European Commission – Eurostat/G6

Contract No. 50721.2013.002-2013.169

'Analysis of methodologies for using the Internet for the collection of information society and other statistics'

D2. Results of the feasibility analysis

March 2014

Document Service Data

Type of Document	D2. Results of the feasibility analysis		
Version:	3	**Status:**	Draft
Created by:	Lefteris Angelis, Dimitris Kalogeras, Michalis Petrakos, Thanasis Priftis, Vasilis Sotiropoulos, Photis Stavropoulos, Michalis Vafopoulos	**Date:**	20/3/2014
Distribution:	European Commission – Eurostat/G4, Agilis S.A.		
Contract Full Title:	Analysis of methodologies for using the Internet for the collection of information society and other statistics		
Service contract number:	50721.2013.003-2013.169		

Document Change Record

Version	Date	Change
1	6/12/2013	Initial release
2	31/12/2013	Revised version
3	20/3/2014	Revised version based on Eurostat's comments received on 15/1/2014

Contact Information

Agilis S.A.
Statistics and Informatics
Acadimias 98 - 100 – Athens - 106 77 GR
Tel.: +30 2111003310-19
Fax: +30 2111003315
Email: contact@agilis-sa.gr
Web: www.agilis-sa.gr

TABLE OF CONTENTS

1. Introduction 4

2. Assessment of technical feasibility ..... Error! Bookmark not defined.

2.1. Introduction ..... Error! Bookmark not defined.

2.2. Network-centric methods ..... Error! Bookmark not defined.

2.3. Web site-centric methods ..... Error! Bookmark not defined.

2.4. User-centric methods ..... Error! Bookmark not defined.

3. Feasibility within the conditions of the ESS ..... Error! Bookmark not defined.

4. Methodological approach ..... Error! Bookmark not defined.

4.1. Production of statistics on the characteristics of business web sites ..... Error! Bookmark not defined.

4.1.1. Relevance ..... **Error! Bookmark not defined.**

4.1.2. Accuracy ..... **Error! Bookmark not defined.**

4.1.3. Coherence and comparability ..... **Error! Bookmark not defined.**

4.1.4. Clarity ..... **Error! Bookmark not defined.**

4.1.5. Timeliness ..... **Error! Bookmark not defined.**

4.1.6. Conclusions about the statistics on the characteristics of business web sites ..... **Error! Bookmark not defined.**

4.2. Production of statistics on the use of Internet by individuals ..... Error! Bookmark not defined.

4.2.1. Relevance ..... **Error! Bookmark not defined.**

4.2.2. Accuracy ..... **Error! Bookmark not defined.**

4.2.3. Coherence and comparability ..... **Error! Bookmark not defined.**

4.2.4. Clarity ..... **Error! Bookmark not defined.**

4.2.5. Timeliness ..... **Error! Bookmark not defined.**

4.2.6. Conclusions about the statistics on the use of Internet by individuals **Error! Bookmark not defined.**

5. Cost-benefit balance ..... Error! Bookmark not defined.

5.1. Web site-centric methods ..... Error! Bookmark not defined.

5.1.1. The site search market ..... **Error! Bookmark not defined.**

5.1.2. Costs ..... **Error! Bookmark not defined.**

5.1.3. Benefits and conclusion ..... **Error! Bookmark not defined.**

5.1.4. To the future ..... **Error! Bookmark not defined.**

5.2. User-centric methods ..... Error! Bookmark not defined.

5.2.1. Costs ..... **Error! Bookmark not defined.**

5.2.2. Benefits and conclusion ..... **Error! Bookmark not defined.**

6. Legal feasibility 4

6.1. Introduction 4

6.2. Legal compatibility analysis 5
6.2.1. Data protection terms and conditions 5
6.2.2. Course of action for NSIs 8
6.3. Data protection legal framework 9
6.4. The *sui generis* Database Right 23
6.5. Conclusion 26
7. Socio-political acceptance 27
8. Conclusions 32
9. References 34
10. Annex 35
10.1. Appendix 1 - Synonym XML definition 35
10.2. Appendix 2 38
10.3. Appendix 3 – Topics for discussion with the NSIs for the assessment of feasibility in the ESS 39

1. Introduction

The first deliverable of project 'Internet as a Data Source', namely deliverable D1 'Definition of Internet data-based indicators', proposed a number of Information Society-related statistical indicators on a) the use of Internet by individuals and b) on the characteristics of the web sites of enterprises. The aim of the present report is to examine whether the proposed indicators and methods for their compilation are feasible from the methodological and the practical point of view.

The feasibility analysis consists of the following elements:

- Technical feasibility (chapter **Error! Reference source not found.**)
- Feasibility within the conditions of the European Statistical System (ESS – chapter **Error! Reference source not found.**)
- Methodological feasibility (chapter **Error! Reference source not found.**)
- Cost-benefit balance (chapter **Error! Reference source not found.**)
- Legal feasibility (chapter 6)
- Assessment of the socio-political acceptance (chapter 7)

It must be noted that each aspect of feasibility is examined in isolation from the others. For example, when assessing the methodological feasibility of the methods, no concern is raised about their legal implications. Cross-references to the different chapters of the report are given when appropriate. Moreover, there are references to two additional deliverables of the project, deliverable D3 which presents the results of two pilot studies and deliverable D5 which discusses the evaluation of the potential of existing data sources to be used as input for official statistics.

The present report closes with the presentation of conclusions in chapter 8.

……..

6. Legal feasibility

6.1. Introduction

The aim of this assessment is to examine whether the automatic data collection methods examined by the project are feasible form the legal point of view.

The issue of collecting and aggregating statistical data has legal implications that relate both to Data Protection and Privacy regulations, and to areas of Intellectual Property Rights and particularly the sui generis Database right in the EU context.

We start by analyzing whether the methods of statistical data collecting and aggregating proposed are compatible with the existing legal framework (section 6.2). The analysis has been based on the exploration of the EU data protection legal framework concerning the processing of statistical data (section 6.3) and of the provisions concerning the sui generis Database right in the EU context (section 6.4).

6.2. Legal compatibility analysis

The object of the present legal analysis is a set of methods under which the Internet shall be used as a data source suitable for statistical purposes and relevant research. More precisely, the examination of the legal feasibility concerns a project that involves:

(a) the installation of a software mechanism in several types of personal computing devices (i.e. desktop computers, tablets, smartphones etc.) with the aim of collecting information on the user's online activities on the Internet, such as duration of Internet usage, hours per day, days per week of Internet usage, visits on web pages etc.
(b) use of a "crawler"-type software to collect and analyse content of corporate web sites, such as the kind of facilities and several categories of information, such as open vacancies for employment, that the site provides to end users.

Overall conclusion: In both cases the user and the private entity (corporation, enterprise etc.) must give their explicit consent for the data collection and processing. If this is received and moreover the sample members have been informed about the data that will be collected and the uses to which they will be subjected, the electronic collection does not differ, from the legal point of view, from the collection of similar data with questionnaires.

The legal assessment will focus on the stages of: data creation, data aggregation or collection stage, enrichment stage and dissemination stage. In each of the stages the aim is to identify the degree to which:

- property rights are created and how their transfer is effected
- if personal data are involved, who conducts their processing, for how long and how they are to be used.

6.2.1. Data protection terms and conditions

The data protection legal framework recognizes the consent of the data subject generally as an appropriate legal basis for the collection and processing of personal data. Nevertheless, there are two crucial factors that should be taken into account in order to ensure that the data subject' consent is an adequate condition for all four stages of the methodology in hand. The first factor refers to the circumstances the data subject opted in and the content of his/her consent. The second factor refers to the cases of data collection and processing that even the proper consent forms only one part of the overall procedure for the lawfulness of the project.

1. The adequate consent

According to the European data protection legal framework, the data subject's consent' is defined as "*any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed*"[1]. The relevant provisions on the lawfulness

1 Article 2 (h) of the Data Protection Directive. Article 2 (g) of the Data Protection Framework Decision in the Framework of the Police and Judicial Cooperation in Criminal Matters. Article 2 (f) of the e-Privacy Directive. Article 2 (h) of the Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December

of data collection and processing are referring to the existence of the "*unambiguous*" consent. For consent to be unambiguous, the procedure to seek and to give consent must leave *no doubt* as to the data subject's intention to deliver consent. In other words, the indication by which the data subject signifies his agreement must leave no room for ambiguity regarding his/her intent. If there is a reasonable doubt about the individual's intention, there is ambiguity.

There are in principle no limits as to the form consent can take. However, for consent to be valid, in accordance with the Directive, it should be an indication. Even if it can be "any" form of indication, it should be clear what exactly can fall within the definition of an indication. The form of the indication (i.e. the way in which the wish is signified) is not defined in the EU Data Protection Framework. For flexibility reasons, "written" consent has been kept out of the final text. It should be stressed that the Directive includes "any" indication of a wish. This opens the possibility of a wide understanding of the scope of such an indication. The minimum expression of an indication could be any kind of signal, sufficiently clear to be capable of indicating a data subject's wishes, and to be understandable by the data controller. The words "indication" and "signifying" point in the direction of an action indeed being needed (as opposed to a situation where consent could be inferred from a lack of action)[2].

More specifically, in the field of personal data collection and processing for statistical purposes the data subject's "informed" consent requires3 that the persons questioned shall be informed of the following elements:

(a) the compulsory or optional nature of the response and the legal basis, if any, of the collection,
(b) the purpose or purposes of the collection and processing
(c) the name and position of the person or body in charge of the collection and/or processing,
(d) the fact that the data will be kept confidential and used exclusively for statistical purposes,
(e) the possibility of obtaining further information on request.

