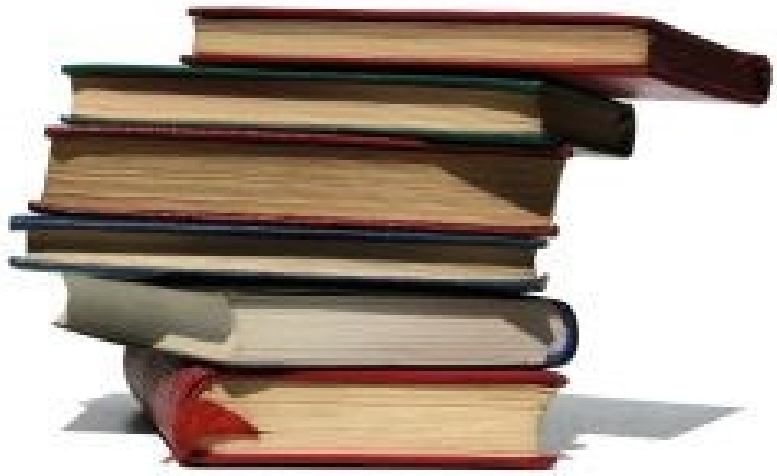
ANEC POCKET GUIDE

Using Consumer Data

Consumer Representatives Guide on Privacy



Raising standards for consumers

Disclaimer: this pocket guide is intended for ANEC membership and ANEC representatives in particular.

1. Objective of the guidance paper

The use of Information and Communication Technology (ICT) in consumers' lives is now very widespread and still growing. ICT is no longer confined to laptops and desk top computers in the home with a growing number of smartphone apps like health and fitness, as well as fridges, cars and transport systems, payment cards, televisions and more. When connected to the Internet, these are frequently used to collect data in real time and provide data concerning the consumers/citizens who own or use them. Service providers, retailers and others can make use of this data for marketing, research and many other purposes. The analysis of the data collected from consumer/citizens brings with it real data privacy concerns.

The intent of this paper is to assist consumer representatives address privacy issues related to personal data analysis on technical committees dealing with "smart" products and services, where ICT technology has become a fundamental part of the capabilities of the products.

The aim is to ensure that good data analysis practice standards support the privacy principle that **"Personal data analysis processes should be designed to protect individuals' privacy"**

Where personal data is processed in a manner that it is analysed to inform or influence decisions then precautions are needed to protect privacy. This principle impinges on:

- Governance
- Identifiability
- Creation of large data sets that collectively represent much more sensitive personal data than individual data items do by themselves
- Accuracy of analysis, especially false positives and false negatives which impact individuals
- Use of personal data analysis for personal risk management within health, finance and many other types of service
- "Big data" applications

2. Data Analysis and Privacy

2.1 What is data?

For the purposes of this document "data" is used in its scientific context and defined in dictionaries as

"Numerical or other information represented in a form suitable for processing by computer."

Note: Big Data is dictionary defined as "data sets, typically consisting of billions or trillions of records that are so vast and complex that they require new and powerful computational resources to process". The privacy issues are the same as for 'ordinary' data processing and analysis but for larger data sets with a challenge of scale.

2.2. What is data analysis?

A good definition of data analysis has been provided on wikipedia:

"Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making." http://en.wikipedia.org/wiki/Data-analysis

2.3. What is personal data?

A good definition of personal data has been provided by the International Standards Organisation in ISO/IEC 29100.

Personally Identifiable Information "PII is any information that (a) can be used to identify the PII principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal"

Note: Fuller guidance on the nature of personal data is provided in the Consumer Representatives Guide on Domestic Privacy.

3. Key Privacy issues associated with data analysis

Organisations analyse huge volumes of personal data in ways that impact on daily life. Such analysis may be undertaken for many reasons and much benefit to individuals and society can be generated.

Privacy issues associated with data analysis come about when personal information is collected from individuals and then analysed, sometimes shared and distributed, allowing organisations and others to draw conclusions about individuals and take decisions about them. Consumers and Citizens will benefit from better privacy protection from standards that address the issues that arise from data analysis, see sections 3.1 to 3.10 below.

3.1 Balancing the right to privacy with the public interest

3.1.1 Governance

Benefits to society and individuals can be derived from data analysis where PII may be used and where not to use PII would hinder the usefulness of the analysis to the individual and the public interest. There may often be a tension between the benefits derived from such PII analysis and an individual's privacy. In such cases good governance is a key element to ensuring that the rights to individual privacy are balanced with public interest issues.

In these situations good governance to protect an individual's privacy should be transparent, proportional and fair while also including the public interest. This is particularly true when data analysis incorporates personal identifiers.

