What are ‘Personal Data’?

A study conducted for the UK Information Commissioner

Sharon Booth
Richard Jenkins
David Moxon
Natasha Semmens
Christopher Spencer
Mark Taylor
David Townend

The University of Sheffield

2004

What are "Personal Data"?

Contents:

General Introduction

"Personal Data" under the Directive 95/46/EC and also in the UK Data Protection Act 1998 is not well defined. The definitions used do not immediately lead to a conclusive understanding of the term, or an immediate indication of the concept underpinning the term. This research, commissioned by the UK Information Commission, is designed to assist the Commissioner in developing a robust, coherent and defensible definition and understanding of "personal data" to help to develop the data protection environment in the UK. It is not designed to produce a single answer to the question "what are personal data?" or to favour a particular approach. It is designed to examine the meanings of the term employed by other data protection authorities both in Europe and beyond to see if other jurisdictions have and employ clear definitions of the term, and it is designed to approach the term in an inter-disciplinary manner again to see if definitions of the terms can be found outside the law. If no definitions were immediately forthcoming, then the purpose of the study was then to develop key elements necessary for the understanding of the term "personal data" and to inform the construction of a practical, coherent and justifiable definition for UK law within the context of the Directive.

To that end a team of researchers from across the University of Sheffield (Dr Semmens, Dr Taylor, and Mr Townend from the Department of Law, Professor Jenkins from Sociological Studies, and Professor Spencer from the Department of Psychology) and two research assistants funded by the UK Information Commission (Ms Booth and Mr Moxon,) have undertaken the work.

Initially, the focus of the work was very wide. Issues such as "anonymisation", "relevant filing system", and the nature of information were considered relevant to the study. Following consideration of the issues, it became very clear that such issues are not central to the concept of "what are personal data?", but rather concern the relationship of data that has been defined as personal data to the obligations of the Directive or UK law. For example, the anonymisation of data is a process that operates upon personal data, it does not, of itself, assist in the definition of personal data. Further, during the research process *Durant v Financial Services Authority*[1] was decided. While this presented a definition from the Courts of "personal data" and required a response from the Information Commissioner, it was decided that this report would retain its focus on the theoretical understanding of the term "personal data" without becoming an extended commentary on the *Durant* decision. The research team has not ignored the decision, however, and a supplementary report applying the theoretical positions devised in this report will be appended to this report.

The research was undertaken through a literature review in law, sociology and psychology. The reviews of sociology and psychology are presented first in the report as they informed, alongside the legal thinking, the whole understanding of how to think about the problem. The concepts understood in the two disciplines differ from the concepts employed in law to answer the question, and it is valuable to hear them first so that their influence resonates throughout the reading of the research. Thereafter, an empirical survey of how the term personal data is understood and applied by data protection authorities in other jurisdictions was undertaken. The survey produced interesting results that, given the nature of the sample and method used in the time available, could realistically be best considered as a

[1] [2003] EWCA Civ 1746

pilot study (i.e. the results should be treated with a degree of caution so as not to read too much into the information given). However, the results were of a sufficient quality to inform the creation of theoretical "ideal types" and "composite types", the vehicle used to explain and offer the choices for defining "personal data" within the jurisdictions considered.

The Purpose and Drive of the Directive

The UK law is set against the imperatives and discretions created through the European Directive 95/46/EC. To some extent, therefore, the principles of the Directive must inform and answer the question "what are personal data?". The underpinning principle of the Directive is the protection of the rights and freedoms of individuals particularly in respect of privacy. This purpose stands within the broader purpose of the Union itself, the creation of a single market. The relationship of data protection and the single economic market is not one of opposition. Rather, personal data is protected in the EU to further create the single market by developing conditions of confidence in data subjects as consumers and citizens for the unrestricted transfer of their data. The single market is furthered when individual citizens trust that there is a regime of protection across the Union (and for transfers beyond) that protects and respects their personal data. Thus, the interpretation of the term "personal data" must be in accord with the primary driving principle, i.e. the protection of the fundamental rights and freedoms of the individual, in particular in respect of privacy. Therefore the definition of "personal data", with its position of gate-keeping the access to the protection given through the obligations and rights relating to the processing of "personal data", must respect the dignity of the individual.

One of the defining elements that make data "personal data" is the necessary link to the fundamental rights and freedoms of the individual data subject's dignity. "What are personal data?" is answered: "that information that relates to the individual's dignity, such that violation of that data would result in a violation of the fundamental rights and freedoms of that individual, especially his or her privacy." How though this answer is to be understood and applied in practice is not clear This study seeks to articulate a number of positions that might assist in understanding what data is so linked to the dignity of an individual that it requires the protection of the law.

Acknowledgements

The research team would like to thank all of the Data Protection Authorities who participated in the survey. Since it has been necessary to anonymise the respondents for this report, it is not possible for us to express our gratitude to individual countries. However, we are extremely grateful to all participants for their time and effort.

We would also like to thank the staff at the Information Commission for supporting this project with such enthusiasm.

Summary

A1: An Introduction to the Study

1. The main aim of this project was to inform the construction of a robust, defensible understanding of the term 'personal data'. This produced three key areas of questions: questions about the meaning of terms arising directly from the Directive 95/46/EC and Data Protection Act 1998; questions about the meaning of 'personal data' in other contexts; and questions about adjudicating between definitions and disciplinary contexts.

2. To answer the questions required a two-fold approach: a clear empirical understanding of how the term 'personal data' is understood by Data Protection Authorities across Europe and beyond; an interdisciplinary theoretical framework, informed by the empirical study, within which to understand the operative concept(s).

3. This was achieved through a three-phase study: Phase 1 – a literature review across the relevant disciplines; initial discussions between the relevant interdisciplinary Expert Panel on the theoretical framework; development and distribution of the first questionnaire; Phase 2 – analysis of the first questionnaire; discussion of key themes and apparent conceptual differences, inconsistencies and operational difficulties in relation to the on-going development of the theoretical framework; and development and distribution of the second questionnaire; Phase 3 – analysis of the second questionnaire and the final development of the theoretical framework. The project was completed with a three month time period.

A2: Literature Review

a. Personal Data in Sociological Perspective

4. Sociologically, the subject of 'personal data' can be addressed from two different, albeit related directions. The first focuses on the information that sociologists and other social scientists collect during empirical research. The second addresses wider processes of data construction in public and private sector organisations and the relationship that these have to the everyday lives of real people.

5. Notions of 'construction' or 'social construction' show that the human world exists because of human endeavour and co-ordination, and is culturally varied by time and place. Data are constructs from facts, from social processes of definition, selection, and collection. They serve, and are defined by, interests, and relate to the purpose of their construction. They are only capable of being understood in the social context of their construction and what counts as data will differ from context to context. Data always have a purpose - they serve interests - and those purposes are crucial in their definition, selection and collection and in deciding what is done with and to them. The key point is that data cannot be understood outside of the social contexts of their construction and the interests that they are designed to serve.

6. Social research always concerns personal data. Its truth claims are based upon their relationship back to real persons, whether dead or alive. *In principle*, therefore, there is in all social research a paper trail - or its electronic equivalent - which should, given the availability of sufficient contextual information, enable data to be tracked back to an identifiable person or persons. Without this potential court of last resort, the truth claims of social research, such as they are, could not be contemplated.

7. *In practice*, however, much social research is dependent upon the anonymity of its data. Either to prevent harm to respondents or to encourage openness. As a result, various approaches to anonymising data are integral to the data construction processes of social research. In all cases, however, what is significantly 'personal' and what is not are matters of judgement: name and address might seem fairly straightforward, but beyond that, context is all. The key issue is the link between data and identifiable persons: what is 'personal' and what is not? This is the difficult issue that informs the present attempt to determine whether or not, 'out there', a minimal consensual definition of 'personal data' can be found.

8. Current concerns with the regulation and governance of personal data can be seen to relate to established concerns with aspects of modernity, such as bureaucracy and formal organisations, the state, social control, and surveillance; the identification and accountability of the individual in modernity. A broad sociological discourse has developed concerned with identification and how it serves the interests of the powerful. This reflects and informs public concern about personal data and its management, although reflection on the inefficiency and inability of institutions to perform in the manner attributed to them is rare in the literature. The experience of various traditions in sociology shows cultural and institutional variation concerning definition, selection, collection and use of personal data. Further, discussion on globalisation suggests that standardisation and simplification are uncertain and uneven, and this may be the case for the definition and management of personal data.

b. **Personal Data in Psychological Perspective**

9. "Personal data" is not defined in psychology *per se*, although many relevant concepts can be found in the discipline, for example privacy, personal and social identity, place identity, the child's developing sense of self, and adult self concept.

10. Privacy is a concept relating to the individual, but the environment in which the individual exists is also related to the concept and privacy differs between individuals according to personal characteristics, cultural backgrounds, sex, age, and economical, educational and social backgrounds. Privacy can be considered as the regulation of the interaction between self and others and/or environmental stimuli. It includes different kinds of privacy: solitude, reserve, intimacy, anonymity, not-neighbouring, isolation, and seclusion. Place identity has strong relationship to personal identity and social identity, and is discussed in the literature of many disciplines, concerning the creation of self-identity and the interpretation of the self and its relationship with others. The significance of place within the construction of identity illustrates some of the problems that may be associated with any concept of privacy incapable of accounting for the

interaction between self and contingent environment. It provides a useful illustration of both the difficulty, and the importance, of accounting for context.

11. Scholars have not had much success in systematically analysing the conditions that nourish place identity. Interdisciplinary research indicates that place identity arises in a dialectic involving both the qualities of places and the characteristics and relations of people to places. Synthesizing the subjective with the objective dimensions of place encompasses the context of action through which individuals trace paths and institutional structures are sedimented. Because social relations are dynamic and changing, so too are places. The difficulties associated with analysing and explicating the relevant dynamics however in no way undermines their significance.

12. There is, within the discourse of place identity, a shared assumption that through personal attachment to place or geography, a person acquires a sense of belonging and purpose that gives meaning to his or her life. This affiliation or identification with place is often experienced as a sense of being 'at home'. Some view home as the crucial setting through which basic patterns of social relations are constituted and reproduced and it has been described as an emotional reference point for a sense of self. An appropriate relation to place has been described as providing the continuity and sense of personal history, and contributing toward satisfaction of the need for personal autonomy and ability to effect desired change. These needs continue throughout life, despite changes in age or life stage, or even changes in place of residence.

c. **Personal Data in Legal Perspective**

13. While there is a large literature on data protection, the interpretation of 'personal data' is seldom addressed, and the meaning of the concept tends to be assumed rather than rigorously analysed.

14. Data protection in Europe, through the Directive, is concerned with the effective operation of the single market, but studies show that this may be undermined by significant differences in interpretation of the Directive in national implementations of the Directive. This can be seen in the interpretation of many key elements of the definition of 'personal data'. For example, the distinctions between concepts of 'data' and 'information', of 'natural person' as distinguished from 'legal' or 'living' person, between data about things and data relating to individuals about things, and of concepts of 'processing', 'filing system', and 'anonymisation' all lack clarity between the European jurisdictions and make for significant differences in the operation of the data protection regime.

15. 'Relating to', central in the Directive's definition of 'personal data' has at least two possible interpretations: one that gravitates around the identification of the individual; and one that simply requires an individual's interests to be engaged. The former is made difficult by the possibility of not only 'direct' but 'indirect' identification. 'Indirect' identification, where an individual could be identified from the data or the data and other data, can only be made workable by a concept of reasonableness, as in Recital 26, but conceptually it threatens the possibility of anonymising or pseudonymising data effectively to remove it from 'personal data'.

16. The latter interpretation, data being personal data by simply concerning an individual, makes almost all data [potentially] fall within the ambit of the Directive, moreover, it prompts extraordinarily difficult questions about how such data could be prospectively identified.. This interpretation, however, is more in line with the relationship between data and the construction of personal identity as found in both the sociological and psychological literature. This requires inclusion of the way that an individual thinks about the data in the definition of 'personal data'. Indeed, the Directive allows for the inclusion of certain data as personal data simply because the data subject believes it to be so.

17. As information may affect a person's identity, so information may affect an individual's privacy: privacy, like identity, is open to many interpretations in law, sociology and psychology ('privacy' again relating to the individual's self-identity dependent upon cultural and social expectations). The change in expectations can be seen between countries and times, and is discussed in the literature in relation to tax and personal finance, sexual behaviour, and medical data.