At their request and/or according to the ways and means defined by domestic law, data subjects shall also be informed of the following:

(f) the way in which consent can be refused or withdrawn, in the case of optional surveys and, in the case of compulsory surveys, the possible sanctions this would entail;
(g) where applicable, the conditions of the exercise of the rights of access and rectification,
(h) the categories of persons or bodies to whom the personal data may be communicated;
(i) the guarantees to ensure the confidentiality and the protection of personal data;
(j) the categories of data collected and processed.

When the data subjects are not directly questioned, they shall be informed of the existence of the collection unless this is manifestly unreasonable or impracticable. They shall be able to inform themselves appropriately of the elements listed above. The persons questioned shall be informed at the latest at the time of collection. Under the title "Secondary collection", the Chapter reads that cases of processing or communication for statistical purposes of personal data collected for non-statistical purposes shall receive suitable publicity. The data subjects shall be able to obtain in a suitable way all abovementioned information, unless:

(a) this is impossible or involves a disproportionate effort; or unless

2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data Official Journal L 008 , 12/01/2001 P. 0001 - 0022

2 Article 29 Working Party Opinion 15/2011 on the definition of consent, p. 11. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp187_en.pdf

3 According to Chapter 5 of the Appendix to Council of Europe's Recommendation No. R (97) 18 concerning the protection of personal data collected and processed for statistical purposes

(b) the processing or communication of the data for statistical purposes is expressly provided for under domestic law.

The data subject shall be able to withdraw his or her consent for a single survey, as long as, identification data have not been separated from other data collected, or to suspend at any time and without retroactive effect his or her co-operation in a survey which extends over a period of time. Refusal to reply shall not be penalized unless domestic law provides for sanctions[4].

Personal data processed for a given statistical purpose may be communicated for other statistical purposes as long as these are specified and of limited duration. Communication in accordance with this principle shall be the subject of a written document setting out the rights and obligation of the parties, unless safeguards are provided for by domestic law. The controller shall in particular:

(a) stipulate that the third party may communicate these data only with the express agreement of the said controller;
(b) stipulate that the third party take appropriate security measures and
(c) ensure that any publication of statistical results obtained by this party will anonymize the data unless dissemination or publication manifestly presents no risk of infringing privacy rights.

Sensitive data communication is allowed where provided for by the law, or where the data subjects have given their explicit consent and provided domestic law does not prohibit the giving of the consent.

Consent can only be valid if the data subject is able to exercise a real choice, and there is no risk of deception, intimidation, coercion or significant negative consequences if he/she does not consent. If the consequences of consenting undermine individuals' freedom of choice, consent would not be free. An example of the above is provided by the case where the data subject is under the influence of the data controller, such as an employment relationship. In this example, although not necessarily always, the data subject can be in a situation of dependence on the data controller - due to the nature of the relationship or to special circumstances - and might fear that he could be treated differently if he does not consent to the data processing.

To be valid, consent must be specific. In other words, blanket consent *without* specifying the exact purpose of the processing is not acceptable. To be specific, consent must be intelligible: it should refer clearly and precisely to the scope and the consequences of the data processing. It cannot apply to an open-ended set of processing activities. This means in other words that the context in which consent applies is limited[5].

Consent must be given in relation to the different aspects of the processing, clearly identified. It includes notably which data are processed and for which purposes. This understanding should be based on the reasonable expectations of the parties. "Specific consent" is therefore intrinsically linked to the fact that consent must be informed. There is a requirement of granularity of the consent with regard to the different elements that constitute the data processing: it cannot be held to cover "all the legitimate purposes" followed by the data controller. Consent should refer to the processing that is reasonable and necessary in relation to the purpose. It should be sufficient in principle for data controllers to obtain consent only once for different operations if they fall within the reasonable expectations of the data subject.

According to a preliminary ruling regarding Article 12(2) of the e-Privacy Directive[6], concerning the need for renewed consent of subscribers who had already consented to have their personal data published in

4 According to Chapter 6 of the Appendix to Recommendation No. R (97) 18

5 Article 29 Working Party Opinion 15/2011 on the definition of consent, p. 17.

6 Judgment of the Court of 5 May 2011, Deutsche Telekom AG (Case C-543/09). This case started with the referral made by the German Federal Administrative Court regarding telecom directories and in particular the

one directory, to have their personal data transferred to be published by other directory services the EU Court of Justice held that where the subscriber has been correctly informed of the possibility that his personal data may be passed to a third-party undertaking and s/he has already consented to the publication of those data in such a directory, renewed consent is not needed for the transfer of those same data, *if it is guaranteed that the data in question will not be used for purposes other than those for which the data were collected with a view to their first publication (paragraph 65).*

2. Where the consent is not enough

The Data Protection Directive foresees in Article 8.2(a) that in some cases, to be determined by Member States, the prohibition of the processing of special categories of personal data may not be lifted by the consent of the data subject. This is the case when the operation contains "special categories" of personal data (revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life.).

Collecting data of an end user's visits on the Internet may also contain collection and processing of these sensitive data categories. In several Member States, the appropriate safeguards that allow the collection and processing of sensitive data are formulated as a prior permission issued by the independent Data Protection Authority[7]. Article 8 of the Data Protection Directive obliges the data controller to comply with national law procedures, in the case of sensitive data collecting.

In conclusion, the mere consent will not be the appropriate legal ground for collecting sensitive data. The controller must make sure that all national law procedures applicable to any territory exposed to the project are followed. It must be examined carefully whether the recording of sensitive data abides to the specific national laws or whether the data that will be recorded must be tweaked appropriately.

b. Database right dimension

The copyright issues relating to the methodologies of the project are less complex, since there will be a consent for collecting data from corporate webpages. The webpages may form a "database" of the owner company. In the case of software that pulls data from the webpage, the mere permission of the company will legalize the whole operation. It should be mentioned in the relevant contracts the categories of data that will form part of the operation and the confirmation that the company owns all copyright data of its webpage. In the case of intellectual property rights reservations to third parties (i.e. webpage developers etc.), their consent should be also demanded.

6.2.2. Course of action for NSIs

The NSIs envisaging the application of IaD methods must therefore make sure that all steps of the production processes are compatible with the relevant national and EU legal framework. The following steps must be taken:

1. The legal service of the NSI carries out a thorough review of national and European legislation concerning the collection, storage and processing of personal and enterprise data for statistical purposes.
2. The production units of the NSI that will utilise the IaD methods prepare detailed descriptions of the "business cases". They contain a description of the data sources, of the means that will be

interpretation of Article 25(2) of the Universal Service Directive (2002/22/EC) and Article 12(2) of the e-Privacy Directive (2002/58/EC). It is clearly linked to the special role of directories in the Universal Service Directive.

7 This is the case according to the Greek Law Nr. 2472/1997.

used for data collection, of the data that will be collected, of the statistical purposes that will be served, of the processing they will be subjected too, of possible re-uses in the future (always for statistical purposes), e.g. re-coding for reconstruction of historical data series of new indicators or with new codelists, of the means taken to ensure and protect the anonymity of the statistical units (persons or enterprises).

3. The descriptions are scrutinised by the legal service and revisions are proposed.
4. The descriptions are finalised and are submitted to the national bodies responsible for data protection issues.
5. Taking these bodies' comments into account revised descriptions are produced and the production units examine whether the resulting production processes are still satisfactory from the statistical point of view.

6.3. Data protection legal framework

a. The Data Protection Directive

The main piece of personal data protection legislation at EU level is Directive 95/46/EC[8]. According to article 3 para. 1, the Directive shall apply to the processing of personal data wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system. Furthermore, according to Article 3, the Directive shall not apply to the processing of personal data:

- *in the course of an activity which falls outside the scope of Community law, such as those provided for by Titles V and VI of the Treaty on European Union and in any case to processing operations concerning public security, defense, State security (including the economic well-being of the State when the processing operation relates to State security matters) and the activities of the State in areas of criminal law,*
- *by a natural person in the course of a purely personal or household activity.*

Article 2 of the Directive contains a list of definitions regarding the concept of the terms used at its provisions. The most important definition clarifies the mere notion of personal data. According to Article 2 (a),

"*personal data' shall mean 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 or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*".

The concept of personal data has been extensively analyzed by the Working Party composed by the representatives of the European data protection authorities, the European Commission and the European

8 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 L 281, 23/11/1995 P. 0031 - 0050

Data Protection Supervisor that was established by Article 29 of the Directive ("The Article 29 Working Party"). According to the Working Party, there are four essential elements that should be examined in order to clarify whether the information in hand is "personal data": i) "*...any information...*", ii) "*...relating to...*", iii) "*... identified or identifiable...*", iv) "*...natural person...*"[9]. In the course of the analysis of the third element, the Working Party concluded that in general terms, a natural person can be considered as "identified" when, within a group of persons, he or she is "distinguished" from all other members of the group. Accordingly, the natural person is "identifiable" when, although the person has not been identified yet, it is possible to do it (that is the meaning of the suffix "-able"). This second alternative is therefore in practice the threshold condition determining whether information is within the scope of the third element. Identification is normally achieved through particular pieces of information which we may call "identifiers" and which hold a particularly privileged and close relationship with the particular individual. Examples are outward signs of the appearance of this person, like height, hair colour, clothing, etc… or a quality of the person which cannot be immediately perceived, like a profession, a function, a name etc. The Directive mentions those "identifiers" in the definition of "personal data" in Article 2 when it states that a natural person *"can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity".*

In the same Opinion, the Working Party gave an example on the gray areas between personal data and statistical data:

"*Apart from their general obligation to respect data protection rules, in order to ensure anonymity of the statistical surveys, statisticians are subjected to a specific duty of professional secrecy, and under those rules it is forbidden for them to publish non anonymous data. This obliges them to publish aggregated statistical data which cannot possibly be attributed to an identified person behind the statistics. This rule is particularly relevant concerning the publication of census data. In each situation a threshold should be determined under which it is deemed possible to identify the persons concerned. If a criterion appears to lead to identification in a given category of persons, however large (i.e. only one doctor operates in a town of 6000 inhabitants), this "discriminating" criterion should be dropped altogether or other criteria be added to "dilute" the results on a given person so as to allow for statistical secrecy.*"

Turning back to the Directive, there are specific provisions that relate to the processing of personal data for statistical purposes. Article 6 contains principles relating to "data quality". According to these legally binding principles, Member States shall provide that personal data must be, inter alia, collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of data for historical, *statistical* or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards. Furthermore,

9 Opinion 4/2013 on the concept of personal data, Article 29 Data Protection Working Party, WP 136, http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

according to the same Article, personal data must be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Member States shall lay down appropriate safeguards for personal data stored for longer periods for historical, *statistical* or scientific use.