Note 1: for an example of more detailed principles of governance, transparency and fairness principles applicable in the case of National Government surveillance see "International Principles on the Application of Human Rights to Communications Surveillance - FINAL VERSION 10 JULY 2013"

Note 2: To illustrate those aspects of data analysis privacy and governance with respect to the right to be forgotten, a commentary is given in Annex 1 on the (current at time of publishing) requests

and judgements on the right to be forgotten from Internet search engines like Google.

3.1.2 Engaging stakeholders

Where there are significant privacy effects on consumers and impacting the public interest then good practice governance processes should include stakeholder engagement where :

- There are clear criteria for being a stakeholder
- The criteria are available and easily accessible in the pubic domain
- Stakeholder identification and selection uses only these criteria
- Application to become a formally recognised Stakeholder is open to all
- Any organisation or individual who meets the criteria set should become a stakeholder
- When stakeholder applicants are not accepted they should be informed of the reasons why
- There should be processes to deal with appeals with respect to stakeholder engagement that are transparent to the public

3.2 Anonymization: Ensuring that when PII is used in data analysis, individuals' personal identifiers are not used when individual identity is not required for the analysis (see also 3.6).

Any data analysis process not requiring individuals to be identified should include steps to ensure that input data is anonymous.

3.3 Re-identification: Being able to identify individuals by combining different sets of data relating to them.

There are risks that apparently anonymous data sets when analysed together may create a high degree of individual identifiability that is not the purpose of the analysis.

For example an analysis of anonymous data sets that include age, sex and full UK post codes (which could be between 15 and 80

households) may effectively identify some individuals down to one or two people.

To address this issue standards should include post data processing checks for the degree of identifiability created by the analysis. Where necessary, the data should be re-anonymised ("k-anonymity").

Each area of processing governance will need to establish unacceptable degrees of re-identifiability within each of their own contexts. This is a key factor that should be made transparent by the governance process. For example being one in 25 people is different from being one in 25,000. See also section 3.9.

Where the aggregate analysis is dynamic, and individual input items may be changed, then the degree of individual identification needs to be re-checked at suitable intervals.

3.4. Profiling: Building up large personal profiles

Data analysis can be used to bring together more and more data about individuals increasing the amount of personal information in a profile. While individual data elements of the profile, like the breakfast cereals usually purchased, may be not very privacy-sensitive items, the combination of large quantities of personal data can reveal a great deal about individual's day to day life to a degree that is seen as intrusive by many.

When profile building is inherent in the data analysis then the associated processes should check for any extra privacy-sensitivity of the enhanced profile and individuals should be notified if that sensitivity increases significantly.

3.5 Data fitness for purpose.

3.5.1 When data is not fit for purpose, then analysis can often lead to inaccurate results affecting individuals, especially when decisions affecting these individuals are taken as a result of the analysis.

Examples include:

- Diagnostic data interpretation tables input for assisted diagnosis analysis becoming out of date when used to analyse person health data

- Using data for analysing where there are children in heavily indebted households, should the data definitions of the data employed not embrace all segments of severe household indebtedness.

When data analysis is a key part of a standard, then good practice requirements should ensure that information is recorded about each data set to enable fitness for purpose checks and that the record is associated with the data set.

The data set records should include

- what the data set includes
- why it was collected
- who was responsible for the collection,
- data collection methodologies,
- analysis methods used for derived data in the sets (see also 3.5.3),
- any personally identifiable information collected
- privacy risks and controls

A good example of this approach is the "USAID Open Data Privacy Analysis Template."
A Mandatory Reference for ADS Chapter 508
New Reference: 03/07/2013
Responsible Office: M/CIO/IA
File Name: 508mah_030714

3.5.2 If there is a significant change in purpose between why the data was collected originally and how it is analysed subsequently, then the output data for some individuals can provide "false positives" such as identifying someone as a credit risk when they are not, and "false negatives" for example not identifying some people for social care when care is needed.

When PII is used for data analysis other than that for which it was originally collected then governance standards should explicitly include how to deal with 'false positives" and "false negatives".

3.5.3. Where the PII analysis methods used are significant to decisions about consumers, for example for the receipt of benefits, or the granting of mortgage applications then good practice standards should ensure that such processing algorithms are consumer transparent ideally being published openly in the public domain or via other means such as being available on consumer request or as part of good governance to stakeholder groups including consumer movement representatives.

3.6 Existing customer or client data analytics

3.6.1 Normally the analysis of PII within the service delivered to an individual for that individual's benefit is perfectly reasonable for that purpose.