18. The identity and privacy of the individual are traditionally seen as well protected by the anonymisation of personal data, thereby placing it beyond the scope of the Directive as non-personal data incapable of identifying the individual subject. This protection is discussed in the literature in relation to the effect and scope of pseudonymisation, anonymisation and anonymous data. The broader effect of anonymisation on dignity is, however, not widely discussed.

19. The Directive poses difficult questions concerning the relationship of privacy, identification and dignity. It also poses difficult questions about the relationship between privacy, and the freedom of expression and the freedom of information. Here, again, the literature (and the case law) is inconclusive. The definitions of what constitutes a 'relevant filing system' and 'processing' are also problematic, both being areas of wide divergence under the implementation of the Directive. However, they both relate to information once it has been defined as personal data rather than being part of the concept of personal data.

20. **Key Findings:**

    While the current literature on data protection and concepts relating to personal data in other related disciplines deal with several of the themes that are crucial in coming to an understanding of 'what are personal data?', it is clear that there is no one uncontested and coherent definition of 'personal data'. None of the issues discussed above are settled in any way.

A3: The Empirical Study of Data Protection Authorities.

21. Data Protection Authorities in 39 countries were invited to participate in the survey. 18 agreed to complete the first questionnaire and of those 11 completed the second questionnaire. For analysis purposes, these were divided into three groups: Group 1,

the eight participants in the EU; Group 2, seven jurisdictions outside the EU, but wishing to comply with the Directive for trade purposes or requiring compliance for accession; Group 3, three countries outside the EU with no requirement of compatibility. *NB. For the purposes of this report, all countries have been anonymised. Each country is referred to according to a numerical label (e.g. Country 1).*

22. Questionnaire 1 (Q1) used a variety of question techniques to assess: if it was possible to create a list of data to be seen as *always*, *sometimes* or *never* 'personal data', whether there is consistent implementation across the EU of the key terms of the Directive's definition of personal data and whether these are common terms beyond the EU, whether these terms produce practical difficulties of interpretation, how 'direct' and 'indirect' identification and 'personal filing system' is interpreted, and the impact of anonymisation upon personal data.

23. Questionnaire 2 (Q2) was designed to allow clarification by the participants of some of their responses in Q1, and to test their response to the emerging theoretical framework. This was achieved through direct questions, scenarios, and statements from the framework for discussion.

Questionnaire 1 Results

Formal key terms within the Directive 95/46/EC definition of 'personal data'.

24. Group 1 countries have generally adopted the same terms found in the Directive's definition of 'personal data' straight into national law. Thus, 'personal data', 'data', 'information', 'identified or identifiable', 'natural person', 'directly or indirectly' are terms common throughout the EU jurisdictions. While 'relating to' is not common throughout, its concept is found in all jurisdictions. The terms, however, are not often further defined by the jurisdictions. The Authorities do not report this to be causing practical interpretation difficulties, with the exception of the term 'relating to'.

25. Group 2 countries all use and define 'personal data' or 'personal information' and some report difficulties in the interpretation of this. With the exception of Country 20, all use 'relating to' and 'identified or identifiable' without interpretational difficulty. The jurisdictions define the scope of personal data as relating to 'living data subjects', 'living individual', or by reference to their Civil Codes, although three jurisdictions use the term 'natural person' without further definition.

26. In Group 3, all three jurisdictions have a central concept of 'personal information' or 'personal data', and this is defined in their legislation. One has a formal definition of 'data' and 'information'; two reported difficulty in defining those terms. 'Relating to' was not used, and two Authorities saw difficulty in mapping the concept of the same onto their national concepts. 'Natural person' is not used, neither is 'directly or indirectly' identifiable, the jurisdictions talking about personal information being 'about' an individual, the individual being identifiable or 'apparent'. One of the jurisdictions uses a completely different set of definitions and criteria. However, there is much conceptual similarity between the jurisdictions.

27. **Key Findings:**

- **Between the jurisdictions surveyed there is confidence in understanding the terms found in the Directive, demonstrated by a lack of need for definition or by a lack of difficulty in defining or interpreting the terms. There is a large degree of similarity in defining 'personal data', with consistency in the use of terminology.**

- **Despite the 'on paper' similarity of definitions discovered, Data Protection Authorities demonstrate a remarkable lack of consistency in their approaches to the classification of data types as 'personal data'. These divergences in approach are to be found both within and outside the EU.**

Personal Data Filing Systems

28. Card indexes, electronic databases, electoral registers, registers of births/marriages/deaths, membership lists of voluntary organisations, and telephone directories were more likely to be classed as *always* being 'personal filing systems'. Most agreed that organisational filing systems, photo albums, diaries, archived minutes of meetings, CCTV footage, and organisational websites would only be personal filing systems in certain circumstances. The majority agreed that a newspaper is highly unlikely to be classed as a 'personal filing system'.

29. Group 1 take a different approach from Group 2 with regard to the operationalisation of the concept of 'personal filing system'. Group 1 tend to take a more consistent approach, with most of the disagreement being attributable to the *always* vs. *sometimes* division. In contrast, the Group 2 tend to take a more diverse approach and are more likely to operationalise the concept in different ways.

The Concepts of Personal Data evident from Q1

30. Some of the differences and apparent inconsistencies between the answers and approaches given in the Q1 questions discussed above could be explained by different interpretations between the respondents about whether the putative data subject had already been identified. However, the degree of difference and apparent inconsistency suggests that there are more fundamental differences as well. This could be attributable to the ambiguity and lack of detailed definition in key terms seen as central to building the concept of 'personal data'. This does not necessarily lead to a conclusion that there is confusion in the concept, as each jurisdiction could operate with a different but internally coherent concept of 'personal data'. This was tested by asking the different countries to explain the *circumstances* in which different data types would be considered *always*, *sometimes*, or *never* personal data, and through the comments made by the respondents to the scenarios and statements in Q2.

31. This analysis shows that the countries have slightly different concepts of what constitutes 'personal data'. Three positions clearly emerge: identificatory potential as a prerequisite; a relationship other than one of identificatory potential as prerequisite; and, identification and effect as prerequisites.

Identificatory Potential as Prerequisite

32. Some countries agree with the statement that 'information can only be personal data if it can identify an individual'. They isolate and privilege a relationship of identification between information and the data subject and then use the presence or absence of that relationship in classifying data as personal data. Thus, if data is anonymised it is no longer personal data as it ceases to identify an individual. That some countries agree with this concept of personal data, however, does not produce the same classification of particular data types as *always*, *sometimes* or *never* personal data. This could be, in part, a matter of using different operational concepts within the broad concept to define how data identifies the individual (for example, using the concept of possibility of absolute anonymisation of data or the concept of likelihood of the data identifying the person in the circumstances): whether an authority locates individual cases according to the theoretical, or the actual, possibility of identification produces very different classifications of particular data.

33. The informational context within which the information is located is considered key by some data protection authorities: the physical context of information may have much less significance. This informational context may include chronological context – holding the data at the same time as other data that produces an identification of the individual. The significance of context led a number of countries to support the view that making a definitive list of "personal data" items was impossible.

34. A number of countries that supported the statement that the making of a definitive list was impossible also listed certain types of data as always personal data. This, perhaps, shows the need to create strong practical guidance from strong conceptual positions. The apparent contradiction may perhaps be reconciled within a realisation of the need for practical guidance amongst theoretical uncertainty but such reasoning was rarely provided expressly.

A Relationship other than one of Identificatory Potential as Prerequisite

35. Here, while "personal data" must relate to an identifiable individual, the data itself might not have to identify the individual but simply relate to the individual in some other way: it could affect, or be linked to, an individual in a way other than identifying them. This position, however, does not seem to be consistently maintained by those countries that support it: identificatory potential seems to continual resurface as a relevant consideration. This is, perhaps, a sensible position, as to hold that all information *relating to* an individual is personal data runs the risk of making all data personal data and thus requiring no distinction between *data* and *personal data*. However, there remains a position that non-identifying data that can be linked to identifying data is personal data, employing a concept that the data is "about" the individual (a relevant link to the data subject). It remains unclear however how such a link may be made.

36. Again the concept of context often appears important in determining the relevance of the data to the data subject (although this position is not universally held). The concept that personal data is data that "concerns" or "affects" an individual appears, *prima facie,* much more inclusive than the identificatory potential concept, and registers many more data types as *always* personal data within responses to the questionnaires. There is however some reason to think that, while the data might not need to contribute to the

identification of the individual in practice, were it not possible for it to do so in principle, it may be difficult to sustain the claim that the information is 'personal'.

Identification and Effect as Prerequisites

37. In this third concept, only data that both clearly identifies and also clearly affects individuals in other ways are capable of being personal data. The effect produced should be directly on the individual's *privacy*. The countries all indicated that they are following the *Durant* decision, where identification of the individual was not sufficient rather the data had to be focussed on the individual affecting the person's privacy. Data protection authorities that agreed with this concept suggested that classifying data types as always, sometimes or never personal data was difficult but not impossible. However, the application of the concept to particular data types produced many qualifications relating to the purpose of the use of the data with a strong importance placed upon identification. It also seemed to produce rather unexplainable classifications.

38. Given this, *always* classifications may indicate a presupposition of context within which to make the classification. Interestingly, in this concept compared with the previous concept, there was greater spread of classification of the *always*, *sometimes*, and *never*, categories. A conclusion could be that while the countries have strong conceptual understandings, abstract classifications of particular types (list making) is not easy and does not produce consistent classifications.

What do we learn from the results of the survey?

39. The results of the literature survey, Questionnaire 1 and Questionnaire 2 seem to converge. There is very little consistent underlying understanding of the concepts, causing a lack of clarity of the concept of personal data both within and outside the EU, between countries adopting similar conceptual frameworks, and within individual countries. However, the conceptual uncertainty is not often noted by the respondents as a matter of concern.

40. Three underlying concepts are frequently being applied to answer the question "what are personal data?" These are the capacity of the data to: identify an individual; affect an individual; and, identify and affect an individual. However, these concepts appear under-developed and applied inter-changeably. This matters particularly within the EU as the Directive seeks to produce a harmonised system for data protection and the inconsistent application of different concepts of "personal data" threaten this intention. It also produces an uncertain environment for data subjects and data controllers making planning and prediction of decisions very difficult. This confirms the need to develop a robust, express, theoretical and defensible framework within which the concept of "personal data" can begin to be understood.

Part C – Developing a theoretical framework to inform an understanding of the term "personal data"

41. The theoretical framework is developed through a series of "ideal types", drawing on the observations and conclusions from the empirical and literature work of this study – ideal types inspired by practice. These are illustrative of alternative approaches towards understanding the term "personal data".

42. The development of a theoretical framework pre-supposes a point of departure and a method of proceeding. The integrity of the framework depends on the choice of these two elements, and an error-free application of the method in constructing each ideal type. A justifiable starting point for this project would seem to be the formal definitions in the Directive and the usage of the data protection authorities' definitions.

43. The ideal types are developed by isolating the differences in approach shown through the responses to the two project questionnaires. They do not draw distinctions that were not evident within the responses, for example no distinction was clear between "direct" and "indirect" identification, "data" and "information", or "anonymisation". They all, to some extent, have to take into account and accommodate latent ambiguities within the ideal type: does the type concern the *actual* or the *possible*?

44. Data protection authorities appear to operate under two general conceptualisations of personal data: the "context independent" concept and the "context dependent" concept. The former concept allows for a list to be drawn of data types that are *always* or *never* personal data: the latter produces a list of *sometimes*, where all data could be personal data in the right circumstances. The two concepts can be subdivided each into two further conceptual variations:

'Unique Identifier' (context independent) Model
Personal Data is data which may be uniquely related to an individual. Due to the uniqueness of the data, it is impossible for it to be anonymised in such a way as to render it impossible for it to continue to be related to an identifiable person. Context is irrelevant.

'Context Independent Affects' Model
Personal Data is data which is capable of affecting an individual in a relevant way. It is possible to anticipate whether data of particular types will affect an individual in a relevant way without taking account of context.

'Context Dependent Identifier' Model
Personal Data is data which may identify an individual. All data is capable of being personal data, as any data is capable of identifying an individual in the right circumstances.

'Context Dependent Affects' Model
Personal Data is data which may affect an individual in a relevant way. All data is capable of being personal data, as any data is capable of affecting an individual in a relevant way in the right circumstances.