According to the Article 29 Working Party interpretation of these provisions[10], they "*should not be read as providing an overall exception from the requirement of compatibility, and it is not intended as a general authorisation to further process data in all cases for historical, statistical or scientific purposes. Just like in any other case of further use, all relevant circumstances and factors must be taken into account when deciding what safeguards, if any, can be considered appropriate and sufficient. In addition, as in other situations, a separate test must be carried out to ensure that the processing has a legal basis in one of the grounds listed in Article 7 and complies with other relevant requirements of the Directive*". The Article 29 Working Party concludes that there may be three different scenarios for further analysis:

- Scenario 1: unidentifiable personal data: data are anonymised or aggregated in such a way that there is no remaining possibility to (reasonably) identify the data subjects. Full anonymisation (including a high level of aggregation) is the most definitive solution. It implies that there is no more processing of personal data and that the Directive is no longer applicable.
- Scenario 2: indirectly identifiable personal data: partial anonymisation or partial de-identification may be the appropriate solution in some situations when complete anonymisation is not practically feasible. In these cases, various techniques (including pseudo-anonymisation, key-coding, keyed-hashing, using rotating salts, removal of direct identifiers and outliers, replacing unique IDs, introduction of 'noise', and others) should be used to reduce the risk that data subjects can be re-identified, and subsequently, that any measures or decisions can be taken in their regard. In addition, there will also often be a need to complement these techniques with other safeguards in order to adequately protect the data subjects. These include data minimisation, as well as appropriate organisational and technical measures, including effective 'data silo-ing', to ensure functional separation.
- Scenario 3: situations where directly identifiable personal data are needed due to the nature of the research. Directly identifiable personal data may be processed only if anonymisation or partial anonymisation is not possible without frustrating the purpose of the processing, and further provided that other appropriate and effective safeguards are in place. Among the appropriate safeguards which may bring additional protection to the data subjects, the following could be considered:

    - taking specific additional security measures (such as encryption);
    - in case of pseudonymisation, making sure that data enabling the linking of information to a data subject (the keys) are themselves also coded or encrypted and stored separately;
    - entering into a trusted third party (TTP) arrangement in situations where a number of organisations each want to anonymise the personal data they hold for use in a collaborative project;
    - restricting access to personal data only on a need-to-know basis, carefully balancing the benefits of wider dissemination against the risks of inadvertent disclosure of personal data to unauthorized persons. This may include, for example, allowing read-only access on controlled premises. Alternatively, arrangements could be made for limited disclosure in a secure local environment to properly constituted closed communities. Legally enforceable confidentiality obligations placed

10 Opinion 3/2013 on purpose limitation, adopted on 2 April 2013, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

on the recipients of the data, including prohibiting publication of identifiable information, are also important. It is important to note that in high-risk situations, where the inadvertent disclosure of personal data would have serious or harmful consequences for individuals, even this type of access or restriction may not be suitable.

In addition,

- further processing of personal data concerning health, data about children, other vulnerable individuals, or other highly sensitive information should, in principle, be permitted only with the consent of the data subject;
- any exceptions to this requirement for consent should be specified in law, with appropriate safeguards, including technical and organisational measures to prevent undue impact on the data subjects (in case of doubt, the processing should be subject to prior authorisation of the competent data protection authority); exceptions should only apply with regard to research that serves an important public interest, and only if that research cannot possibly be carried out otherwise.

In Article 7 the Directive sets out the criteria for making data processing legitimate. There are six different legal grounds that permit the processing of personal data:

(a) the data subject has unambiguously given his consent; or
(b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; or
(c) processing is necessary for compliance with a legal obligation to which the controller is subject; or
(d) processing is necessary in order to protect the vital interests of the data subject; or
(e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed; or
(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject.

In the case of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life, there is a specific regime for the lawful processing. According to Article 8 of the Directive, processing of such special categories of data shall be prohibited by the Member States, with five concrete exemptions:

(a) the data subject has given his explicit consent to the processing of those data, except where the laws of the Member State provide that the prohibition may not be lifted by the data subject's giving his consent; or
(b) processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law in so far as it is authorized by national law providing for adequate safeguards; or
(c) processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent; or
(d) processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political,

philosophical, religious or trade-union aim and on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects; or

(e) the processing relates to data which are manifestly made public by the data subject or is necessary for the establishment, exercise or defence of legal claims.

Directive 95/46 provides for specific obligations to data controllers. One of the general transparency obligations is to provide information to the data subject, when the data have not been obtained from him or her. According to Article 11, when the data have not been obtained from the data subject, Member States shall provide that the controller or his representative must at the time of undertaking the recording of personal data or if a disclosure to a third party is envisaged, no later than the time when the data are first disclosed, provide the data subject with at least the following information, except where he already has it:

(a) the identity of the controller and of his representative, if any;

(b) the purposes of the processing;

(c) any further information such as

- the categories of data concerned,
- the recipients or categories of recipients,
- the existence of the right of access to and the right to rectify the data concerning the data subject

in so far as such further information is necessary, having regard to the specific circumstances in which the data are processed, to guarantee fair processing in respect of the data subject.

According to Article 11 para. 2, the abovementioned obligation shall not apply where, in particular for processing for *statistical purposes* or for the purposes of historical or scientific research, the provision of such information proves impossible or would involve a disproportionate effort or if recording or disclosure is expressly laid down by law. In these cases Member States shall provide appropriate safeguards.

Data processing for statistical purposes is therefore recognized as a legitimized interest that may restrict data protection principles, according to national legislation. This is stipulated in Article 13 para. 2 of the Data Protection Directive, which states that subject to adequate legal safeguards, in particular that the data are not used for taking measures or decisions regarding any particular individual, Member States may, where there is clearly no risk of breaching the privacy of the data subject, restrict by a legislative measure the rights provided for in Article 12 when data are processed solely for purposes of scientific research or are kept in personal form for a period which does not exceed the period necessary for the sole purpose of creating statistics.

b. The e-Privacy Directive

While Directive 95/46 is of a general nature, there are specific EU provisions for the protection of privacy and data protection in the field of electronic communication. The e-Privacy Directive 2002/58/EC[11] contains a set of legally binding rules concerning some fields of data processing in the electronic communications sector. The e-Privacy Directive was amended by Directive 2009/136/EC[12]. There are no specific rules governing data collection for statistical purposes in this legal framework. As a result, the general provisions on data collection for statistical purposes apply also in the electronic communications network.

Nevertheless, one should keep in mind that the e-Privacy Directive contains specific rules on mechanisms of data collection in the digital environment. From this point of view, there are provisions that may have a direct impact in assessing mechanisms that collect data from the Internet or other digital networks.

According to Article 1 para. 1 of the e-Privacy Directive, its provisions provide for the harmonization of the national provisions required to ensure an equivalent level of protection of fundamental rights and freedoms, and in particular the right to privacy and confidentiality, with respect to the processing of personal data in the electronic communication sector and to ensure the free movement of such data and of electronic communication equipment and services in the Community.

Article 3 defines the scope of the e-Privacy Directive as follows:

> *This Directive shall apply to the processing of personal data in connection with the provision of publicly available electronic communications services in public communications networks in the Community, including public communications networks supporting data collection and identification devices.*

Article 4 para. 1 ("Security of processing") states that the provider of a publicly available electronic communications service must take appropriate technical and organizational measures to safeguard security of its services, if necessary in conjunction with the provider of the public communications network with respect to network security. Having regard to the state of the art and the cost of their implementation, these measures shall ensure a level of security appropriate to the risk presented. According to para. 2, in case of a particular risk of a breach of the security of the network, the provider of a publicly available electronic communications service must inform the subscribers concerning such risk and, where the risk lies outside the scope of the measures to be taken by the service provider, of any possible remedies, including an indication of the likely costs involved. According to para. 3, in the case of a personal data breach, the provider of publicly available electronic communications

11 Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). Official Journal L 201 , 31/07/2002 P. 0037 - 0047

12 Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws Text with EEA relevance. Official Journal L 337 , 18/12/2009 P. 0011 - 0036

services shall, without undue delay, notify the personal data breach to the competent national authority.

Article 5 ("Confidentiality of the communications") obliges the Member states to prohibit listening, tapping, storage or other kinds of interception or surveillance of communications and the related traffic data by persons other than users, without the consent of the users concerned, except when legally authorized to do so in accordance with Article 15 para. 1.This provision does not affect any legally authorized recording of communications and the related traffic data when carried out in the course of lawful business practice for the purpose of providing evidence of a commercial transaction or of any other business communication. Member States shall ensure that the storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her *consent*, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing. This shall not prevent any technical storage or access for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or as strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service.