Examples of analysis of PII for an existing customer or client include:

- medical diagnosis,
- product recall,
- individual pricing offers to existing customers
- assessing an individual's credit risk
- locating restaurants near to an individual (mobile phone 'app')

3.6.2. If data is shared in order to provide an existing client with service, i.e when data is obtained by a sharing agreement or from open data sources to enhance customer / client service, then it is important to check whether a new processing purpose is being undertaken for the shared data as compared to the original data collection purpose consented to for that shared data. If a new purpose is involved for the shared data, then standards should ensure that the individual provides consent, or not, to the new processing purpose for that data.

3.6.3. Where the analysis undertaken results in significant decisions being taken about the individual then the algorithms used should be available at a minimum for stakeholder scrutiny. Confidentiality requirements may apply to stakeholders about the specifics of such algorithmic processing when they relate to commercial services,

however confidentiality requirements should allow non specific disclosure of concerns about the nature of the algorithms used.

3.6.4. National security processing should be included in 3.6.3 requirements unless such requirements are explicitly excluded by national laws.

3.7 Analysis of PII from open data

If open data is utilised in the data analysis processing that is used to deliver service to an existing customer then good practice standards should ensure that mechanisms are provided to pass to the source of the open data:

- any removal of consent associated with the personal data
- corrections to inaccuracies in the personal data that is available from open sources
- requests to be forgotten

Note: privacy data sharing transparency and traceability requirements that would underpin such notification mechanisms from the data subject to the open data provider or data sharer are dealt with in the Consumer Representatives Guidance paper on data sharing transparency.

3.8. Data analytics to identify or target an individual

3.8.1. Data analysis may be undertaken on data collected from individuals, and/or from other sources, in order to identify individuals who meet assigned criteria or characteristics, and then to take action with respect to those individuals.

Examples of 'targeting' purposes include :

- marketing using analysis of web browsing behaviour for individual targeting of adverts
- police and national security
- disaster recovery management e.g. finding the vulnerable people to be rescued

Good practice governance requirements (see 3.1) should be included in standards addressing analysis that targets individuals.

3.8.2. If the targeting purpose is to solicit commercial or financial benefit, or to be the target of campaigning activity of any type, then privacy control by the individual, such as opt out and spam controls, should be incorporated into :

- the technology and methods used to acquire input data
- any action arising from the data analytics processing such as sending promotional material

3.9. Data analytics to identify groups of people

3.9.1 This category of data analysis is undertaken to identify clusters of individuals with common characteristics where subsequent action does not target individuals.

Examples include:

- analysis for new product development
- marketing promotions (where action is aimed at the group not individuals)
- service performance analysis
- government policy development
- provision of environmental warnings

The privacy issues of items 3.1 to 3.5 should be kept in mind when the analysis purpose is to identify groups of people.

3.9.2. If communication is intended to the target group then non individually targeting/identifying methods such as TV broadcast, web site access, or newspapers should be used. If communication methods are considered that require individual addressability then the requirements identified in section 3.8 apply.

3.10. Data analytics for systems

3.10.1. When data analysis is undertaken to manage large scale systems such as:

- traffic management,
- police resources scheduling,
- bank ICT infrastructure performance analysis,
- energy provision management,
- placement and use of new buildings (commercial and public)

The privacy issues of items 3.2. to 3.5 should be kept in mind when the analysis purpose is for system management.

- end of main document -

Prepared by Peter Eisenegger

ANEC ICT Working Group

Annex 1 – An initial analysis of the Right to be Forgotten and Internet Search Services using open data where PII analytics is undertaken on that data.

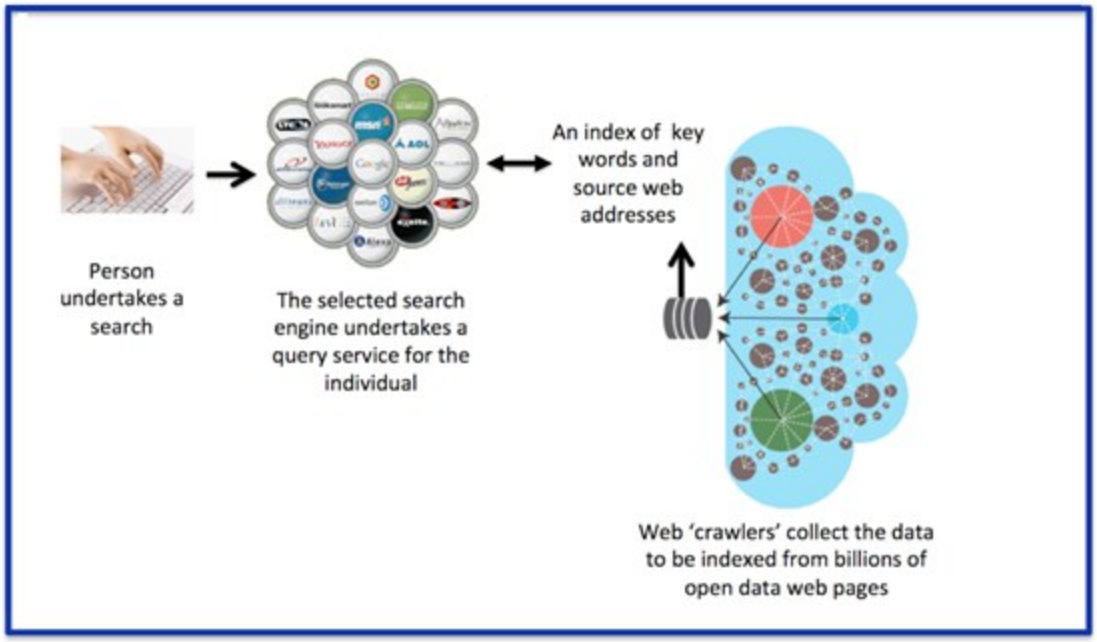
A1 - 1 A simple description of Internet searching