'**Unique Identifier' (context independent) Model** – *the impossibility/improbability of anonymisation*

45. Here the data is "personal" because it links directly to an individual without reference to any other information. Where additional information is required, it is not in a pre-existing database of knowledge, but is created entirely *ab initio*. The data yields identifying information without reference to a pre-existing context. Few data types fit into this model, however DNA may be a notable and useful exception.

46. DNA sequences are unique to the individual (with the exception of identical twins) and a DNA profile, practically unique, can be created from an analysis of the sequence, allowing the theoretical possibility of matching the sample to the individual. In actuality however, given the resource implications associated with an entirely *ab initio* search (which are prohibitive), the practical necessity remains for a reference to additional information i.e. an existing "interpretative framework" (such as may be provided by an existing DNA database). The "interpretative framework" links the various samples and target material through information extrinsic to the sample.

47. Even the DNA case, then, cannot identify the individual in a wholly context independent way. However, the DNA sample may contain all that is substantially needed to provide a context within an individual may be identified. Working from this example, an ideal type can be constructed around the relative difficulty or impossibility of anonymisation of the data. That data that resists anonymisation, and therefore contains all that is substantially needed to identify the individual, under this model is more likely to be considered "personal data".

48. When other information is required of necessity to supplement data to enable identification (i.e. data is 'attributed' to an individual by a 'database') then, not only is its position as a *unique* identifier vulnerable but also the data is more easily anonymised. In such circumstances single pieces of data will be incapable of contributing both to the relevant context *and* enabling identification. 'Unique identifiers' are more likely, in practice, to be deliberately constructed through the assemblage of 'portfolios' containing a number of pieces of data (e.g. PIN numbers *together with* plastic card numbers). The portfolio, when taken together, will produce a unique identifier.

49. The Unique Identifier Model has advantages in that it allows individuals to reliably categorise information as "personal data" in a relatively context independent fashion. The likely presence of the context acknowledged significant (i.e. the interpretative framework capable of recognising and using the identifier in question) may be estimated with a degree of accuracy and reliability that may not be associated with the other decision-making models. Its disadvantages are that it overestimates the extent to which any given piece of data may be unique independent of context and therefore, it underestimates the significance of context (as most data, and certainly all attributed identifiers, will only be unique in a given context).

'Context Dependent Identifier' Model – *the possibility/probability of identification*

50. Here, *any* information that can be used to identify an individual may be "personal data". There is no necessity for the data to be "unique" and the elements may, individually, be widely shared: it is their combined ability to point to a single individual that is crucial. It is not the uniqueness of the data *per se* that makes it personal data, rather it is the

availability of an informational context within which that data may function as the 'key' to the construction of a portfolio that operates as a unique identifier. Those using such a model may focus upon either the bare possibility, or the actual probability, of accessing such additional information.

51. The model must take a position as between the availability of an *actual* context and a *possible* context (and the likelihood of realising that possibility). Including theoretical possibility would greatly extend the amount of data qualifying as "personal data". Clearly, certain types of data inform about the individual and thereby identify the individual. Other types of data, however, may not of themselves inform about or identify the individual, but may, in a certain context enable the identification of an individual when added to the mass of information (knowledge that a number plate is issued in the UK alone cannot identify an individual, but in the context that one knows that the only individual driving a British car at a French campsite is Mr Smith will enable his identification by locating a car with a British number plate). Thus, operating a theoretical possibility creates all data as potentially enabling data. This suggests that the actual possibility in the instant case is more significant and useful.

52. While certain contexts are more likely to arise where there is an actual possibility of identifying the individual, there is extreme difficulty in accurately prejudging whether any particular piece of data will actually enable identification: whether the relevant context will present in the future. The model recognises the significance of the informational context and allows determination of what is more or less likely to enable identification by taking account of relative availability of relative contexts. However, it does not allow one to draw up a definitive list of "personal data"; it is not possible to limit the ways in which the data may relate to an individual beyond those limits imposed by the possibility of identification (and, significantly, any information may identify an individual in appropriate contexts); and if a predictive judgement is made on whether a particular piece of data will be considered "personal data" then fallible prediction must be made of whether the relevant context will be present.

'Context Independent Affects' Model – *the possibility /probability of relevant effect*

53. To be "personal data", data must be capable of affecting an identifiable person in a material way, and the notion of what is a relevant effect permits various interpretations. The favoured version amongst the respondents in this survey is that personal data is only data that affects an identifiable person's privacy. According to this model whether specific data types may have such an effect can be assessed in a context independent fashion.

54. This is problematic first because it relies upon an apparently untenable concept of privacy. It would appear to presuppose that specific types of information may affect dissimilar individuals' privacy in similar ways. This appears inconsistent with the concept of privacy familiar within either the sociological, psychological, or even, the legal perspective. However, a more generous reading of this Model could recognise that, while what constitutes 'privacy' may be a product of an interaction with social context, it may nevertheless be possible, *given a particular social context*, to anticipate a particular data type's ability to affect an individual's privacy. Thus, while not ignoring context entirely, it would dramatically reduce its everyday significance. There are however further difficulties in determining which data are attributed such significance by

society, the relationship of the individual to society, and to which society the individual relates. This suggests that such a context independent model is highly problematic, although, it should be noted that if the difficulties can be overcome the model may protect an individual's privacy more effectively than a model centred solely upon the notion of identification.

55. The model potentially allows the construction of a list of data types that will *always* or *never* be personal data within a specific social context and offers more specific protection of an individual's privacy than an ideal type centred simply on a notion of identification. However, it fails to explicitly recognise that context considerations are to some extent unavoidable when assessing an individual's privacy; it again has the potential to include all data; it relies upon judgments about the likelihood of a particular piece of information having an effect upon individuals which may prove incorrect in specific cases.

'Context Dependent Affects' Model – *the possibility/probability of relevant effect*

56. Here "personal data" is data that is capable of impacting upon an individual's privacy, determined by contingent circumstances in each case: it is necessary to take into account the specific context as that determines the meaning and value of any specific data for an individual's privacy. It is more consistent with the sociological and psychological literature, recognising privacy as an interaction between an individual and others (and/or environment). Again the difficulty of whether the type concerns the actual or theoretical is significant as the latter again produces an inclusion within "personal data" of all data as potentially affecting an individual. Here, however, a restriction of the scope of personal data to that which *actually* affects the individual's privacy may not restrict the range of information that might constitute personal data given the intensely subjective process of interaction between self and environment that could include almost any data.

57. In order to assess if the circumstances had an impact on the individual, given the subjectivity of the assessment it could only be made after the individual actually experienced them, making predictive judgement very difficult. In order to prospectively assess if privacy is likely to be affected, it may be necessary to have regard to the "likely contexts" in order to assess the likely effect that a particular piece of data will have on that privacy,. A classificatory model based around effect may then need to *assume* relatively stable contexts: taking note of what data is usually available to others within a particular context and their possible use in that context, and the various uses to which the data could be put and the impact of such uses upon the individual's privacy. However, such prediction would be inherently fallible..

58. This ideal type acknowledges the significance of context to the assessment of whether a particular piece of data will affect an individual's privacy and recognises that for an individual what information may affect his or her privacy may not be readily anticipated by another in advance. The type also acknowledges that any information that might affect an individual in a relevant way would constitute their personal data. However, it does not allow the construction of a list of data types that will *always* or *never* be "personal data" and any information might affect an individual in a relevant way given idiosyncratic vulnerabilities to the acquisition and use of information of different types.

The relation of the ideal types to one another

59. While the ideal types are extreme versions of four aspects, their elements can be combined to produce composite concepts of personal data. Countries may operate with such composite concepts, for example operating with a concept that requires personal data to both identify and affect an individual. The ideal types remain useful both as ideal types and also as providers and clarifiers of elements to such composite concepts.

The "Identifies and Affects" Concept

60. Here data must be assessed in two different ways before it can be considered "personal data": does it identify an individual and does it affect an individual? The assessment could be either dependent or independent of context, although the countries using this type in the sample operated a context dependent approach (without indicating whether they believed identification and/or effect to be context dependent variables). This shows the complexity of composite types.

61. While complexity may be a disadvantage, a more sophisticated approach may avoid some of the disadvantages associated with particular ideal types. For example, linking the Context Dependent Affect Model with a requirement that data should (actually) identify the individual might help avoid the difficulties associated with the model otherwise encompassing all data as personal data. Similar restriction may result from linking the two context dependent ideal types: to produce a concept of personal data that while context dependent was effectively limited via internal reciprocal qualification.

Conclusions:
The Significance of Ideal Type and Composite Types for Decision-making Strategies

62. From the empirical surveys and the literature review it is clear that data protection authorities do not use clearly defined concepts and models to establish which data are personal data. However, there are significant elements that are observable within the literature, in the ways that different authorities make practical decisions about the classification of data types and then in the ways that they talk about their decision-making processes. From these significant elements, a series of ideal types can be established characterising ways of defining personal data according to the weight placed on single significant elements. It is then possible to make composite types by mixing the single significant elements together. This may have the effect of reducing the disadvantages associated with the operation in practice of the ideal types. From this, by way of conclusion, the implications and possible outcomes and experiences that a data protection authority relying on one of the ideal types in decision-making will experience can be considered and the difficulties of the disadvantages of each type can be assessed. This process produces no single definition of personal data, rather it gives data protection authorities, and in particular the UK authority, a theoretical approach to resolving the conceptual question "what are personal data?" according to its own policy position.

63. The Unique Identifier type is problematic as very little data could be classed as a unique identifier. This could be solved by classifying data according to its degree-of-fit

with the ideal of a unique identifier: assessing how closely it resembles the characteristics of a unique identifier. The context within which the data resembles a unique identifier becomes very important. The result is that all data is classified on a continuum relating to how close it is to being a unique identifier according to the circumstances in which the particular data exists. Thus, the same data type can move position on the continuum according to changing contexts. The higher incidence of classifications close to unique identifier may strengthen a presumption that the data type itself is "personal data", however, where this line should be drawn is a challenge to each data protection authority.

64. In order to draw the line and use the unique identifier concept, a data protection authority must establish three things: the relevant context within which to judge the "uniqueness" of a piece of data; the appropriate height at which to set the "bar" of uniqueness; and, whether a specific piece of data is sufficiently unique within that context to "clear the bar". Invariably, data operates as a unique identifier because, within the context, it enjoys a unique status and functions as an identifier. This influences profoundly the distinction between "direct" and "indirect" identification. The distinction no longer lies in the data itself, rather it resides in the context within which the data is perceived. Whether a specific data will enable identification either directly or indirectly depends upon whether sufficient information is already present to attribute the identifier with "unique status" and to "directly" enable identification.

65. If the context is included in this model, then the authority must also establish whether it is simply the existence or the accessibility of the additional relevant or necessary information that is significant. If accessibility is crucial, then the question is "who must be able to access?"

Context Independent Affects

66. The Context Independent Affects type has similar problems. The list of personal data is difficult to create even when "effect" is narrowed to "effect on privacy" as the type depends upon a reliable prediction of the effect of particular information upon individuals' privacy and this is difficult to achieve in advance. By taking into account social context and removing particular contingencies and individual circumstances, and effectively limiting context, the calculation is easier to make.

67. Defining an individual's society and the effects of that society on information and its relationship to the individual's privacy is difficult, and this can stretch the possibility of applying the model. The need to recognise the context shows the difficulty of maintaining a coherent context independent type. Further, the model does not assist in distinguishing "direct and indirect identification" as identification is not central to this concept of personal data, yet the issue is central to the Directive.

Context Dependent Strategies

68. The Context Dependent models are not free from difficulty, however. The distinction between "actually" identifies and "could possibly" identify is problematic as the "could possibly" position leads to the potential for all data to be "personal data" in the correct context. The model therefore requires a consideration of "actual" circumstances,

making prediction difficult. Developing a middle ground where the context producing either identification or other effect on the individual would "probably" arise, then data in such a context would be "personal data". This would allow for the creation of a list of personal data, although it would be best considered indicative only.

69. Basing the definition of "personal data" on one model is problematic. Ignoring the role of context causes problems of coherence, but relying simply upon context leads to unpredictability. The composite approach may offer a solution. However, developing a strategy that mixes together a number of ideal types or creates a bespoke type may create unpredictability and offers more opportunity for internal incoherence and inconsistencies in materially similar cases. Further, the development of the bespoke type is particularly difficult and does not guarantee a trouble-free solution. The process of identifying the relevant and significant elements for a particular data protection authority's definition of "personal data" will, however, produce a reasoned (and therefore more transparent and reasoned) position that avoids the absolute unpredictability of no position.