Specific provisions of the e-Privacy Directive regulate the processing of traffic data and location data. According to Article 6 data relating to subscribers and users processed and stored by the provider of a public communications network or publicly available electronic communications service must be erased or made anonymous when it is no longer needed for the purpose of the transmission of a communication. Traffic data necessary for the purposes of subscriber billing and interconnection payments may be processed. Such processing is permissible only up to the end of the period during which the bill may lawfully be challenged or payment pursued. According to Article 9, where location data other than traffic data, relating to users or subscribers of public communications networks or publicly available electronic communications services, can be processed, such data may only be processed when they are made anonymous, or with the consent of the users or subscribers to the extent and for the duration necessary for the provision of a value added service. The service provider must inform the users or subscribers, prior to obtaining their consent, of the type of location data other than traffic data which will be processed, of the purposes and duration of the processing and whether the data will be transmitted to a third party for the purpose of providing the value added service. Users or subscribers shall be given the possibility to withdraw their consent for the processing of location data other than traffic data at any time. Where consent of the users or subscribers has been obtained for the processing of location data other than traffic data, the user or subscriber must continue to have the possibility, using a simple means and free of charge, of temporarily refusing the processing of such data for each connection to the network or for each transmission of a communication.

c. Council Framework Decision on data protection in the framework of police and judicial cooperation in criminal matters

The Data Protection Directive and the e-Privacy Directive contain provisions that apply to the former "First Pillar" according to a former version of the European Union Treaty (namely: the European Community law). After the Lisbon Treaty, the scope of the secondary community legislation obtains a new dimension, which does not fall within the aim of this study to describe. Under the three-pillars system, the European Union adopted a specific set of data protection rules applying in the framework of

police and judicial cooperation in criminal matters. This is the Data Protection Framework Decision[13], which contains specific provisions for data protection in this field.

According to Nr. 6 of the preamble, the Data Protection Framework Decision applies only to data gathered or processed by competent authorities for the purpose of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties. This Framework Decision should leave it to Member States to determine more precisely at national level which other purposes are to be considered as incompatible with the purpose for which the personal data were originally collected. In general, further processing for historical, statistical or scientific purposes should not be considered as incompatible with the original purpose of the processing.

The non-incompatibility principle is stipulated in Article 3 of the Decision ("Principles of lawfulness, proportionality and purpose"):

> "*1. Personal data may be collected by the competent authorities only for specified, explicit and legitimate purposes in the framework of their tasks and may be processed only for the same purpose for which data were collected. Processing of the data shall be lawful and adequate, relevant and not excessive in relation to the purposes for which they are collected.*
>
> *2. Further processing for another purpose shall be permitted in so far as:*
>
> *(a) it is not incompatible with the purposes for which the data were collected;*
>
> *(b) the competent authorities are authorised to process such data for such other purpose in accordance with the applicable legal provisions; and*
>
> *(c) processing is necessary and proportionate to that other purpose.*
>
> *The competent authorities may also further process the transmitted personal data for historical, statistical or scientific purposes, provided that Member States provide appropriate safeguards, such as making the data anonymous.*"

One more exceptional provision for statistical purposes is contained in Article 11 ("Processing of personal data received from or made available by another Member State")

> "*Personal data received from or made available by the competent authority of another Member State may, in accordance with the requirements of Article 3(2), be further processed only for the following purposes other than those for which they were transmitted or made available:*

13 Council Framework Decision 2008/977/JHA of 27 November 2008 on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters

(a) the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties other than those for which they were transmitted or made available;

(b) other judicial and administrative proceedings directly related to the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties;

(c) the prevention of an immediate and serious threat to public security; or

(d) any other purpose only with the prior consent of the transmitting Member State or with the consent of the data subject, given in accordance with national law.

The competent authorities may also further process the transmitted personal data for historical, statistical or scientific purposes, provided that Member States provide appropriate safeguards, such as, for example, making the data anonymous."

d. Council of Europe Treaties

The Council of Europe was established in 1949 to enable governments of the European states to co-operate "*to achieve a greater unity between its members for the purpose of safeguarding and realising the ideals and principles which are their common heritage and facilitating their economic and social progress*" (Article 1 of the Statute of the Council of Europe). The international organization is governed by the Committee of Ministers of Foreign Affairs of the member states, which is advised by the Parliamentary Assembly, and many intergovernmental committees of experts dealing with most aspects of the daily life of European citizens, except defence: human rights, harmonization of law, culture and education, social affairs, public health and the economy. The Council of Europe's activities focus in particular on "topical issues" such as problems linked to drugs, terrorism, refugees and the prevention of torture.

The Council of Europe Convention for the Protection of Human Rights and Fundamental Freedoms was opened for signature in 1950. Article 8 of this Convention states that "everyone has the right to respect for his private and family life, his home and his correspondence". This right can be restricted by a public authority only in accordance with domestic law and in so far as it is necessary, in a democratic society, for the defence of a number of legitimate aims. But the Convention also lays down, in Article 10, the fundamental right to freedom of expression. This right includes explicitly the "freedom to receive and impart information and ideas without interference by public authority and regardless of frontiers". The "freedom to receive information" set out in Article 10 is considered as implying the "freedom to seek information". Articles 8 and 10 are not contradictory but complementary. However, in practice, the exercise of one of these rights can be restricted by the exercise of the other. For this reason, the European Commission and Court of Human Rights have defined in case-law the limits to the exercise of each of these rights and, in particular, the extent to which public authorities have the right to interfere. This case-law has been - and still is - of great importance to the Council of Europe in its work on data protection as the source of criteria for the development of national regulations on data protection. Nevertheless, in the years following the adoption of the European Convention on Human Rights, it became apparent that efficient legal protection of privacy required more specific and systematic development.

The first international legally binding text on data protection was adopted by the Council of Europe Member States in 1981. The European Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data[14] is a "first generation" international treaty that has been ratified by all countries of the European area. Even in this primary piece of legislation, restrictions to national data protection rules for statistical purposes were expressly considered as acceptable. According to the Convention's Article 9, restrictions on the exercise of the rights specified in Article 8, paragraphs b, c and d, may be provided by law with respect to automated personal data files used for statistics or for scientific research purposes when there is obviously no risk of an infringement of the privacy of the data subjects. The Council of Europe Convention served as a model for the drafting of Directive 95/46/EC.

The current impact of the Data Protection Convention with regard to the processing of personal data for statistical purposes is connected mainly to a secondary Council of Europe text that applies the Convention's principles to the special sector of statistical activities. Recommendation No. R (97) 18 concerning the protection of personal data collected and processed for statistical purposes[15] was adopted by the Council of Europe's Committee of Ministers on 30 September 1997. This text replaced Recommendation No. R (83) 10 on the protection of personal data used for scientific research and statistics in so far as that recommendation applies to the collection and automatic processing of personal data for statistical purposes.

According to its preamble, the Recommendation reads that the Committee of the Ministers recognizes that "the production of reliable statistics depends to a great extent on the collection of the most detailed information possible and on the processing of this information by means of increasingly effective automatic data processing technology", while it is also "aware of the fact that such information may concern identified or identifiable persons ("personal data")" and "aware of the need to develop techniques making it possible to guarantee the anonymity of the data subjects" and of "the concern of the international community of statisticians for the protection of personal data, and the development of international recommendations with regard to the professional ethics of statisticians".

The Appendix to Recommendation No. R (97) 18 contains the substantial contribution of this secondary Council of Europe text to the subject matter of data protection in the statistical sector. The Appendix contains a definitions chapter. According to this, "personal data" *"means any information relating to an identified or identifiable individual ("data subject"). An individual shall not be regarded as "identifiable" if the identification requires an unreasonable amount of time and man-power. Where an individual is not identifiable, data are said to be anonymous.*" As "identification data", the Appendix defines those personal data "*that allow direct identification of the data subject, and which are needed for the collection, checking and matching of the data, but are not subsequently used for drawing up statistical results.*" As "sensitive data" the Appendix defines the ones that have been defined as "special categories" of data by the Data Protection Convention": racial origin, political opinions, religious or other beliefs, health, sexual life, criminal convictions "and other data defined as sensitive by domestic law". As "processing" the Appendix defines any operation or set of operations carried out partly or completely with the help of automated processes and applied to personal data, "*such as storage, conservation,*

14 European Treaty Series, No. 108, http://conventions.coe.int/Treaty/en/Treaties/Html/108.htm

15 Text available on https://wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=2001724&SecMode=1&DocId=578856&Usage=2

adaptation or alteration, extraction, consultation, utilization, communication, matching or interconnecting and erasure or destruction." The Appendix contains an additional definition, for the term of "communication". It refers to the act of "*making personal data accessible to third parties, regardless of the means or media used*". There are two different definitions for the terms "statistical purposes" and "statistical results". The first term refers to "*any operation of collection and processing of personal data necessary for statistical surveys or for the production of statistical results. Such operations exclude any use of the information obtained for decisions or measures concerning a particular individual*". The second term means information which has been obtained by processing personal data "*in order to characterize a collective phenomenon in a considered population*".

Chapter 2 to the Appendix defines the scope of the recommendation, which includes the collection and automated processing of personal data for statistical purposes and extends to the statistical results, to the extent that they permit identification of data subjects. The scope chapter provides that no personal data shall be processed in a non-automatic manner in order to avoid the provisions of this recommendation.

Chapter 3 to the Appendix ("Respect for privacy") contains three general principles concerning the right to privacy.

(a) Privacy should be respected in all three stages of personal data collection and processing:
   - when these data are kept for future use;
   - when statistical results are disseminated;
   - when, for reasons of better ensuring that statistical records are representative or for reasons of confidentiality, personal data need to be modified.

(b) Persons involved in a statistical activity that contains personal data collection and processing shall be subject to a duty of professional secrecy by domestic law or practice.

(c) Personal data collected and processed for statistical purposes shall be made anonymous as soon as they are no longer necessary in an identifiable form.