A user inputs a set of characters for the search to analyse and operate on. This is referred to as the search string.

The purpose of the search is not explicit other than to find information on the Internet that makes use of the words and characters input.

The search service, of which there are several like Google and Bing, then accesses indexes containing the words and characters found on web pages and the web addresses to web pages containing those words/characters. On a search the matching indexed items are returned to the individual who is undertaking the search, as web links to the actual contents web pages.

Figure A1 - 1 Simple overview of Internet Search Services



The contents of the indexes are created by software that automatically trawls round web sites, examining the contents of open web pages and other information associated with those pages

such as key words for searches. The words/characters found and the page's web address, are returned to the search engine index.

A1 – 2 The right to be forgotten

If the search string makes use of a personal identifier data type, especially someone's name, then normally the analysis undertaken by the search service will return PII and the right to be forgotten may be invoked by an individual who is the PII "Principal".

A1 – 2.1 The Right to be Forgotten - Open data web sites

Good practice standards for privacy of open data containing PII should ensure that the accuracy, legitimacy, suitability for use by others and governance of that PII rests primarily with the organisation who makes that information available publicly as open data on the Internet. This good practice should include means, provided on the originating web site, to allow individuals to request items of personal information to be forgotten and for good governance of that request to apply (see 3.1).

A1 – 2.2 The Right to be Forgotten - Search services

Through the analysis of PII a search service allows an individual to find likely sources of open data PII relating to themselves. Some of these information sources may be such that an individual may wish to be forgotten. Good practice standards for search services should ensure that the first point of call in a right to be forgotten case should be the organization responsible for publishing that information via their web pages.

However search services do process PII in these circumstances and can legally receive requests to be forgotten for the relevant PII they have in their search indexes.

Normally removal of such information from a search service index would be undesirable as it removes the ability of the individual to check whether the source web pages (which are not the responsibility of the search engine in terms of content) have removed the PII concerned.

There is the further consideration that the open web pages searched may contain a mixture of information, some of it not applicable to the individual concerned, while being in the public interest, and only some of the information on the web page may be relevant to the right to be forgotten request. Examples could be newspaper and media firms' web sites and individuals' blogs. When the reference to a whole page is removed by the search service, as that is how the technology works at the whole page level, then there may be cases where valuable public information is "forgotten" along with that element of PII that should be forgotten.

In practical terms this is why it is preferable and good practice to approach the web site owners first, as they have editorial control over the content of the pages while the search engines can only forget whole pages.

Given the public interest value of open data search capabilities, the governance good practice applied by search services to "right to be forgotten" requests should examine whether the individual has approached the organisation responsible for the source web page and requested the relevant PII to be forgotten. If the individual has done so and the organisation responsible for the source pages has not shown why the public interest is better served by maintaining the PII availability then the search service's good governance practice should support the individual in removing its search access to the web pages concerned.

If a search service none the less does receive "right to be forgotten" requests when the source web site organization cannot be approached by the individual with the right to be forgotten request, then the search service provider should apply good governance processes as described in 3.1 balancing the individual's privacy and right to be forgotten with the public interest.

Overall the "Right to be Forgotten" is a good example, like product safety with the "New Approach", where the law is robust while the implementation of the law can be greatly assisted by a series of product and technology standards to underpin that law.

ANEC in Brief

ANEC is the European consumer voice in standardisation, representing and defending consumer interests in the development of technical standards, in the application of conformity assessment schemes to standards, and in the creation or revision of legislation on products and services. ANEC receives funding from the European Commission and the EFTA Secretariat.

ANEC, the European Association for the Co-ordination of Consumer Representation in Standardisation

Avenue de Tervueren 32, box 27 – 1040 Brussels – +32 (0)2 743 24 70
anec@anec.eu - www.anec.eu

http://twitter.com/#!/anectweet

http://companies.to/anec