70. The key term to understand in creating definitions of personal data within the Directive (and other legislation relying upon the term) is "relating to". It has the greatest impact on classification of data. The definitions of other terms, however, also have significant impacts upon classification. For example, "identification" which can have a variety of definitions: "handshake" identification, where an individual must be capable of being physically located to be identified; or "isolate and affect" identification, where although physical location is not necessary, identification is achieved by isolating the individual from others and deliberately targeting him or her in a particular way, for example within the electronic environment.

71. The choice of the term "identification" goes to the definition of a person. "Handshake" identification requires a natural, living person, however, "isolate and affect" identification much more easily extends to legal persons. The concept of identification could be made easier by employing a composite concept of personal data requiring identification and affect upon the privacy of an individual. This could require identification to relate only to an individual's privacy, although this does not necessarily limit the definition only to handshake identification. Again, the context and circumstances are the crucial aspect of the definition.

72. While ideal types do not by themselves provide a comprehensive understanding of "personal data", they assist in understanding the key elements and in creating a conceptually coherent definition that is justifiable and practical. Using the ideal types as a tool for the formation of a concept of personal data may aid the transparency, accountability and predictability of any data protection authority's decision making strategy and answer to the question "what are personal data?"

Part A – Introduction

The aim of the first part of this report (Part A) is threefold. A brief introduction to the aims of the study will follow. This will identify the key research questions. Then, A2 will provide a brief analysis of the literature relating to the question 'what are personal data?'. Although there is very little legal literature addressing this specific question, we have been able to draw on both international commentary on data protection issues and the multi-disciplinary literature on the concepts of identity and privacy to inform our discussion. It will become clear that the discussion has prompted a series of questions relating to both the conceptualisation and operationalisation of 'personal data'. In A3, we move on to describe the methodology of this study. This section will contain a detailed description of the methodological procedures.

A1: An Introduction to the Study

The main of this project was to inform a robust, defensible understanding of the term 'personal data'. In order to achieve this, a number of key research questions were identified at the start of the project:

1. Questions arising directly from Directive 95/46/EC and the Data Protection Act 1998

    - Is it possible to create a list of information that is necessarily 'personal data' (e.g. is a person's name always personal data?), or, is the classification of information as 'personal data' always dependent upon other factors (e.g. the context of its use or other information in another's possession i.e. 'triggering information')?
    - Should a distinction be drawn between living and deceased individuals: what is a 'natural person'?
    - What is an 'identifiable person'?
    - How must data 'relate' to an identifiable person for that data to be properly termed 'personal data'?
    - What is meant by the distinction between 'direct' and 'indirect' identification?
    - In what circumstances is it 'reasonably likely' to identify a person from data, such that it becomes 'personal data'?
    - 'Identified' by whom: only by the data controller or any person?
    - When does information become 'personal data' and when does it cease to be 'personal data'?
    - When is anonymisation effective in removing data from the status of 'personal data'?
    - Is 'relevant filing system' sufficiently clear to create a clear boundary of 'personal data'?
    - What is the significance, if any, of any distinction that may be drawn between 'data' and 'information'?
    - What is the significance of the different types of identity (as listed in the Directive 95/46/EC)?

- Is the distinction between 'personal data' and information contained within 'personal data' sufficiently clear and appropriate?

2. Questions about the definition of 'personal data' in other contexts

    - Is personal data defined effectively in other legal contexts or jurisdictions?
    - Do other disciplines that show an understanding of individuality and personal identity within the individual and about the individual within society produce coherent definitions of 'personal data'?
    - Is there a logical definition of 'personal data' that derives from an understanding of the relationships between 'information', 'data', 'knowledge' and 'understanding'?

3. Questions about adjudicating between definitions and disciplinary contexts

    - How can one adjudicate between different contexts of understanding of 'personal data'?
    - What is the importance of understanding the context within which each definition is made?
    - Are the definitions that may be found from other jurisdictions and disciplines useful for defining 'personal data' within the context of the Directive if they do not share the same purpose as the Directive?
    - How can an understanding of the competing definitions of 'personal data' be useful in explaining the Directive's necessary interpretation of 'personal data' for its purposes to constituencies that do not share or understand the same context or purpose?

Clearly a very important question in creating a coherent and defendable (i.e. arguable and explainable) position for the question 'what are personal data?' is:

- How clear is the purpose of the Directive within itself, within the European legislative context, within the European Convention on Human Rights, and for the wider world?

Methodology and study design

The project was competed over a period of three months (December 2003 - February 2004). The project was designed in three phases. Each phase had distinct empirical and theoretical elements which ran concurrently and converged within each phase:

Phase One:

The study began with an investigation into the understanding of the theoretical interpretations of 'personal data' that have been developed in a number of key disciplines (law, philosophy, criminology, sociology and psychology). At a roundtable discussion with the members of the Expert Panel, the conceptual development commenced. These discussions informed the development of the first questionnaire.

The first questionnaire was designed and sent to a sample of Data Protection Authorities across Europe and beyond. This asked the Authorities for:

- their opinions on the meaning of 'personal data',
- information and data that have caused difficulties for them in classification,
- their opinions on a number of pieces of information and data that the research team saw as 'trouble cases' for defining the boundaries of personal data.

- their opinion of the basis of their authority and purpose in protecting personal data and the justifications for their classification (e.g. in Europe, one would expect Authorities to reflect the purpose of the Directive in their response, but will other jurisdictions show a similar purpose for the protection of personal data?).

Phase Two:

When the responses to the first questionnaire had been returned, preliminary analysis revealed a number of key themes, as well as several conceptual inconsistencies and operational difficulties. The results were used to inform the next stage of theoretical development which was operationalised through further roundtable discussions with the Expert Panel. Following these discussions, a second questionnaire was designed to explore the differences in approach taken by different Authorities and to test out our emergent theoretical models.

Phase Three:

The responses to the second questionnaire were analysed and used to inform the final stages of the development of the theoretical framework

A2: A Review of the Literature

From the outset of the project, we prioritised the need to approach the research questions from multi-disciplinary perspectives. To restrict our investigations to a purely 'legal' focus would be to prevent an in-depth analysis of the concept of 'personal data'. Thus, this literature review is presented in three sections. We consider the literature relating to 'personal data' in three disciplines: sociology, psychology and law.

a) Personal data in the sociological perspective

Sociologically, the subject of 'personal data' can be addressed from two different, albeit related directions. The first focuses on the information that sociologists and other social scientists collect during empirical research. The second addresses wider processes of data construction in public and private sector organisations and the relationship that these have to the everyday lives of real people.

Before considering each of these, some introductory comments with respect to the notion of 'construction' - or 'social construction'[2] - are necessary. Briefly, this notion reminds us that the complexities of human life don't 'just happen'. The world of humans exists because of the work and co-ordination of humans: it is constructed and it requires construction. Second, it also reminds us of the enormous historical, cultural and local variability of that human world: different collectivities do things differently.

With respect to personal data, to talk about 'social construction' recognises the importance of process. Data are always created or made, they are not simply 'there' as naturally or objectively occurring realities. Thus data are abstractions from the facts; they are not the facts. Data are produced during social processes of definition, selection and collection. This implies, further, that there is no necessary closure of the process: data are always liable to revision and redefinition, in varying degrees.

The next point to note is that the construction of data is never disinterested. Data always have a purpose - they serve interests - and those purposes are crucial in their definition, selection and collection and in deciding what is done with and to them. *Inter alia* the interests concerned may be political, governmental, commercial, medical or scientific. As in the work of investigation agencies, for example, they may also be idiosyncratic and personal. The key point is that data cannot be understood outside of the social contexts of their construction and the interests that they are designed to serve.

Finally, data are constructed in real institutional contexts, whether these are national jurisdictions, international agencies, corporations, or whatever. These contexts may have some things in common - the maths and procedures of statistics, for example, or

[2] Berger, P. L. and Luckmann, T. (1967), *The Social Construction of Reality*, London: Allen Lane; Hacking, I. (1999) *The Social Construction of What?*, Harvard: Harvard University Press; Jenkins, R. (2004a) 'Social Construction', in A. Kuper and J. Kuper (eds.) *The Social Science Encyclopedia*, 3rd edition, London: Routledge, in press; Searle, J. R. (1995) *The Construction of Social Reality*, London: Allen Lane.

computing software and hardware - but they also have their own histories and cultures and face their own present contingencies. Interests aside, therefore, factors such as these will also differentiate 'what counts as data' in one place from 'what counts' in another. Uniformity is not to be expected.

Like other forms of systematic inquiry, empirical social research proceeds via the construction of appropriate data. All social research data are personal data: in the first instance they are about or otherwise connected to real persons, whether alive or dead. The 'in the first instance' is important. While there is a world of difference between the detailed accounts of individual lives that are the basic data of ethnography and the apparently impersonal aggregate data used in international comparative research, each depends, somewhere down the line, on the collection of information, either from specific individuals or about their behaviour. In principle, therefore, there is in all social research a paper trail - or its electronic equivalent - which should, given the availability of sufficient contextual information, enable data to be tracked back to an identifiable person or persons. Without this potential court of last resort, the truth claims of social research, such as they are, could not be contemplated.

At which point, however, the 'in principle' becomes significant. In practice, while much social research deals in data that derive from information that is already in the public domain - politicians' speeches, policy documents and the outputs of various media, for example - much, and perhaps much more, social research actually depends to a significant degree precisely on the anonymity of its data. There are two main reasons for this. Ethically, based on the implied dictum of 'doing no harm', respondents - who supply the information that becomes data - deserve to have their privacy protected as a basic right[3]. This is increasingly a matter of observing formalised research governance protocols. No less important, however, is the more self-interested epistemological argument that informants cannot themselves be expected to tell the truth - as they see it - without some faith that the confidentiality of the research process, and the privacy of their testimony, can be guaranteed.

As a result, various approaches to anonymising data are integral to the data construction processes of social research. In large-scale surveys, for example, data is either collected anonymously at source or is subsequently 'cleaned' of potentially identificatory information. More problematically, in small-scale qualitative studies, not only are pseudonyms created, but empirical details may also be falsified and/or tactically - and tacitly and tactfully - written out of the account, in order to create at least a legal fiction of anonymity. The UK social science data archives at the University of Essex - the ESRC Survey Archive[4] and Qualidata[5] - have developed expertise in the anonymising of original social research material of all kinds, in order to make it available for secondary analysis. Even in the Essex archives original data are not available to others, for source verification.
Due to the constraints set by the personal nature of its data, the integrity of the social research enterprise is reliant on the validity and reliability of its research procedures, on the one hand, and professional standards, on the other (which is where research governance is becoming increasingly important). In all cases, a degree of trust in the

[3] Bryman, A. (2001) *Social Research Methods*, Oxford: Oxford University Press at pp. 475-86
[4] http://www.data-archive.ac.uk/home/
[5] http://www.esds.ac.uk/qualidata/

probity of the researcher(s) and the secure storage of raw data is called for. In all cases, too, what is significantly 'personal' and what is not are matters of judgement: name and address might seem fairly straightforward, but beyond that context is all, and there is a limit - which is itself always contextual - to the degree to which potentially identificatory context can successfully be disguised.

Many of these issues about the data construction process in social research apply to personal data of whatever sort, in whatever institutional contexts. The key issue is the link between data and identifiable persons: what is 'personal' and what is not? This is the difficult issue that informs the present attempt to determine whether or not, 'out there', a minimal consensual definition of 'personal data' can be found.

The other sociological perspective on personal data sends us off in a different direction, and raises some of the other issues summarised in the introductory remarks about social construction: first, the purposes for which data are collected and the uses to which they are put, and second, cultural and institutional similarities and differences in the definition, selection, collection and uses of personal data.