Chapter 4 to the Appendix contains general conditions for lawful collection and processing for statistical purposes. Under the title "*Purpose*", this Chapter stipulates a more concrete application of the purpose limitation principle: "personal data collected and processed for statistical purposes shall serve only those purposes. They shall not be used to take a decision or measure in respect of the data subject, nor to supplement or correct files containing personal data which are processed for non – statistical purposes. Processing for statistical purposes of personal data collected for non-statistical purposes is not incompatible with the purpose(s) for which the data were initially collected if appropriate safeguards are provided for, in particular to prevent the use of data for supporting decisions or measures in respect of the data subject. Under the title "Lawfulness", the Chapter reiterates the legality criteria that were previously stipulated in Article 6 of the Data Protection Directive and the transparency obligations set forth in Section IV of the Data Protection Directive. Consent plays a crucial role when examining the legality of data processing for statistical purposes, while the Appendix adds a provision according to which "personal data may be collected on a compulsory basis with a view to their being processed for statistical purposes only if required by domestic law". According to the proportionality principle, "*only those personal data shall be collected and processed which are necessary for the statistical purposes to be achieved. In particular, identification data shall be collected and processed only if this is necessary.*" Under the title "Sensitive data", the Appendix reiterates that if these data are to be processed for statistical

purposes, they should be collected in a form in which the data subject is not identifiable. In the case the statistical purposes necessitates the identification of the data subjects, domestic law shall provide appropriate safeguards including specific measures to separate identification data as from the stage of collection unless it is manifestly unreasonable or impracticable to do so.

Chapter 5 to the Appendix provides extensive conditions of information to be given to the data subject. Under the title "*Primary collection*", the text reads that the persons questioned shall be informed of the following elements:

(a) the compulsory or optional nature of the response and the legal basis, if any, of the collection,
(b) the purpose or purposes of the collection and processing,
(c) the name and position of the person or body in charge of the collection and/or processing,
(d) the fact that the data will be kept confidential and used exclusively for statistical purposes,
(e) the possibility of obtaining further information on request.

At their request and/or according to the ways and means defined by domestic law, data subjects shall also be informed of the following:

(f) the way in which consent can be refused or withdrawn, in the case of optional surveys and, in the case of compulsory surveys, the possible sanctions this would entail,
(g) where applicable, the conditions of the exercise of the rights of access and rectification,
(h) the categories of persons or bodies to whom the personal data may be communicated,
(i) the guarantees to ensure the confidentiality and the protection of personal data,
(j) the categories of data collected and processed.

When the data subjects are not directly questioned, they shall be informed of the existence of the collection unless this is manifestly unreasonable or impracticable. They shall be able to inform themselves appropriately of the elements listed above. The persons questioned shall be informed at the latest at the time of collection. Under the title "Secondary collection", the Chapter reads that cases of processing or communication for statistical purposes of personal data collected for non-statistical purposes shall receive suitable publicity. The data subjects shall be able to obtain in a suitable way all abovementioned information, unless:

(a) this is impossible or involves a disproportionate effort,
(b) the processing or communication of the data for statistical purposes is expressly provided for under domestic law.

Chapter 6 to the Appendix ("*Consent*") reiterates that consent of the data subject, when required, shall be free, informed and unambiguous and that the data subject shall be able to withdraw his or her consent for a single survey, as long as, identification data have not been separated from other data collected, or to suspend at any time and without retroactive effect his or her co-operation in a survey which extends over a period of time. Refusal to reply shall not be penalized unless domestic law provides for sanctions.

Chapter 7 to the Appendix provides for the rights of access and rectification. Any person may obtain the personal data concerning him or her held by the data controller and, as the case may be, have them

rectified. However, where there is clearly no risk of breaching the privacy of the data subject, this right may be restricted in accordance with domestic law when the personal data are processed solely for statistical purposes and specific appropriate measures exist to prevent any identification by a third party on the basis of individual data or of statistical results.

Under the title "Rendering data anonymous" (Chapter 8), the Appendix introduces the principle that personal data collected for statistical purposes shall be made anonymous immediately after the end of data collection, checking or matching operations, except:

(a) if identification data remain necessary for statistical purposes and the measures prescribed by principle 10.1 have been taken; or
(b) if the very nature of statistical processing necessitates the starting of other processing operations before the data have been made anonymous as long as the safeguards envisaged in principles 15.1. to 15.3 are in force.

Reiterating the fairness of data collection principle, Chapter 9 ("Primary collection of personal data for statistical purposes") to the Appendix underlines that personal data shall be collected only from a person other than the data subject if domestic law provides for it and includes appropriate safeguards, or there is manifestly no risk of infringement of the rights and fundamental freedoms of the data subject. Exemptions are recognized where domestic law includes appropriate safeguards and:

(a) provides for the collection with identification data or
(b) permits the linking of the data collected to identification data for the construction of samples.

According to this Chapter, data on non-respondents relevant to the planning or carrying out of the survey, or information on the reasons for non – response, may be used only in order to ensure the representative quality of the survey. The controller shall take appropriate measures to allow the persons questioned to assure themselves of the authority to act of the person collecting the data.

The Appendix contains also two principles on "Identification data" (Chapter 10). When these data are collected and processed for statistical purposes, they shall be separated and conserved separately from other personal data, unless it is manifestly unreasonable or impracticable to do so. These data may, however, be used to create a file of addresses for statistical purposes if provided for by domestic law, if the data subject has been informed and has not opposed it, or if the data come from a file accessible to the public.

With regard to the conservation of data, Chapter 11 provides that, unless they have been made anonymous, or domestic law provides for these data to be kept for archiving purposes subject to appropriate safeguards, personal data collected and processed for statistical purposes shall be destroyed or erased when they are no longer necessary for those purposes. In particular, identification data shall be destroyed or erased as soon as they are no longer necessary:

(a) for the collection, checking and matching of the data; or
(b) to ensure the representativeness of the survey; or
(c) to repeat the survey with the same people.

Under the title "Communication", Chapter 12 to the Appendix states that personal data collected for statistical purposes shall not be communicated for non-statistical purposes. Nevertheless, personal data processed for a given statistical purpose may be communicated for other statistical purposes as long as these are specified and of limited duration. Communication in accordance with this principle shall be the subject of a written document setting out the rights and obligation of the parties, unless safeguards are provided for by domestic law. The controller shall in particular:

(a) stipulate that the third party may communicate these data only with the express agreement of the said controller;
(b) stipulate that the third party take appropriate security measures, in accordance with principles 15.1 to 15.3 of this recommendation and
(c) ensure that any publication of statistical results obtained by this party will conform with principle 14 of this recommendation.

Sensitive data communication is allowed where provided for by the law, or where the data subjects have given their explicit consent and provided domestic law does not prohibit the giving of the consent.

According to Chapter 13, the principles of this recommendation shall be applicable to the transborder communication of personal data for statistical purposes, under the relevant provisions of the Data Protection Convention (and its Protocol on transborder data flows, that had not entered into force when the recommendation was adopted).

Statistical results shall be published or made accessible to third parties only if measures have been taken to ensure that the data subjects are no longer identifiable on the basis of these results, unless dissemination or publication manifestly presents no risk of infringing the privacy of the data subjects (Chapter 14).

With regard to security of personal data, Chapter 15 reiterates general principles concerning the relevant obligations of the data controller. If data must be retained in an identifiable form, organisational and technical resources, in particular automated resources, shall be used to prevent unauthorized identification of the data subject. Measures shall be taken to prevent re-identification of data subjects and use for non-statistical purposes of personal data collected for statistical purposes. Professionals, firms or bodies in charge of producing statistics shall develop techniques and procedures ensuring the anonymity of data subjects. According to Chapter 16 ("Codes of ethics"), entities in charge of producing statistics should adopt and publish codes of professional ethics which meet the principles set out in this recommendation, in particular:

(a) on the other categories of persons and bodies which have access to the personal data;
(b) on the measures to be taken for the protection, confidentiality and security of these data as well as measures to respect statistical ethics;
(c) on the controllers of statistical processing.

According to Chapter 17, in order to ensure broad access of information tools and to technical knowledge appropriate to effective protection of personal data collected for statistical purposes, competent governmental bodies should collaborate closely in the development of these tools and technical knowledge, and should set up international programmes of co-operation, exchanges of experience,

transfer of knowledge and technical assistance. According to Chapter 18, member states give one or more independent authorities responsibility for ensuring the application of the provisions of domestic law giving effect to the principles laid down in the recommendation.

For a deeper analysis of the principles laid down in the recommendation, an Explanatory Memorandum[16] is also available.

6.4. The *sui generis* Database Right

A database right is a special formulation of copyright legal provisions that exist to recognize the investment that is made in compiling a database, even when this does not involve the "creative" aspect that is reflected by copyright. In European Union law, database rights are specifically coded laws on the copying and dissemination of information in computer databases. These rights were first introduced in 1996. The relevant legally binding instrument is Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

The Database Directive contains a definition for the "database", according to which this term shall mean a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means. It is expressly stipulated that protection under the Database Directive shall not apply to computer programs used in the making or operation of databases accessible by electronic means.

According to Article 2, there are some limitations on the scope of the Database Directive: it shall apply without prejudice to Community provisions relating to:

(a) the legal protection of computer programs;
(b) rental right, lending right and certain rights related to copyright in the field of intellectual property;
(c) the term of protection of copyright and certain related rights.

The object of the legal protection provided for by the Directive is described in Article 3. Databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright. No other criteria shall be applied to determine their eligibility for that protection. The copyright protection of databases provided for by the Directive shall not extend to their contents and shall be without prejudice to any rights subsisting in those contents themselves.

Article 4 to the Directive defines the database authorship. The author of a database shall be the natural person or group of natural persons who created the base or, where the legislation of the Member States so permits, the legal person designated as the right holder by that legislation. Where collective works are recognized by the legislation of a Member State, the economic rights shall be owned by the person holding the copyright. In respect of a database created by a group of natural persons jointly, the exclusive rights shall be owned jointly.

