection/prevention uses.

Procedures and protocols for the collection of data

- Protection of stored data in locked room

Procedures and protocols for managing data transfers

- Procedure for storage and copying of images in the event of suspected crimes

Fig. 3. Sample FAPIA Report

A. Other Outcomes

FAPIA can also be used to provide the information needed for a range of other reports, especially systems such as Privacy Impact Assessments as is required by a number of organisations as a result of government regulations. While the precise structure and content of the PIA will vary, often markedly, the comprehensiveness of the FAPIA system will mean that in most cases the information needed has been collected by the FAPIA process and is readily available.

The main outputs of the FAPIA process is an assessment of the impact on privacy of an internet application. This set of impacts could be used to inform part of the risk register for the system. The Risk Register (RR) is a risk management tool commonly used in risk management and compliance. It acts as a central repository for all risks identified by the organisation. Although risk registers are used extensively they are often criticised because they can lead to ritualistic decision-making [14], illusion of control [15], and the fallacy of misplaced concreteness [16]. Used correctly however they can be a useful tool in managing the risks associated with a project. This suggests that a similar tool would prove equally useful in managing the privacy concerns of applications

IV. Conclusion

The Framework for the Assessment of Privacy of Internet Applications (FAPIA) is a tool that if used, enables Internet system developers to build systems that respect the privacy of the people who are involved in the system while still achieving the business needs for which the system was developed and should do so without having to bolt-on systems after completion (or near completion). FAPIA focusses on Internet systems, and as such provides specific internet-linked prompts that developers can use.

References

[1] The Internet of Things Council, http://www.theinternetofthings.eu/, Accessed: 9/5/2015.

[2] Weiser, M., Brown, J.S.: The coming age of calm technology. (1996) http://www.ubiq.com/hypertext/weiser/acmfuture2endnote.htm. Accessed: 29/5/2015

[3] "Personal Area Networks", http://www.techopedia.com/definition/5079/personal-area-network-pan, Accessed: 9/5/2015.

[4] Laufer, R. S., & Wolfe, M, "Privacy as a concept and a social issue: A multidimensional developmental theory". Journal of Social Issues, 33, 1977, pp22–42

[5] Culnan, M. J., "Protecting Privacy Online: Is Self-Regulation Working?," Journal of Public Policy and Marketing (19:1), 2000, pp. 20-26

[6] Malhotra N., Kim S., Agarwal J. "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model", Information Systems Research, Vol 15, Issue 4, 2004, pp336 – 355.7 6-0a International Association For Impact Assessment

[7] IAIA. International Association For Impact Assessment, http://www.iaia.org/. Accessed: 9/5/2015

[8] 8# 6a Flaherty, D., "Privacy impact assessments: an essential tool for data protection" [2000] PrivLawPRpr 45; (2000) 7(5) Privacy Law and Policy Reporter 85; http://www.austlii.edu.au/au/journals/PLPR/2000/45.html Accessed: 9/5/2015

[9] 9# 6a1 Stewart, B., "Privacy impact assessments" [1996] PrivLawPRpr 39; (1996) 3(4) Privacy Law & Policy Reporter 61. http://www.austlii.edu.au/au/journals/PLPR/1996/39.html. Accessed: 9/5/2015

[10] 10# 6a2 Clarke R., "Privacy impact assessment: Its origins and development" computer law & security review 25, 2009, pp123–135

[11] 11# 6a3 Clarke R. "Privacy impact assessments", 1998, Xamax Consultancy Pty Ltd. Available at: http://www.xamax.com.au/DV/PIA.html. Accessed: 9/5/2015

[12] "European Convention on Human Rights", Council of Europe, http://www.echr.coe.int/Documents/Convention_ENG.pdf, Accessed: 9/5/2015

[13] "Directive 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 such data," Official Journal of the European Communities, 1995, L 281/31, http://ec.europa.eu/justice_home/fsj/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf. Accessed: 9/5/2015

[14] Drummond, H. "MIS and illusions of control: an analysis of the risks of risk management. Journal of Information Technology, 2011, 26, 259–267. doi:10.1057/jit.2011.9

[15] Lyytinen, K. "MIS: the urge to control and the control of illusions – towards a dialectic". Journal of Information Technology, 2011, 26, 268-270 doi:10.1057/jit.2011.12

[16] Budzier, A. "The risk of risk registers – managing risk is managing discourse not tools". Journal of Information Technology, 2011, 26, 274-276 doi:10.1057/jit.2011.13

Appendices

The Data Protection Directive [13] of the EU provides the following set of principles for the processing of data and related principles of privacy. These principles underpin this Framework.

Data protection principles

1) Personal data shall be processed fairly and lawfully.

2) Personal data shall be obtained only for specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.

3) Personal data shall be adequate, relevant and not excessive in relation to the purposes for which they are processed.

4) Personal data shall be accurate and, where necessary, kept up to date.

5) Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.

6) Personal data shall be processed in accordance with the rights of data subjects.

7) Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.

8) Personal data shall not be transferred to a country or territory outside the EU unless that country or territory ensures an adequate level of protection