With respect to the purposes and uses of data - the interests that it serves - the current concern with the regulation and governance of personal data can be addressed from within long-standing and convergent sociological concerns about aspects of modernity such as bureaucracy and formal organisations, the state, social control, and surveillance. Beginning with Max Weber's famous description in 1905 of modern organisational rationality and bureaucracy as an 'iron cage'[6], progressing through Michel Foucault's use of the image of Bentham's Panopticon as an analogy for modern 'disciplinary' society[7], to more recent concerns about the 'network society' created by the Internet[8] and the 'surveillance society' facilitated by new technology[9], there is an extensive discussion within sociology and allied disciplines of the increasing identifiability and accountability of individuals in modernity. The underlying theme is that these developments, dependent upon ever more efficient and extensive systems for collecting and processing personal data, serve the interests of the powerful, whether in the state or the private sector. Many of the substantive topics within this broad sociological discourse are concerned with identification in one or other of its manifestations: the role of the census of population in creating social categories[10], for example, or the administrative allocation to individuals of public resources and penalties. [11]

This critical social science discourse reflects and informs public concern about personal data and its management (a public mood within which the role of the Information Commissioner and, indeed, the present exercise can be located). However, there are other sides to the matter that are sociologically less prominent (and which return us to issues of process). The positive side of the balance sheet - or at least the question of how

[6] Weber, M. (1976) *The Protestant Ethic and the Spirit of Capitalism*, London: George Allen and Unwin at p. 181.
[7] Foucault, M. (1977) *Discipline and Punish: The Birth of the Prison*, London: Allen Lane at pp. 195-228
[8] Castells, M. (1996) *The Rise of the Network Society*, Malden, Mass.: Blackwell.
[9] Lyon, D. (ed.) (2003) *Surveillance as Social Sorting*, London: Routledge.
[10] Kertzer, D. I. and Arel, D. (eds.) (2002) *Census and Identity: The Politics of Race, Ethnicity and Language in National Censuses*, Cambridge: Cambridge University Press.
[11] Jenkins, R. (2004b) *Social Identity*, 2nd edition, London: Routledge, in press. at p. 164-8

one could possibly run a modern society without bureaucratic management and control - is, for example, rarely addressed. What's more, the efficiency and reach of the 'iron cage', the Panopticon and the 'information society' are overestimated: individuals routinely resist and subvert the formal objectives and procedures of bureaucracies; formal procedures beget their informal alternatives; organisational size, complexity and extent create spaces within which monitoring can be evaded and make procedures more difficult to enforce; the individuals on whom systems depend are often incompetent, lazy or disaffected; and the many 'irrational' dimensions of everyday human life are ubiquitous and significant within organisations. [12]

Omnipresent and omnipotent bureaucratic rationality is not only unlikely; it is impossible to imagine. As recent tragic events such as the Soham case have illustrated, policing the population efficiently via personal data remains an aspiration rather than a reality. Perhaps paradoxically, cases such as these throw a harsh and perplexing light on the civil liberties arguments about personal data, and may be used as a further argument for more effective data management.

The cultural and institutional variability with respect to the definition, selection, collection and use of personal data is one of the issues that this report is attempting to understand better. Not least because of its undeniable difficulty, the thoroughgoing comparative study of phenomena such as 'personal data' is, as yet, still in its infancy: Hofstede[13] is one pioneering example from the management studies literature, the social policy literature, although helpful, tends to be couched in generalities rather than detail, while anthropology has long since eschewed serious comparative analysis as an epistemological mind-field.

One area of discussion that may have something to offer in this respect is globalisation. Ritzer's famous 'McDonaldisation' thesis[14], uses the fast food industry as a metaphor for the cultural and bureaucratic standardisation that is attendant upon the global liberalisation of markets. [15] It is useful here because it suggests that the standardisation and simplification of process and product are simultaneously a response to globalisation and one of its drivers. The same is probably true with respect to global information systems and technologies, in this context personal data management: while hardware and software are important factors, the imperatives of globalisation, such as they are, should not be underemphasised.

'Such as they are' is an important qualification in the above, however. One of the consistent themes in the discussion of 'McDonaldisation' is that simplification and, especially, standardisation are, in fact, uncertain and uneven: there is resistance[16] and local ways of doing things have a way of infiltrating and colouring the process. [17] More generally, 'glocalisation' - the assertion of local distinctiveness - is every bit as much a

[12] *Ibid* at p. 179-83
[13] Hofstede, G. (2001) *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organisations Across Nations*, ThoCountry 42nd Oaks: Sage.
[14] Ritzer, G. (2000) The McDonaldization of Society: An Investigation Into the Changing Character of Contemporary Social Life, second edition, ThoCountry 42nd Oaks: Pine Forge.
[15] Although it is often overlooked that Ritzer's analysis has many antecedents in the literature on the capitalist labour process: see, for example, Braverman, H. (1974) *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*, New York: Monthly Review Press.
[16] Smart, B. (1999) *Resisting McDonaldization*, London: Sage.
[17] Watson, J. L. (ed.) (1998) *Golden Arches East: McDonald's in East Asia*, Palo Alto: Stanford University Press.

consequence of the complexities of globalisation as standardisation. [18] This is likely to be the case for the definition and management of personal data.

b) Personal data in the psychological perspective

"Personal Data" would not be found in the index of the typical textbook of psychology, but many concepts which can be seen as relevant to legal discussions of the concept would be found, allied to some burgeoning fields of research. Examples of such concepts would include: privacy, personal and social identity, place identity, the child's developing sense of self, and adult self concept.

Privacy

Although privacy is the condition of the individual[19], the environment in which the individual exists is also related to the concept of privacy[20]. The definition of privacy varies for each individual due to different personal characteristics, cultural backgrounds, sex, age, economical, educational and social backgrounds[21]. In brief, privacy can be considered as the regulation of the interaction between the self and others and/or environmental stimuli[22]. The most basic need for privacy can be stated as the optimization of social contact with both in-coming and outgoing information and avoiding unwanted crowding within the environment[23]. Before now, the concept of privacy in psychology literature was considered to be one-dimensional[24]. In Westin's[25] theoretical analysis on the functions of privacy, it was suggested that there are four different kinds of privacy. These four kinds are solitude, reserve, intimacy, and anonymity. Afterwards, Marshall[26] empirically determined Westin's four states of privacy and found two additional states: not-neighbouring, and seclusion.

---

[18] Robertson, R. (1992) *Globalization: Social theory and Global Culture*, London: Sage at p. 173

[19] Chapin, F. S. (1951) "Some housing factors related to mental hygiene" *Journal of Social Issues*, 7, 164-171.; Westin, A. F. (1967) *Privacy and Freedom*, New York: Atheneum; Weiss, P. (1983) *Privacy*. Carbondale, IL: Southem Illinois University Press; Schoemen, F. D. (1984) "Privacy: Philosophical dimensions of the literature", in F. D. Schoeman, (Ed.), *Philosophical Dimensions of Privacy: An Anthology* Cambridge: Cambridge University Press, pp. 1-33; Gavison, R. (1984) Privacy and the limits of law", in F. D. Shoeman, (Ed.), *Philosophical Dimensions of Privacy: An Anthology*, Cambridge: Cambridge University Press, pp. 346-402; Newell, P. B. (1998) "A cross-cultural comparison of privacy definitions and functions: A system approach" *Journal of Environmental Psychology*, 18, 357-371.

[20] Chermaye¡, S. & Alexander, N. Y. (1963) *Community and Privacy: Toward a New Architecture of Humanism*, New York: Doubleday; Hall, T. (1969) *The Hidden Dimension* New York: Doubleday Co.; Canter, D. & Canter, S. (1971) "Close together in Tokyo" *Design and Environment*, 2, 61-63.;Duvall-Early, K. & Benedict, J. D. (1992) "The relationships between privacy and different components of job satisfaction" *Environment and Behaviour*, 24, 670-679.

[21] Altman, I. (1975) *The Environmental and Social Behavior* Monterey, CA: Brooks/Cole; Altman, (1976) "A conceptual analysis" *Environment and Behavior*, 8, 7-29; Altman, I. (1977) "Privacy regulation: Culturally universal or culturally specific" *Journal of Social Issues*, 33, 66-84.; Newell, P. B. (1994) "A system model of Privacy" *Journal of Environmental Psychology*, 14, 65-78; Newell, P. B. (1995) "Perspectives on privacy" *Journal of Environmental Psychology*, 15, 87-104; Newell, P. B. (1998) "A cross-cultural comparison of privacy definitions and functions: A system approach" *Journal of Environmental Psychology*, 18, 357-371.

[22] Pedersen, D. M. (1997) "Psychological functions of privacy" *Journal of Environmental Psychology*, 17, 147-156; Kupritz, V. W. (1998) "Privacy in the workplace: The impact of building design" *Journal of Environmental Psychology*, 18, 341-356; Newell, 1998, *ibid*.

[23] Altman, 1975; Kupritz, 1998

[24] Pedersen, D. M. (1987) "Sex differences in privacy preferences" *Perceptual and Motor Skills*, 64, 1239-1242.

[25] 1967

[26] Marshall, N. J. (1974) "Dimensions of privacy preferences" *Multivariate Behavioral Research*, 9, 255-272.

Later, Pedersen[27] stated six states of privacy as a result of an empirical study. In Pedersen's classification, seclusion and not-neighbouring are not considered, and instead, like Westin's classifications, he defines solitude, reserve, anonymity, and intimacy. Differing from Westin's classification, Pedersen[28] extended Westin's research and found isolation, which was similar to solitude but more strict, and he divided Westin's intimacy into two: intimacy with friends and intimacy with family. Solitude is the condition of being alone and unobserved by others, and it is a condition which is either desirable or neutral. In solitude there is no need to be geographically removed from others. For Pastalan, the distinguishing characteristics of solitude were solitariness and physical isolation[29].

Personal identity and social identity, and their relationship to place identity

The concern with the construction of self in the world and its connectedness to place and the environment is reflected in the growing literature on place and identity. Although places and their attendant meanings contribute to identity in complex ways, previous research on place identity has typically focused on two broad functions: display and affiliation. With regard to place identity as display, researchers have documented how people use places to communicate qualities of the self to self or the other; places may be integrally involved in the construction of both personal identities--unique configurations of life history items that differentiate the self from the other--and social identities--groups of attributes associated with persons of a given social category[30]. Scholars have also explored how people use places to forge a sense of affiliation through attachment[31]. Such an identification with a place often involves emotional ties to place, but it may also involve a sense of shared interests and values. Home provides the opportunity for both display and affiliation functions. Scholarship on place identity has been forthcoming in the fields of sociology, psychology, architecture and geography. Each discipline has a unique perspective on what place identity is. Environmental psychologists base their definition of place identity on cognition, or the action of knowing or consciousness:

...a sub-structure of the self-identity of the person consisting of, broadly conceived, cognitions about the physical world in which the individual lives. These cognitions represent memories, ideas, feelings, attitudes, values, preferences, meanings, and conceptions of behaviour and experience which relate to the variety and complexity of physical settings that define the day-to-day existence of every human being. At the core of such physical environment-related cognitions is the "environmental past" of the person; a past consisting of places, spaces and their properties which have served instrumentally in the satisfaction of the person's biological, psychological, social and cultural needs.[32]

Sociologists define place identity as an interpretation of self:

...that uses environmental meaning to symbolize or situate identity. Like other forms of identity, place identity answers the question - Who am I? - by countering - Where am I? or

[27] Pedersen, D. M. (1979) "Dimensions of privacy" *Perceptual and Motor Skills*, May, 1291-1297.
[28] *Ibid*.
[29] Newell, 1995
[30] Goffman, E. (1963) *Stigma* New York: Anchor Books.
[31] Altman, I., & Low, S. (eds) (1992) *Place Attachment* New York: Plenum Press.
[32] Proshansky, H. M., Fabian, A. K., & Kaminoff, R. (1983) "Place Identity", *Journal of Environmental Psychology*, 3, 57-83 at p. 59.