16 http://www.coe.int/t/dghl/standardsetting/dataprotection/EM/EM_R(97)18_EN.pdf

According to Article 5 (“Restricted acts”) in respect of the expression of the database which is protectable by copyright, the author of a database shall have the exclusive right to carry out or to authorize:

(a) temporary or permanent reproduction by any means and in any form, in whole or in part;
(b) translation, adaptation, arrangement and any other alteration;
(c) any form of distribution to the public of the database or of copies thereof. The first sale in the Community of a copy of the database by the rightholder or with his consent shall exhaust the right to control resale of that copy within the Community;
(d) any communication, display or performance to the public;
(e) any reproduction, distribution, communication, display or performance to the public of the results of the acts referred to in (b).

Article 6 provides for exceptions to restricted acts. The performance by the lawful user of a database or of a copy thereof of any of the acts listed in Article 5 which is necessary for the purposes of access to the contents of the databases and normal use of the contents by the lawful user shall not require the authorization of the author of the database. Where the lawful user is authorized to use only part of the database, this provision shall apply only to that part. Member States shall have the option of providing for limitations on the rights set out in Article 5 in the following cases:

(a) in the case of reproduction for private purposes of a non-electronic database;
(b) where there is use for the sole purpose of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved;
(c) where there is use for the purposes of public security of for the purposes of an administrative or judicial procedure;
(d) where other exceptions to copyright which are traditionally authorized under national law are involved, without prejudice to points (a), (b) and (c).

In accordance with the Berne Convention for the protection of Literary and Artistic Works, this Article may not be interpreted in such a way as to allow its application to be used in a manner which unreasonably prejudices the rightholder's legitimate interests or conflicts with normal exploitation of the database.

The “sui generis” database right is stipulated in Article 7 of the Directive. Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database. The sui generis right may be transferred, assigned or granted under contractual license. It shall also apply irrespective of the eligibility of that database for protection by copyright or by other rights. Moreover, it shall apply irrespective of eligibility of the contents of that database for protection by copyright or by other rights. Protection of databases under the right provided for in paragraph 1 shall be without prejudice to rights existing in respect of their contents. For the purposes of this Directive:

(a) 'extraction` shall mean the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form;

(b) 're-utilization` shall mean any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission. The first sale of a copy of a database within the Community by the rightholder or with his consent shall exhaust the right to control resale of that copy within the Community; public lending is not an act of extraction or re-utilization.

The repeated and systematic extraction and/or re-utilization of insubstantial parts of the contents of the database implying acts which conflict with a normal exploitation of that database or which unreasonably prejudice the legitimate interests of the maker of the database shall not be permitted.

Article 8 provides for rights and obligations of lawful users. The maker of a database which is made available to the public in whatever manner may not prevent a lawful user of the database from extracting and/or re-utilizing insubstantial parts of its contents, evaluated qualitatively and/or quantitatively, for any purposes whatsoever. Where the lawful user is authorized to extract and/or re-utilize only part of the database, this paragraph shall apply only to that part. A lawful user of a database which is made available to the public in whatever manner may not perform acts which conflict with normal exploitation of the database or unreasonably prejudice the legitimate interests of the maker of the database. A lawful user of a database which is made available to the public in any manner may not cause prejudice to the holder of a copyright or related right in respect of the works or subject matter contained in the database.

Exceptions to the sui generis right are mentioned in Article 9 to the Directive. Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorization of its maker, extract or re-utilize a substantial part of its contents:

(a) in the case of extraction for private purposes of the contents of a non-electronic database;
(b) in the case of extraction for the purposes of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved;
(c) in the case of extraction and/or re-utilization for the purposes of public security or an administrative or judicial procedure.

The right provided for in Article 7 shall run from the date of completion of the making of the database. It shall expire fifteen years from the first of January of the year following the date of completion. In the case of a database which is made available to the public in whatever manner before expiry of the period provided for, the term of protection by that right shall expire fifteen years from the first of January of the year following the date when the database was first made available to the public. Any substantial change, evaluated qualitatively or quantitatively, to the contents of a database, including any substantial change resulting from the accumulation of successive additions, deletions or alterations, which would result in the database being considered to be a substantial new investment, evaluated qualitatively or quantitatively, shall qualify the database resulting from that investment for its own term of protection.

According to Article 11 of the Directive, the right provided for in Article 7 shall apply to database whose makers or rightholders are nationals of a Member State or who have their habitual residence in the territory of the Community. This shall also apply to companies and firms formed in accordance with the law of a Member State and having their registered office, central administration or principal place of business within the Community; however, where such a company or firm has only its registered office in

the territory of the Community, its operations must be genuinely linked on an ongoing basis with the economy of a Member State. Agreements extending the right provided for in Article 7 to databases made in third countries and falling outside the abovementioned provisions shall be concluded by the EU Council acting on a proposal from the EU Commission. The term of any protection extended to databases by virtue of that procedure shall not exceed that available pursuant to Article 10.

According to Article 13, the Database Directive shall be without prejudice to provisions concerning in particular copyright, rights related to copyright or any other rights or obligations subsisting in the data, works or other materials incorporated into a database, patent rights, trade marks, design rights, the protection of national treasures, laws on restrictive practices and unfair competition, trade secrets, security, confidentiality, data protection and privacy, access to public documents, and the law of contract.

According to Article 14, protection pursuant to the Database Directive as regards copyright shall also be available in respect of databases created prior to the date referred to Article 16 which on that date fulfill the requirements laid down in this Directive as regards copyright protection of databases. Notwithstanding the abovementioned provision, where a database protected under copyright arrangements in a Member State on the date of publication of the Directive does not fulfill the eligibility criteria for copyright protection laid down in Article 3 the Directive shall not result in any curtailing in that Member State of the remaining term of protection afforded under those arrangements. Protection pursuant to the provisions of the Directive as regards the right provided for in Article 7 shall also be available in respect of databases the making of which was completed not more than fifteen years prior to the date referred to in Article 16 (1) and which on that date fulfill the requirements laid down in Article 7. The protection provided for in the abovementioned provisions shall be without prejudice to any acts concluded and rights acquired before the date referred to in those paragraphs. In the case of a database the making of which was completed not more than fifteen years prior to the date referred to in Article 16 (1), the term of protection by the right provided for in Article 7 shall expire fifteen years from the first of January following that date.

Article 15 states that any contractual provision contrary to Articles 6 (1) and 8 shall be null and void.

The Database Directive provides no mandatory public-interest exceptions comparable to those recognized under domestic and international copyright laws. An optional exemption concerning "illustrations for teaching or scientific research" applies to extractions but not reutilization[17].

6.5. Conclusion

The overall project seems to be compatible with relevant data protection and database right rules. The prior consent and permissions should comply with the abovementioned provisions. The compliance is a matter of properly drafted Terms of Service to which the end user and the companies may opt in, before the installation / operation of the data collection software to their devices or web pages. The examination of the Terms of Service by the independent Data Protection Authorities in the territories exposed to the project would also provide for an additional confirmation of the legal compatibility.

17 "The role of scientific and technical data and information in the public domain", Proceedings of a symposium, National Research Council of the National Academies, Washington DC 20001, 2003, *Jerome Reichman*, Discussion Framework, p.82

7. Socio-political acceptance

In Europe the right to privacy is enshrined in the European Convention of Human rights[18] reflecting an approach in society that values privacy and personal dignity on par with freedoms unlike in other regions like the US. Recent events, like the Snowden revelations for large government operations that collect data on individuals at a huge scale worldwide have increased public awareness on the issue of privacy with respect to governments and big data holders, especially in Europe.

In this section we will examine attitudes of stakeholders (individuals and businesses) towards a system of data collection that collects data for statistical purposes from their day to day actions.

User centric approach

In user centric approaches the individual is the reference unit and the stakeholder. Based on our pilot exercise participating users were also asked whether they had reservations about installing a data collection application and to describe them. Most of the respondents (38/40 i.e. 79%) did not have reservations and 10 (21%) provided some.

We have identified four issues that should be considered as generating (justified or not) reservations for participating in a user centric data collection system.

- Intrusiveness. Obtaining too much information. Statistical data collections for official statistics, while handling sensitive information about individuals and households, are rarely intrusive. One common exception is the information on income that is well known that generates both frustration and has relatively low response rate[19]. The low response rate for some aspects of income indicates that some respondents have limits on the kind of information that they are ready to provide in official statistics surveys. An application that is installed in the devices that a person uses to access the internet may record and dispatch information that is really too sensitive for many persons to accept even if they trust that it will only be used for statistical purposes. This issue has been brought up by 2 members of our sample, one was worried whether the content of chats and other personal communication with family and friends was recorded and another gave a general statement that the computer is used for personal matters and was sceptical about sharing these uses with others. It is therefore important that information that is collected and transmitted to the NSI is as little sensitive as it can eg. reporting category of websites visited and not individual sites.
- Confidentiality is also an issue that concerns users. Confidentiality protection is enshrined in statistical law in all countries and for all statistics produced. From comparative results in certain countries, there is more trust on behalf of the public that their data is kept confidential than to Statistical institutes in general[20] but it is nevertheless an issue that needs to be addressed.
- Security. Installing an application that collects information in the background and then sends it to another computer over the internet poses the security risk that it can be intercepted by a third party or that it can be used as a back door to gain access to their computers. This worry has been reported by three respondents that report concerns about their computer security (two were questioning whether their passwords are safe) as well as their concern that third parties could obtain

[18] Article 8 stipulates that "Everyone has the right to respect for his private and family life, his home and his correspondence" subject to certain restrictions.

[19] An assessment of survey errors in EU-SILC, ISSN 1977-0375, p.32, available at http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-10-021/EN/KS-RA-10-021-EN.PDF

[20] Public image survey of Statistics Denmark, 2011, available at http://unstats.un.org/unsd/dnss/docViewer.aspx?docID=2759

personal information. It is essential that these fears are taken seriously in the design of the software as well as assuring users for the safety of the operation.

- Transparency. What exactly the software does after it is installed in a device may also worry respondents. Two respondents express them, revealing a need to explain the way the application functions in a clear, comprehensive and specific manner (what it does and what it does not). It may also be useful to allow for verification of these claims by revealing the source code of the software although this may compromise security.

When users were asked to name conditions required for accepting to participate in such a survey they mostly named confidentiality issues (16/26), i.e. preserving anonymity. Security was also reported by some respondents (3/26). Issues related with the use of the device (slowing down, leaving traces, ease of installing and uninstalling) were indicated by three respondents (3.26). Two respondents noted that their participation depended on whether they were interested on the scope of the survey and one mentioned the degree of trust to the responsible institution. Another issue that was brought up by several respondents (7/26) was a potential incentive that they required in order to participate.

Incentives, whether monetary or nonmonetary can be considered as an inducement offered by the survey designer to compensate for the absence of factors that might otherwise stimulate cooperation--e.g., interest in the topic of the survey or a sense of civic obligation.[21] Although Official Statistics Institutes are reluctant to use incentives, which, among other issues, may render a survey very expensive they should contemplate their use when requiring installation of software in respondents devices.

- Such software uses computational resources of the device. Although it should have only a small effect on device performance this is certainly not zero. Incentives can be seen as some sort of partially renting the respondent's equipment.
- Transmission of data via 3G/4G networks may entail actual costs for cooperation that users are entitled to ask compensation for.

Site centric approach

The data collection tool chosen for the pilot used Google's Custom Search Engine in order to detect keywords in the enterprises websites' content that has already been indexed by Google. 63 enterprises (approximately 20%) have been randomly selected among those listed in our inventory. Among the 63 randomly selected enterprises, two enterprises do not have active websites and have been excluded from the analysis. Thus, the eligible sample is 61 enterprises. In the course of this assessment, we have contacted the owners of the 61 randomly selected websites in order to investigate whether they are willing to accept and implement the proposed new method of data collection. We have prepared a questionnaire, which outlined the proposed method and indicators and posed five questions in order to collect their opinions about them.

Out of the 61 selected websites that were contacted, 27 (44,3%) websites' owners have replied and 16 (26,2%) have refused to take part. The rest of the 18 (29,5%) websites owners never replied.

[21] Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. The ANNALS of the American Academy of Political and Social Science, 645(1), 112-141.

On first question asking for preference for data collection (automatic vs questionnaire based) the responses from website managers was divided. Considering that computing the indicators and filling up the questionnaire results in some burden we can ascertain that website managers have some reservations about allowing such automatic data collection tools.

1st question: **Having read the accompanied document that lists the specific indicators related to your website, which way would you prefer in order to provide data for those indicators to an Official Statistical Institution?**	**N**	**%**
Via automatic collection from my website without my interference	12	44.4
Via an appropriate questionnaire	12	44.4
Via either of the first two ways	1	3.7
None of the first two ways	1	3.7
I do not know/No answer	1	3.7
Total	**27**	

Some of those that opposed automatic collection (3 out of 13) did not mind if collection was implemented manually yet still from the statistical institute and not themselves, although most retained their objection and wanted to have responsibility for data referring to their sites.

2nd question: **If an employee from an Official Statistical Institution manually visited your website and recorded the requested data, would you still be opposed?** *(only for those who answered "Via an appropriate questionnaire" or "None of the first two ways" in the 1st question)*	**N**	**%**
Yes	10	76.9
No	3	23.1
I do not know/No answer	-	-
Total	13	

When asked about the reasons for opposing automatic collection most respondents (7/13) did not elaborate, two refused citing general reasons while the rest suggested that they want to be fully informed of the data content as well as the data collection process, while some also noted the need some verification.

3rd question: **Can you please specify the reasons, why you do not wish to automatically collect data from your website?** *(only for those who answered "Via an appropriate questionnaire" or "None of the first two ways" in the 1st question)*	**N**	**%**
No reason	7	53.8
I do not think it is necessary	1	7.7
I do not want the collected data from my site to be published by an Official Statistical Institution or to be known to my competitors	1	7.7
I want to be informed every time about which data will be used and the nature of the survey	1	7.7
I would agree to an automatic data collection if only the requested data was the one that it is described in your document. If more data is going to be collected, such as measuring website's traffic then I am opposed.	1	7.7
In order always to be able to verify the information/data is going to be requested	1	7.7
We want to know, every time, the requested information	1	7.7
Total	13	

4th question: **In order to give your permission for an automatic data collection from your website, would you require some kind of a confidentiality guarantee?**	**N**	**%**
Yes	14	51.8
No	10	37.0
I do not know/No answer	3	11.1
Total	27	

Respondents willing to cooperate mostly required some sort of bilateral agreement. Only two were satisfied with general confirmation and assurances on behalf of the statistical institute. Most of those requiring some sort of agreement wanted a cooperation agreement (9) rather than a confidentiality agreement (3). Only one respondent required financial compensation as part of the cooperation agreement.

	5th question: **What kind of confidentiality guarantee would you require?** *(only for those you answered "Yes" in 4th question)*	**N**	**%**

Confidentiality Assurance	Written confirmation that the data will be used only for the purposes of this research and will not be used for other purposes or disclosure to third parties	1	7.1
	Assurance of anonymity	1	7.1
Confidentiality agreement	Confidentiality agreement	2	14.3
	Privacy policy agreement	1	7.1
Cooperation agreement	Cooperation agreement	7	50.0
	Written agreement that data will not be used for commercial purposes and copywrite will be protected	1	7.1
	Financial compensation and a cooperation agreement	1	7.1
	Total	14	

From our small sample of website managers it seems that about half will not cooperate with an automatic survey (although some of them might be turned if they have full information on the collection process and access to the data transmitted). Those that can potentially agree see themselves as partners and not just respondents and require bilateral agreements rather than self-imposed rules and commitments from the National Statistical Institute.

8. Conclusions

Two separate production processes, one web site-centric and the other user-centric have been examined in this report:

- the production of statistics on the characteristics of business web sites, based on data collected with the help of crawlers or search engines that rely on earlier crawling from the said web sites.
- the production of statistics on the use of Internet by individuals, based on data collected with the help of monitoring software installed on the users' devices.

The two processes have been examined from several angles.

Technically they are both feasible. Software components are available in several forms and the software technologies needed for development from scratch are commonplace. The capacities needed for development and maintenance are quite easy to find in the job market even if not already available to the NSIs.

The processes are also acceptable in the ESS, according to the small sample of NSIs that were interviewed. The NSI most opposed to these processes was mainly not aware of their details and potential, and expressed concerns about the additional workload that they would impose. In general however, NSIs are at least curious about these methods and see their potential. Some of them are already studying them.

The two processes diverge in the conclusions about their methodological feasibility. The both produce very relevant, timely and rich-in-detail statistics. Compared to the current ICT surveys the web-site centric process has a much narrower scope: it substitutes and expands a small subset of the current survey's indicators, while the user-centric process can reproduce most current indicators. The user-centric process thus also offers great savings in response burden. Both have accuracy issues: the web site-centric one suffers from measurement errors, in its keyword-based implementation and possibly by non-response. The user-centric one mainly suffers from non-response, manifested as refusals to participate or switching off of the monitoring software occasionally.

The two processes also achieve different cost-benefit balance. The web site-centric process seems to have too high costs for the benefits it offers, especially if one takes into account that it covers a small subset of current indicators and has reduced accuracy. The user-centric approach seems to be more expensive than the current ICT survey but reduces response burden and production times considerably. Unfortunately there was no detailed cost information about these processes or the current ICT surveys so as to make a more precise assessment.

The processes are compatible with current European legislation, as long as NSIs inform explicitly individuals and enterprises about the collected data and the uses they will be subjected to and they obtain the sample units' consent. In principle the processes do not differ from traditional surveys that collect sensitive business or personal data.

In user centric approach we found that most users want to cooperate and will do so if they are satisfied that their privacy and anonymity will be preserved and their use of their devices will not be affected in a

substantial way. Incentives may help to further increase cooperation. Regarding the site centric approach, a large part of websites (about half) will refuse cooperation and those that can potentially agree see themselves as partners and not just respondents and require bilateral agreements rather than self-imposed rules and commitments from the National Statistical Institute.

Overall, the user-centric process is the more feasible of the two. It can replace the current ICT survey to a great extent for a not much higher cost. The same cannot be said for the web-site process. As envisaged it collects a small subset of the current survey's indicators. A variation, namely the collection of data from enterprise servers, which was outside the scope of the project, can supplement this process and can deliver a much larger set of highly relevant ICT and other enterprise data.

9. References

[Beach2010] Beach, A., Gartrell, M., Xing, X., Han, R., Lv, Q., Mishra, S., &Seada, K. (2010, February). Fusing mobile, sensor, and social data to fully enable context-aware computing.In *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications* (pp. 60-65).ACM.

[Kang2011] Kang, J. M., Seo, S. S., & Hong, J. K. (2011, September).Usage pattern analysis of smartphones. In *Network Operations and Management Symposium (APNOMS), 2011 13th Asia-Pacific* (pp. 1-8). IEEE.

[Koster 1995] Martijn Koster, Robots in the Web: threat or treat? ConneXions, Volume 9, No. 4, April1995. http://info.webcrawler.com/mak/projects/robots/threat-or-treat.html

[Miller 2012] Miller, G. (2012). The smartphone psychology manifesto.*Perspectives on Psychological Science*, *7*(3), 221-237.