Where do I belong? From a social psychological perspective, place identities are thought to arise because places, as bounded locales imbued with personal, social, and cultural meanings, provide a significant framework in which identity is constructed, maintained, and transformed. Like people, things, and activities, places are an integral part of the social world of everyday life; as such, they become important mechanisms through which identity is defined and situated.[33]

Although the connection of place and identity has received the greatest attention outside geography, the geographical work is very diverse. Related to the work done in psychology and sociology, psychoanalytic studies of place and space have been produced by those claiming to practice "psychogeography". Political, cultural, social and health geographers[34] have been slowly taking part in the examination of place identity as well. For example, Cutchin[35] very thoroughly explores place identity in his theory of experiential place integration, which is an "active developmental process based on the enhancement of security, freedom and identity and meaning in place". Identity in this case is defined as the coherence of a self in its relation to another person, social group, community or environment:

People gain identity in conjunction with the physical and socio-cultural aspects of place. At the same time self identity is constructed by interactions in place, the identity of the place for selves is remade.... Identity is developed in and attached to place. Development of identity is not automatic...actions, roles and responsibilities create identity in a continuous fashion.[36]

Clearly, place identities affiliate the self with significant locales, bringing a sense of belonging and order to one's socio-spatial world[37]. Scholars have not had much success in systematically analysing the conditions that nourish place identity. Interdisciplinary research indicates that place identity arises in a dialectic involving both the qualities of places and the characteristics and relations of people to places. Synthesizing the subjective with the objective dimensions of place encompasses the context of action through which individuals trace paths and institutional structures are sedimented. Within this definition, actors with histories and goals, hopes and fears behave within the given, complex set of paths and projects, deriving a concept of place as a "historically contingent process" in which individual and institutional practices - in their reflexive relationship - create and recreate places[38]. Here the focus is very much on dialectical process described in structuration theory, the everyday shaping and reproduction of human agents and social structures[39]. In a similar fashion, Massey states, "a `place' is formed out of the particular set of social relations which interact in a particular location". Because social relations are dynamic and changing, so too are places. Pragmatism takes this dynamic further by arguing that place is where several levels of action occur together simultaneously; there are large-scale events that envelope us and often constrain us, interpersonal or group

[33] Cuba, L. J., & Hummon, D. M. (1993) "Constructing a sense of home: Place affiliation and migration across the life cycle" *Sociological Forum* 8, 547-570, at p. 112
[34] Cutchin, M. P. (1997) "Physician Retention in Rural Communities" *Health and Place* 3, 25-41.
[35] *Ibid*, at p. 39
[36] *Ibid*, at p. 21
[37] Relph, E. (1976) *Place and Placelessness*, London: Pion.
[38] Pred, A. (1984) *Place, Practice and Structure*, Cambridge: Polity Press.
[39] Giddens, A. (1979)

interactions, and relatively autonomous self-actions. The interlocking and continuous set of actions creates a situation of emerging experience. Human action is not simply in response to institutions, norms and persons - it is completely embedded in them in the ebb and flow of place-based events. This view of place is neither subjective nor objective - it merges "both sides" of place by training attention on the action between elements sharing a locale. Place is deeply woven into human experience, and place identity is understood to be the construction of self in the world and its connectedness to place and the environment. Geographical theorists have employed a number of concepts associated with place identity. Those most notably are the humanistic geographers Tuan[40], Relph[41], and Buttimer[42]. Although each of them defines "place" somewhat differently, two of their underlying assumptions are shared. The first assumption that is basic to the work of these humanistic geographers is that this sense of "rootedness" or "centeredness" is an unselfconscious state. In describing the essence of place, Relph states that:

"The essence of place lies in the largely unselfconscious intentionality that defines places as centers of human existence. There is for virtually everyone a deep association with and consciousness of the places where we were born and grew up, where we live now, or where we have had particularly moving experiences. This association seems to constitute a vital source of both individual and cultural identity and security."[43]

Similar to the complexity of examining other human affective ties to the material environment[44], the characterization of place identity as an unselfconscious state by these theorists creates some problems; the perspective implies that place identity in its full meaning cannot be described or communicated. A thorough description or communication of place identity may be through the absence of experiencing it, as suggested by Relph's[45] notion of "placelessness", and Fried's[46] concept of "grieving for a lost home". The second shared assumption is that through personal attachment to place or geography, a person acquires a sense of belonging and purpose that give meaning to his or her life. This affiliation or identification with place is often experienced as a sense of being "at home"--of being comfortable, familiar, and "really me" here[47]. Without exception, the home is considered to be the "place" of greatest personal significance in one's life - "the central reference point of human existence"[48]. According to Buttimer, place identity, or the sense of belonging, is a function of the degree to which the activities important to a person's life are centred in and around the home. Buttimer implies that a particular balance between "home" and the surrounding geography, or "horizons of reach" is necessary for the maintenance of self-identity and emotional well-being[49]. These ideas are reiterated in the literature on the meaning of home. Home is central in the lives of most people. Most agree that the idea of home involves more than where one resides. Nevertheless, in some studies, the concept has been defined by a single characteristic such as size, individual possessions, or informal unstructured atmosphere. The more accepted view, however, is

[40] 1980
[41] 1976
[42] 1980
[43] Relph, 1976, p. 43
[44] Tuan, Y.F. (1974) *Topophilia* Englewood Cliffs: Prentice Hall.
[45] (1976)
[46] (1963)
[47] Relph, 1976; Seamon, D. (1979) *A Geography of the Life World*, New York: St Martins.
[48] (Relph, 1976, p. 20)
[49] Buttimer, 1980

that the concept can only be understood in terms of its many cultural and psychological aspects[50]. Examining the home as a site of shared symbolic meaning[51] draws upon the humanistic ideas that reconstitute landscape. In addition to being a geographical location, home is also the crucial setting through which basic patterns of social relations are constituted and reproduced[52]. As home is usually the foremost place in peoples' lives, home has received considerable attention[53], but only recently has this attention been in connection with health care. Home is deemed a non-traditional health care setting[54], where health-promoting properties represent focal centres for unique healing properties and reputations found in, but not out of, place. This approach is informed by and draws together a number of strands in the recent work on the relationships between healing spaces (where the focus on place involves an interest in the context of an experienced place) and the broader processes of health care restructuring. Those environments that bring about a strong positive sense of place for individuals can also be described as authentic landscapes, just as those associated with placelessness is described as unauthentic[55]. Meaning, value and experience are found in those environments that have a strong sense of place. Sense of place defines the identity, significance, meaning, intention, and the felt value that are given to places by individuals[56] as a result of experiencing it over time[57]. As to the best way in meeting the purpose of making the link between a sense of place and health, Kearns and Gesler[58] suggest that Eyles[59] makes the most useful articulation of place:

"Sense of place, he proposed, is an interactive relationship between daily experience of a (local) place and perceptions of one's place-in-the-world. This conceptualisation sees place as simultaneously centre of lived meaning and social position. Place involves an interactive link between social status and material conditions and can be used to interpret a range of situated health effects that imply a link between mind, body, and society."[60]

As Gesler summarizes, "Places provide meaning for people in many different ways: through identity and feelings of security, as settings for family life and employment, as locales for aesthetic experience"[61]. It is through lived experience that moral value, and aesthetic judgements are transferred to particular sites that, as a result, acquire a spirit or personality. It is this subjective knowledge that gives such places significance, meaning and the felt value for those experiencing them. The field of care[62], or appreciation by non-visual senses - such as smell, hearing, touch and taste - of such places, is also associated with the unique placefulness given to them. Tuan describes "Topophilia" as the affective

[50] Walmsley, D.J. & Lewis, G. J. (1993) *People and Environment*, London: Longman; Saegert, 1985; Sixsmith, 1986; Hayward, 1975
[51] Relph, 1976
[52] Walmsley & Lewis, 1993.
[53] Tuan, 1974
[54] Abel & Kearns, 1991; Williams, 1998
[55] Relph, 1976
[56] Pred, 1983
[57] Relph, 1976; Tuan, 1976
[58] 1998, p. 6
[59] 1995
[60] 1998, p. 6
[61] Gesler, 1992
[62] Tuan, 1974

bond between people and place or setting'[63]. Similarly, Relph describes this bond as existential insidedness: "the most fundamental form of insidedness... in which a place is experienced without deliberate and self-conscious reflection, yet is full of significance"[64]. Cosgrove[65] surmises that "home is perhaps that place where most of us experience true existential insidedness". Just as some environments have negative connotations, experienced environments that have a strong positive sense of place have a therapeutic effect "on human attitudes and behaviour"[66]. Gesler[67] describes landscapes endowed with a strong sense of place as being known only from within over long periods of acquaintance. This knowingness is exemplified in the home, where "networks of interpersonal concern"[68] have existed for an extended period of time. One health application of a strong sense of place is psychological rootedness, usually achieved through a long-standing and possibly ongoing relationship with a certain place. Somerville[69] is one of the many psychologists who have explored the role of home in human experience, arguing that home is physically, psychologically, and socially constructed, where individual meanings of home - such as privacy or identity or familiarity - can be internally explicated as a physical/psychological/social construct[70]. It is associated with those environmental features endowed with meaning that are related to one's life course. Implicit in this multifaceted view of home is the assumption that home allows person-environment transactions that satisfy basic human needs[71]. Sixsmith[72] describes home as an emotional reference point for a sense of self. Rowles[73] goes so far as to state that the need for home is a fundamental human imperative. Others have referred to the need for continuity and a sense of personal history[74], the need for personal autonomy and ability to effect desired change[75]. These needs continue throughout life, despite changes in age or life stage, or even changes in place of residence. Clearly, home has special meanings, and those meanings are important to one's feeling of well-being[76]. Through measuring the restorative qualities of favourite places, Korpela and Hartig[77] found that home was the most favoured, followed by water. Rowles[78] indicates that the phenomena constituting the psychological aspects of place attachment enhance well-being and even, at least speculatively, add years to life. There is consensus that this applies to patients being cared for at home, but less is known about family caregivers, which are increasingly represented as elderly given the aging demographic structure. Gerontologists have recognized the subjective meaning of home to older persons, recognizing that older persons wish to remain independent[79], value their homes in terms of family tradition[80],

---

[63] *Ibid*. at p. 4.
[64] Relph, 1976, p. 55
[65] 1978, p. 69
[66] Jackson, 1989
[67] 1992
[68] *Ibid* at p. 738.
[69] 1997
[70] Casey, 1993; Doyle, 1992; Ahrentzen, 1992; Fogel, 1992
[71] Lawton, 1985
[72] 1986
[73] 1978
[74] Shumaker & Conti, 1985
[75] Lawton, 1985
[76] Namazi, Eckert, Rosner, & Lyon, 1991
[77] Korpela, K. and Hartig, T. (1996) "Restorative Qualities of Favourite Places" *Journal of Environmental Psychology*, 16, 221-233.
[78] 1978
[79] Kummerow, 1980

derive status from homeownership[81], and use denial of poor conditions as a positive adaptive mechanism[82]. Empirical research has proven the subjective value of home to older persons, explained by competence in a familiar environment, traditional family orientation and memories, the status value of home ownership, and a cost versus comfort trade-off factor[83]. The question of whether the experience and meaning of home maintains its sense of place for family caregivers over the care-giving process has yet to be determined. A number of social scientists have suggested the need for such research, noting that the role of material aspects of housing and of societal and individual forces in the production and reproduction of the meaning of home has been neglected in the literature[84]. Given the emphasis on the negative effects of providing informal care in the home (in particular, the restriction on autonomy, leisure activities, as well as adverse effects on psychosocial and physical health), especially for women[85], a more critical reading of home as place is required.

c) Personal data in the legal perspective

While there is a significant body of literature on data protection, the question 'what are personal data?' is seldom addressed. Indeed, a widespread, definitive understanding of the concept 'personal data' has been assumed by commentators and policy makers alike. However, when one examines the debates and questions which have started to emerge following the implementation of Directive 95/46EC, it becomes apparent that there is a strong case for a rigorous re-consideration of the conceptual foundations of 'personal data'.

The philosophy underlying Directive 95/46EC is outlined by Bainbridge (1996). He describes how the aims and objectives of the Directive are deeply rooted in the historical development of data protection policy in Europe, emphasising the importance of the need to strengthen the internal market. He explains that by harmonising data protection law across the community, member states would be prevented from restricting or prohibiting the movement of personal data within the community. This, he notes, "mirrors the equivalent principles that apply to goods, services, persons and capital".[86]

However, a detailed analysis of the implementation of Directive 95/46/EC across Europe (provided by Korff [87]) suggests that differences in interpretation of the Directive have the potential to become significant obstacles to the internal market. His 2002 study examined the differences in the implementation of the Directive across member states, assessing textual divergences (between the Directive and national laws, and between national laws themselves) and the practical effects of these divergences. Although he only discovered

---

[80] Langford, 1962
[81] Baer, 1976
[82] Lawton, 1980
[83] O'Bryant (1981)
[84] Despres, 1991; Rubinstein, 1990; Somerville, 1997
[85] Brannen, 1992; Ginn & Arber, 1999
[86] Bainbridge, D (1996) *EC Data Protection Directive* London, Butterworths, at p. 42
[87] Korff, D (2003), *EC Study On Implementation of Data Protective: Comparative Study of National Laws* http://europa.eu.int/comm/internal_market/privacy/docs/lawreport/consultation/univessex-comparativestudy_en.pdf

minor variations in approach in implementations of the key terms of the Directive, it is recognised that some of these differences are likely to lead to more serious differences in practice,

> "As a result of seemingly minor additions or variations, some data will be regarded as "personal" in some countries, but not in others; some processing systems will be regarded as (sufficiently structured) "filing systems" to fall with the law in one country, but as insufficiently structured or easily searchable – and thus outside the law – in another" [88].

These findings were confirmed by Beyleveld *et al.* in 2003. Although their project, PRIVIREAL[89], was restricted to the implementation of the Directive in relation to medical research and the role of ethics committees, they confirm the differences in approach to the operationalisation of the key terms. An analysis of both studies has enabled us to identify a number of questions which have emerged from the implementation of the Directive. These questions are discussed in detail below.

Is there a meaningful distinction between the concepts 'data' and 'information'?

In accordance with the approach of the Directive, it seems that very few countries have made a clear distinction between 'data' and 'information'. It should, of course, be noted that the UK *does* draw a distinction. The Data Protection Act 1984 s1(2) stated that " 'Data' means information recorded in a form in which it can be processed by equipment operating automatically in response to instructions given for that purpose". This suggests that data is a particular (sub-) category of information.. The distinction between data and information continues to be drawn in the 1998 Act (s1(1)) although the boundaries of the (sub-) category were extended. However, Korff rightly notes that the distinction drawn between data and information *per se* seems to have had little effect in practice[90]

This distinction between 'data' and 'information', drawn by the Act, seems out of step with other ways of conceiving the difference between the two concepts. Rather than understanding data to be a (sub) category of information, information is more commonly understood to represent that which may be gathered through the 'processing' of data. The implication of this is that 'data' may exist prior to processing: prior to information. Rouille-Mirza and Wright suggest that 'data' can be seen as inert symbols, signs or measures, while 'information' is data which has been acted upon by a decoding device, which they suggest may take the form of a brain. Thus it is use of such a device which the core to the distinction. This means that 'data can be information, but there is a time when data can exist where information does not- when it is just inert symbols prior to processing'[91]. Using the example of DNA, they illustrate how this may have implications where certain types of 'data' are taken to fall within the Directive, placing restrictions on its use. DNA can be seen as data 'with the potential to become information'[92] and as the Directive draws no distinction between data and information, DNA (even if it remains as 'raw' data

[88] Korff, D *op cit* p.11
[89] This project was funded by the European Commission.
[90] Korff, D (2003), *EC Study On Implementation of Data Protective: Comparative Study of National Laws* at p. 13
[91] Rouille-Mirza, J and Wright, J, (2003) PRIVIREAL Issues Paper s2.1.2
[92] Ibid.

and no information has been gathered through its processing) is presumably included. This, they suggest, would have massive implications for medical research.

How clear is the concept of the 'natural person'?

Two separate questions seem to have emerged from the debates surrounding the term 'natural person'. The first is whether a 'natural person' must necessarily be alive or whether the term might include the deceased? The second question relates to the status of 'legal persons'; does the concept of the 'natural person' include, for example, companies?

With regard to the first question, Korff notes that some countries do extend data protection to include the deceased, while others (including the UK) do not[93]. However, as Rouille-Mirza and Wright point out, the fact that this extension is made does not necessarily mean that countries have interpreted 'natural person' to include the dead in practice. They suggest it is difficult to afford rights to the dead as it is impossible for them to enact the 'active rights' contained in the Directive. It would, therefore, be up to somebody else (a relative, a legal representative) to enact rights on their behalf.[94] In any event, the current UK position has been confirmed by the judge in *D v L* who saw personal data as "information relating to a living individual who can be identified from that information."[95] Note here the emphasis on 'living', which follows the comments of Lord Mostyn in the debate on the Data Protection Bill in the House of Lords.[96]

On the issue of 'legal persons', Korff notes that some countries do extend the concept of the data subject to 'legal persons'. Country 3, for example, includes data on companies (held by credit reference agencies or blacklisting companies) to qualify as 'personal data'. Similarly, Country 1, Country 12, and Country 13 have extended data protection to 'legal persons' in some circumstances.

The Confederation of British Industry (CBI) has issued strong warnings about the adoption of this approach[97]. The CBI argues that Recital 24 suggests that legal persons should not be dealt with under data protection legislation. The contrary approach adopted by some countries has caused confusion, which the CBI sees as creating obstacles to the single market. It is suggested that, in order to ensure a truly single market, we should adopt a consistent approach and that countries which have afforded rights to legal persons should be persuaded to restrict data protection to 'natural persons'.

Is the concept of 'personal data' relative?

As Korff notes, the definition of personal data contained in the Directive can be read as being 'relative'. This means that potentially 'any data that can conceivably be linked to an

[93] Korff, D (2003) *EC Study On Implementation of Data Protective: Comparative Study of National Laws* at p. 30
[94] Rouille-Mirza, J and Wright, J, (2003) PRIVIREAL Issues Paper s1.1.1
[95] *D v L* [2003] EWCA Civ 1169
[96] Hansard, February 2nd 1998, at Column 438
[97] Confederation of British Industry, '*CBI Official Response to the Review of Directive 94/45/EC*' (August 2002) http://www.cbi.org.uk/ndbs/PositionDoc.nsf/fb66d262805fa2f58025673a0058587b/c70a78a03cd38caf80256c37005c2f44/$FILE/cocolaw9546ec0802.pdf

individual (in whatever way, by whoever) [can] be regarded as personal'.[98] Thus, in the Directive's definition of 'personal data', a great deal hinges on the phrase 'relating to'. The way that this phrase is interpreted has dramatic repercussions as to what is or is not classed as personal data.

When interpreted very narrowly, the term can be restricted to data which is capable of identifying an individual, either by itself or in combination with other data. Identification, in this context, can be direct or indirect. However, if the Directive is interpreted to include *indirect* identification by anyone, the nets are thrown wide open and even this 'narrow' interpretation can encompass a huge range of information. Accordingly, Rouille-Mirza and Wright argue that "…for data to be ever fully anonymous there can be no instance anywhere in the country or even the world, where information that can be used to link anonymised data to the individual exists". They suggest that the 'reasonable means' test suggested in Recital 26 is a potential 'practical solution' to define the limits of indirect identification.[99]

If we choose to interpret the term 'relating to' more widely, the waters are muddied further. We may take the term to include any data which may 'affect' the individual in some way, regardless of its capacity to identify

The relationship between the legal, the sociological and the psychological perspectives

If 'affect' is attributed significance by the law then there is clear potential for overlap between an understanding of the term from the legal perspective, and those understandings associated with either the sociological or psychological perspective. For example, in order to try to understand the value attached to personal information by an individual, we might consider the literature on Identity Theory.

Identity is a concept that extends beyond mere external identification and is a construct linked to the concept of privacy. According to Giddens, each individual has a 'personal identity' and a 'social identity' - these are separate elements.[100] This is a classical sociological position. However Jenkins has more recently argued that these two elements are in fact "entangled"[101] (see also Cooley[102] and Mead[103]). Individuals define themselves (and others) in the process of social interaction. It is this combination of internal and external that gives a person (and collectivities) his/her 'identity'. Jenkins calls this the "internal- external dialectic of identification". [104] Exactly what constitutes an individual's identity will inevitably vary across culture and time[105] and is subject to expectations

[98] Korff, D, (2003) *EC Study On Implementation of Data Protective: Comparative Study of National Laws* at p. 14
[99] Rouille-Mirza, J and Wright, J, (2003) PRIVIREAL Issues Paper s1.1.2
[100] Giddens, A, 1991) *Modernity and Self Identity: Self and Society in the Late Modern Age* Cambridge, Polity, at p. 91
[101] Jenkins, R, (1996) *Social Ide*ntity London, Routledge at p. 19
[102] Cooley, CH, (1962) *Social Organisation: A Study of the Larger Mind* New York, Schoken; Cooley, CH, (1964) *Human Nature and the Social Order* New York, Schoken
[103] Mead, GH,(1934) *Mind, Self and Society From The Standpoint of a Social Behaviourist* ed. CW Morris Chicago, Chicago University Press
[104]Jenkins, R, (1996) *Social Identity* London, Routledge at p. 20
[105] Cushman 1991: noted by Gover, Mark R, and Gavalek, J, Persons and Selves (1996): *The Dialectics of Identity* Michigan State University

regarding issues such as gender, class and race[106] as well as consisting of 'features that distinguish him or her from the group'. [107]

Social Identity Theory[108] in social psychology suggests that we construct our own identity in the following ways: C*ategorisation* (defining our behaviour by reference to the norms of the groups we belong to); *identification* (defining ourselves from the identity of the groups we belong to (social identity) and also in terms of our identity as an individual (personal identity)); and *social comparison* (in order to evaluate ourselves we compare ourselves with similar others, and identify ourselves with a prestigious group).[109]

If social and group identity is based on common social categories, then the person is an individual and simultaneously has a number of categories to which they can belong. We engage in 'self categorisation'[110] by perceiving ourselves as a unique individual (personal identity) and a member of different groups (social identity) at different times.[111] Although individuals can belong to some groups by choice, other group memberships are ascribed and therefore not necessarily recognised or acknowledged by the individual, so there may be a vast difference between characteristics that form our own sense of identity, and those ascribed to us. Whether it is only 'personal' information that might form part of our personal identity is debatable. Even information that might be seen as 'relating to' others can form part of a person's sense of self and assist in creating her understanding of her own identity. [112]

The law has recognised that the individual's thoughts on what is their personal data should be given some weight. The Directive recognises that there are some 'sensitive' types of information, but even a recorded conversation (audio information) can be personal to someone if they think it should be private, regardless of subject. Individuals' claims that information is personal to them even if they are not the subjects of it have succeeded in law. [113] However, if the data subject alone decides what their personal data is, it provides them with the potential to protect that data (or information) which is stigmatised by society-precisely the type of data (e.g. police charges) we may want to access in certain situations (e.g. job interviews).

Privacy has usefully been described as 'the selective control of access to the self or one's group'.[114] Kupritz has identified three central themes of privacy: retreat from people, control over information and regulation of interaction. [115] Although privacy is often viewed

---

[106] Sampson 1993 noted by Gover and Gavalek, ibid.
[107] Deborah Larson: 'Comments on Paul Kowert' www.mershon.ohio-state.edu
[108] Tajfel and Turner, 1979 noted in *'Social identity and Self-Categorisation*' www.anu.au/psychology/social/socident.htm
[109] Festinger 1954, ibid.
[110] Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., Wetherell, M.S. (1987) *Rediscovering The Social Group: A Self Categorisation Theory* Oxford, Blackwell
[111] Article on Social identity and Self-Categorisation www.anu.au/psychology/social/socident.htm
[112] *Odievre v France* [2003] 1 FCR 621 where the European Court rejected a decision that information about a birth was private to the parents, deeming it to be accessible to the applicant - whose birth it had been - under Article 8 ECHR rights.
[113] Gu*nn-Russo v Nugent Care Society and Another* [2001] EWHC Admin 566, CO/4370/2000, [2002] Fam Law 92
[114] Altman 1975, noted by Pedersen, D (1997) *Psychological functions of privacy*, Journal of Environmental Psychology No 17 at pp. 147 – 156
[115] Kupritz, V.W. (1998) *Privacy in the workplace: The impact of building design* Journal of Environmental Psychology No. 18 at pp. 341 – 356

as a physical state, its 'permeability' may also be achieved by 'leakage' of information by technological means.[116] Indeed, the ability to control the flow of one's personal information, arguably the central tenet of the concept of privacy, may be understood as a necessary good for autonomous action. It is well documented that the desire for privacy is deeply rooted in natural (human and animal) instincts. Westin[117] describes in some depth the role privacy plays in both the animal kingdom and function of primitive societies, suggesting that the need for privacy results in social norms which are evident in most societies. For Westin, the individual's ability to control the flow of information about him/herself is the key to understanding social structure in all societies,

> "The point is that kinship rules and interaction norms present individuals with a need to restrict the flow of information about themselves to others and to adjust these regulations constantly in contacts with others. This need is fundamental to individual behaviour with intimates, casual acquaintances, and authorities" [118]

Privacy, then, is an essential component of individual autonomy, at least in a democratic society. [119] Concepts of 'privacy' and 'personal' can be seen to be purely dependent upon the social and cultural mores of the time and context. It has been suggested that "the requirement that disclosure or observation of information or conduct would be highly offensive to a reasonable person of ordinary sensibilities is in many circumstances a useful practical test of what is private."[120] Presumably, what the 'reasonable person' deems as private is also dependent on time and place. [121]

The complexity of the concept of privacy is clear, and this is reflected in English law. A UK Government document highlights the difficulties of definition:

> "Domestic law does not provide a single definition of the term privacy and therefore what might be included in a right to privacy. Definitions of privacy are most often attempted by reference to its opposite – distinguishing that which is rightfully private from that which is public. …While it may be possible to identify certain matters that may generally be private or included in a right of privacy any definition is inevitably subjective since it will depend upon an analysis of all the relevant facts and circumstances of the case." [122]

---

[116] Shapiro, S. (1998) *Places and Spaces: The Historical Interaction of Technology, Home, and Privacy* The Information Society No. 14 at pp. 275 - 284

[117] Westin, A. (1967) *Privacy and Freedom*, London: The Bodley Head.