[Mokh2007]Mokhonoana, P. M., & Olivier, M. S. (2007, September). Acquisition of a Symbian smart phone's content with an on-phone forensic tool. In *Proceedings of the Southern African Telecommunication Networks and Applications Conference* (pp. 1-7).

[Rofouei 2012] Rofouei, M., Wilson, A., Brush, A. J., &Tansley, S. (2012, May). Your phone or mine?: fusing body, touch and device sensing for multi-user device-display interaction. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (pp. 1915-1918).ACM.

[Shepard2011] Shepard, C., Rahmati, A., Tossell, C., Zhong, L., &Kortum, P. (2011).LiveLab: measuring wireless networks and smartphone users in the field. *ACM SIGMETRICS Performance Evaluation Review*, *38*(3), 15-20.

[Souza 2010] de Souza, M., Carvalho, D. D. B., Barth, P., Ramos, J. V., Comunello, E., & von Wangenheim, A. (2010, August).Using acceleration data from smartphones to interact with 3D medical data. In *Graphics, Patterns and Images (SIBGRAPI), 2010 23rd SIBGRAPI Conference on* (pp. 339-345). IEEE

[Vafopoulos 2011] Vafopoulos, M. (2011). The Web economy: goods, users, models and policies. Foundations and Trends® in Web Science, 3(1-2), 1–136. doi:http://dx.doi.org/10.1561/1800000015

[Wagner2013] Wagner, D. T., Rice, A., & Beresford, A. R. Device Analyzer: Large-scale mobile data collection.

[Weitzner et all 2008] Weitzner, D., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., &Sussman, G. J. (2008). Information accountability. Communications of the ACM, 51(6), 82–87

.

10. Annex

10.1. Appendix 1 - Synonym XML definition

```
<?xml version="1.0" encoding="UTF-8"?>
<Synonyms start="0" num="20" total="20">
<Synonym term="B8b">
<Variant>privacy policy</Variant>
<Variant>terms of use</Variant>
<Variant>Privacy Statement</Variant>
<Variant>Conditions of use</Variant>
<Variant>Terms and Conditions</Variant>
<Variant>Terms &amp;amp; Co</Variant>
</Synonym>
<Synonym term="B8g">
<Variant>jobs</Variant>
<Variant>vacancies</Variant>
</Synonym>
<Synonym term="B8p1">
<Variant>cart</Variant>
<Variant>shopping basket</Variant>
</Synonym>
<Synonym term="N11a">
<Variant>widgets</Variant>
<Variant>Facebook</Variant>
<Variant>LinkedIn</Variant>
<Variant>Yammer</Variant>
<Variant>Twitter</Variant>
<Variant>Follow us</Variant>
<Variant>Share this page</Variant>
<Variant>Follow</Variant>
<Variant>Like us</Variant>
</Synonym>
<Synonym term="N11b">
<Variant>Blogs</Variant>
<Variant>Follow</Variant>
</Synonym>
<Synonym term="N13">
<Variant>Wiki</Variant>
</Synonym>
<Synonym term="N14">
<Variant>Creative commons (licence)</Variant>
<Variant>rss (feed)</Variant>
</Synonym>
<Synonym term="N18">
<Variant>Workflow Engine</Variant>
</Synonym>
<Synonym term="N1a">
<Variant>url</Variant>
<Variant>Website</Variant>
</Synonym>
<Synonym term="N1b">
<Variant>e-mail</Variant>
<Variant>Email</Variant>
```

<Variant>E-mail</Variant>
<Variant>email</Variant>
<Variant>eMail</Variant>
</Synonym>
<Synonym term="N1c">
<Variant>telephone</Variant>
<Variant>telephone number</Variant>
<Variant>Phone</Variant>
<Variant>Tel.</Variant>
<Variant>Fax</Variant>
<Variant>Tel/Fax</Variant>
<Variant>T:</Variant>
<Variant>tel</Variant>
<Variant>TELEPHONE</Variant>
</Synonym>
<Synonym term="N1d">
<Variant>address</Variant>
<Variant>Postal Address</Variant>
<Variant>Post code</Variant>
<Variant>P.O. box</Variant>
</Synonym>
<Synonym term="N22">
<Variant>Online chat</Variant>
</Synonym>
<Synonym term="N2a">
<Variant>Language</Variant>
<Variant>Greek</Variant>
<Variant>EL</Variant>
</Synonym>
<Synonym term="N2b">
<Variant>English</Variant>
<Variant>EN</Variant>
</Synonym>
<Synonym term="N3">
<Variant>Last Update</Variant>
<Variant>Last Updated Dated</Variant>
</Synonym>
<Synonym term="N4">
<Variant>Signin</Variant>
<Variant>login</Variant>
<Variant>Login</Variant>
<Variant>register</Variant>
<Variant>Create an Account</Variant>
<Variant>openID</Variant>
<Variant>registration</Variant>
<Variant>Subscribe</Variant>
</Synonym>
<Synonym term="N5">
<Variant>sitemap</Variant>
<Variant>site map</Variant>
<Variant>SITEMAP</Variant>
<Variant>Sitemap</Variant>
<Variant>Site Map</Variant>
</Synonym>
<Synonym term="N6">
<Variant>analytics</Variant>

<Variant>googleanalytics</Variant>
</Synonym>
<Synonym term="N9">
<Variant>mpeg</Variant>
</Synonym>
</Synonyms>

10.2. Appendix 2

Tools (open/free) for mobile data collection

iPhone Analyzer:(http://www.crypticbit.com/zen/products/iphoneanalyzer) allows you to forensically examine or recover date from in iOS device. It principally works by importing backups produced by iTunes or third party software, and providing you with a rich interface to explore, analyse and recover data in human readable formats. Because it works from the backup files everything is forensically safe, and no changes are made to the original data.

BitPim:(http://www.bitpim.org/) is a program that allows you to view and manipulate data on many CDMA phones from LG, Samsung, Sanyo and other manufacturers. This includes the PhoneBook, Calendar, WallPapers, RingTones (functionality varies by phone) and the Filesystem for most Qualcomm CDMA chipset based phones. To see when phones will be supported, which ones are already supported and which features are supported

VIAFORENSICS: (https://viaforensics.com/resources/tools/) viaForensics has developed a number of free mobile and computer forensics tools.

Mobile Internal Acquisition Tool (MIAT): (http://computerforensics.champlain.edu/blog-tags/mobile-internal-acquisition-tool**). The tool is presented in**[Distefano2008].It seems that is freely available after request to authors.

TULP2G(http://tulp2g.sourceforge.net/): forensic framework for extracting and decoding data.

Commercial tool:

Lantern: (http://katanaforensics.com/): Well-known tool for iPhone, iPod, iPad. New releases support Android devices.

10.3. Appendix 3 – Topics for discussion with the NSIs for the assessment of feasibility in the ESS

Introduction

The project has several objectives related to the employment of modern and enhanced methodologies for producing official statistics from non-traditional data sources such as the Internet or Big Data.

The discussion with a selected group of National Statistical Institutes (NSIs), indicated by Eurostat / Unit G6, will provide input for assessing the feasibility of producing official statistics about the information society based on data obtained with two specific types of measurement:

1. **User-centric**: Automatic recording, with some sort of benevolent monitoring software, of data generated while individuals use the internet with personal devices such as computers, tablets and smartphones.
2. **Enterprise website-centric**: Automatic extraction, with some sort of benevolent web crawler, of data available in the websites of business enterprises about functionalities the websites offer to users and about characteristics of the enterprises (e.g. engagement in e-sales, price lists of products, vacancies, etc.).

Both types of measurement would be used only with the explicit consent of the targeted individuals or enterprises respectively. Moreover, data collected with them could be complemented with data collected with more “traditional” methods (e.g. surveys, data extraction from registers, etc.).

The following list contains the topics to be discussed with the NSIs. It is not a questionnaire but a roadmap of the discussion.

Activities of the NSI in this area

Discussion about statistical production activities of the NSI that involved user-centric or website-centric measurements similar to those described in the introduction. It does not matter whether they are still on-going or whether they are test activities or regular production ones.

1. Description of activities
    a. Target indicators
    b. Target population / statistical units
    c. Collected variables
    d. Sample design / sampling frame / sample selection procedure / sample size
    e. Measurement mode → please identify cases where combinations of automatic measurements and traditional survey methods were used
    f. Response rates
    g. Data processing and data analysis requirements

    h. Hardware and software used
2. Additional information
    a. Reasons for undertaking these activities
    b. Problems encountered
    c. Notable experiences
    d. Effort and cost
    e. The NSI's / your "verdict" about the activities?
    f. Are they still on-going?

Opinion about these methods of measurement

It is of interest to have the NSI's opinion about the feasibility and applicability of these methods in the context of the European Statistical System (ESS).

1. If there has been no such activity / If there were activities but they have been stopped, why is that?
2. Future plans, schedules
3. Opinion about the feasibility of the methods of measurement
    a. Legal barriers foreseen
    b. Quality of statistics (coverage of target population, coverage of the phenomena intended to be measured, non-response, precision, comparability, relevance of the produced indicators)
    c. Did you have to deal with or have you thought about issues such as:
        i. use of the regular individuals' and enterprises' sampling frames in surveys that will use the automatic measurement methods
        ii. tracking of individual users in the case of multi-device use / multi-user use of the same device
        iii. the possible association of one enterprise with multiple websites
    d. Expected degree of acceptance by targeted users and enterprises and by the public in general. Potential to alleviate fears about breach of privacy.
    e. Technical barriers, relevant competences required.
    f. Which are the biggest advantages of these methods?
    g. Which are their greatest problems?
4. Comparison with traditional surveys and production methods.
5. Likelihood of such methods being adopted for regular statistical production in the ESS.
6. Are other organisations in the country engaged in such activities, even if only for research?