[118] Westin, A. (1967) *Privacy and Freedom*, London: The Bodley Head, at p.14

[119] Alldridge, P. and Brants, C. (2001) *Personal Autonomy, The Private Sphere and the Criminal Law*, Oxford: Hart Publishing

[120] *Country 36n Broadcasting Corp v Lenah Game Meats Pty Ltd* (2001), quoted by Lindsay J, at paragraph 188 of *Douglas v Hello!* [2001] QB 967

[121] The tax laws in Sweden provide a current example. In Sweden, a person's yearly income tax return is publicly accessible complete with photograph, information that in some cultures might be considered too personal for general publication. In the UK, asking a person for information about their salary was considered highly vulgar and an invasion of privacy just 50 years ago, but is much less so now. Discussion of sexual behaviour has also become more acceptable; witness the plethora of TV programmes with people only too keen to divulge intimate details of their 'personal' lives.

[122] *Privacy and data-sharing: The way forward for public service* The Performance and Innovation Unit www.number-10.gov.uk

Despite the doubts that may be raised over the appropriate meaning to be attached to the terms 'privacy' or 'identity', it does appear that, within UK law at least, the term 'relating to' has been associated with a meaning that extends beyond simple 'identification'. In *Durant v Financial Services Authority*[123] Auld LJ stated that "not all information retrieved from a computer search against an individual's name or unique identifier is personal data within the Act. Mere mention of the data subject in a document held by a data controller does not necessarily amount to his personal data. Whether it does so in any particular instance depends on where it falls in a continuum of relevance or proximity to the data subject."[124] Clearly, this is a reading of 'relating to' that requires something more than mere identification for data to be classed as personal.

Over the years the law's understanding of where the boundaries of that 'relevance' or proximity might fall has not always been consistent. In a now dated case (*R v Brown*[125]), vehicle ownership information was considered to be 'personal data' under the 1984 Data Protection Act, but in *Farrer v Secretary of State*[126], the judge felt that vehicle details were "data which in my opinion barely touches that privacy [the claimant's private life]". Similarly, vehicle details may not be personal data even if other identifying details obtained alongside them in the same context are.[127]

Further issues within the legal perspective:

Are the terms 'processing', 'filing system' and 'anonymisation' in need of clarification?

Anonymisation

Recital 26 of the Directive states that the principles of data protection do not apply to anonymised data. Latham J in the original *Source Informatics*[128] case stated that "if anonymity is guaranteed, [the patients'] privacy would not be invaded". Brown LJ felt that "Council Directive (95/46/EC) can have no more application to the operation of anonymising data than to the use or disclosure of anonymous data (which of course, by definition, is not 'personal data' and to which, therefore, it is conceded that the Directive has no application)". Brown LJ also felt moved to remark on the "striking paucity of authority on the … anonymisation of confidential information and its subsequent use in anonymised form". [129]

However, Beyleveld and Townend argue that "the only times that data rendered non-personal can be said to be beyond the scope of the principles of protection is where the data no longer has a history that can link it to an identifiable data controller who obtained the personal source data from the data subject or where it is known that the source data

[123] [2003] EWCA Civ 1746, [2003] All ER (D) 124 (Dec).
[124] Ibid., paragraph 28ff
[125] [1996] 1 AC 543
[126] [2002] EWHC 1917 Admin
[127] *R v Rees* Court of Appeal, Criminal Division 20 October 2000 paragraph 17
[128] *R v Department of Health ex parte Source Informatics Ltd* [2001] QB 424
[129] ibid. at paragraph 44

was given for unlimited purposes".[130] This suggests that to ever truly anonymise data, and therefore render a subject unidentifiable, is extremely difficult.

Processing and the 'Filing System'

Both the terms 'processing' and 'filing system' have come under closer scrutiny recently, particularly within the UK context.

Article 2(c) of the Directive defines a 'personal data filing system' as

> "…any structured set of personal data which are accessible according to specific criteria, whether centralized, decentralized or dispersed on a functional or geographical basis".

This suggests that for the filing system to exist it must contain personal data; if the data is not personal, then a relevant filing system does not exist.

Korff notes that both the UK and Finland have attempted to narrow the definition in the Directive, with the result that within their domestic law "structured sets of data may fall outside the concept of filing system, even though they would not elsewhere[131]". Indeed, the Data Protection Act (s1(1)) defines a 'relevant filing system' as "any set of information relating to individuals to the extent that, although the information is not processed by means of equipment operating automatically in response to instructions given for that purpose, the set is structured, either by reference to individuals or by reference to criteria relating to individuals, in such a way that specific information relating to a particular individual is readily accessible". This would, indeed, appear to confuse the issue.

Korff suggests that the UK Information Commissioner has tended to take a flexible approach to the concept of the filing system in order to avoid difficulties. However, the decision in *Durant v FSA* may well limit the Commissioner's flexibility. Moreover, the introduction of the Freedom of Information Act is also likely to have an impact upon this situation.

In the recent case of *Lindqvist*[132]*,* the concept of 'processing' was considered. The case dealt with the internet display of data concerning others, loaded by an individual for a non-commercial purpose. The case was brought on the basis of the provisions of the Directive by complainants whose information had been published on the internet without their consent. It was held that some of the data (references to a medical condition of one individual, together with names and other information which made the complainants identifiable) constituted 'personal data'. Loading the information onto the internet was considered to be processing for the purposes of Article 3 of the Directive, in that entering it into the computer at all was considered to be processing. The fact that the defendant was a private individual and not involved in economic activity did not mean that the activity of

[130] Beyleveld, D, and Townend, D, *When Is Personal Data Rendered Anonymous? Interpreting Recital 26 of Directive 95/46/EC* (World Congress on Medical Law, August 2002), article published in (2004) *Medical Law International*, 6(2), 73-86.
[131] Korff, D (2003*) EC Study On Implementation of Data Protective: Comparative Study of National Laws* at p. 20
[132] Case C-101/01

publishing information on the internet fell outside the scope of Community Law. In addition, the activity did not fall under the exceptions cited in Article 3(2) of the Directive.

Is it possible to balance the Freedom of Expression v Right to Privacy?

The tension between freedom of expression as enshrined in the ECHR and the Data Protection Act's attempt to preserve a right to privacy was exposed in *Campbell v MGN Ltd*[133]. The claim was made that exemptions from the 1998 Act (for journalistic publications) only apply to pre-publication processing, and that publication itself is outside the scope of the Act. The freedom of expression rights of journalists would be compromised if this were not the case. MGN contended that the Act is incompatible with the ECHR in that it has created a law of privacy and a fundamental enhancement of Article 8 ECHR at the expense of Article 10 (para 91). Much of the defendant's (MGN) argument "was founded on the submission that it was virtually impossible for journalists to comply with the requirements of the Act" (para 74). The final judgement disagreed with the defence of inapplicability of the 1998 Act to journalistic publication, as not to do so would rob the Act of a good deal of its use and force.

In *Lindqvist*[134], the defendant's freedom of expression under Article 10 of the ECHR was also raised as an issue - the European court left it to the national court to find a balance between rights and interests that may have conflicted. It was decided that the Directive does not 'bring about a restriction that conflicts with the general principles of freedom of expression or other freedoms and rights'.

Key Findings

While the current literature on data protection deals with several of the themes that are crucial in coming to an understanding of 'what are personal data', it is clear that there is no one uncontested and coherent definition of 'personal data'. None of the issues discussed above are 'settled' in any way.

[133] [2002] EWCA Civ 1373, [2003] QB 633
[134] Case C-101/01

A3: Methodology of the Survey

At the start of the project, an invitation to participate in the project was distributed via email to 39 countries (this list was compiled via an internet based search for email contact details for Data Commissioners). Of those who replied, 15 agreed to complete the first questionnaire. [135] We received multiple responses from Germany (Federal Germany and three Länder). Therefore, in total, we received 18 responses. Of these 18, 11 respondents completed the follow-up questionnaire (Questionnaire 2).

For the purposes of this report, all countries have been anonymised and shall be referred to by numbers only. Numerical labels were assigned at random to countries at the initial sampling stage and therefore range from 1-42 (39 countries plus 3 German Lander).

The jurisdictions surveyed were divided into three groups for the purposes of analysis:

Group 1 - 8 EU jurisdictions

Group 2 - 7 jurisdictions outside the EU. These either wish to have 'adequate' data protection compatible with EU legislation for trade (under Article 25(6) of the Directive), or be compliant with the legislation in order to join the EU in future.

Group 3 - 3 countries outside the EU with no requirement of compatibility.

[135] Iceland specifically declined to take part in the study and Spain had to be excluded from the sample since it was not possible to respond in English. All remaining countries failed to respond to the research team directly. Following phase 1 of the project, the Information Commission placed a call for further participation. At this later stage we received questionnaires from three countries. The Netherlands contacted the office of the Information Commission to check the legitimacy of the project but, although it received a positive response from the IC, did not complete a questionnaire.

The Questionnaires

This section contains a brief description of the questions contained in Questionnaires 1 and 2 (Q1 and Q2). Full copies of both questionnaires are provided in Appendices 3 and 4. Questionnaires were distributed, completed and returned electronically.

Questionnaire 1

Questionnaire 1 was designed to collect a wide range of data relating to the definitions and interpretations of key terms within the Directive ('personal data', 'information', 'relating to', identified or identifiable', 'natural person' and 'personal filing system'), with a particular emphasis on how these interpretations work in practice.

Question 1 was designed to test whether it is possible to draw up a list of data types which are always, sometimes or never personal data. Respondents were presented with a list of 35 data types (see Appendix 3) and asked to classify each type as *always*, *sometimes* or *never* personal data. Where a respondent said that a data type was *sometimes* personal data, s/he was asked to give examples of situations in which the data would NOT be classed as personal. It should be noted that the question was posed in such a way that it did not assume that an individual had already been identified.

Question 2 was designed to assess how the Directive has been implemented in different EU jurisdictions, focussing on a series of key terms ('**personal data**', '**data**', '**information**', '**relating to**', '**identified or identifiable**' and '**natural person**'). Where the jurisdiction was not a member of the EU, respondents indicated whether these key terms or similar terms have been included in individual pieces of legislation. Question 3 asked respondents to identify any problems/issues arising from the practical interpretation of the key terms highlighted in question 2.

Question 4 explored the interpretation of the distinction between 'direct' and 'indirect' identification by asking about formal definitions and personal understanding of the terms.

Questions 5 and 6 were designed to investigate the interpretation of the term 'personal filing system'. Question 5 focussed on formal definitions of the term 'personal filing system'. Question 6 was developed to attempt to understand the use of the term through the use of examples (requiring respondents to say whether different examples of filing systems would be classed as 'personal' *always*, *sometimes* or *never*).

Question 7 asked respondents to indicate whether the process of anonymisation is capable of transforming 'personal data' into 'non-personal data'. Also, each jurisdiction was asked whether a distinction is drawn between different methods of anonymisation.

Questionnaire 1 ended with an open question, inviting respondents to make any additional questions/comments about the interpretation of the term 'personal data'.

Questionnaire 2

The second questionnaire was developed following the preliminary analysis of the first questionnaire and the conclusion of the first stage of the development of the theoretical framework. There were two key aims of the second questionnaire:

1. Clarification of responses to Questionnaire 1, with a view to developing more fully our understanding of the use of the term 'personal data'.
2. As the results from the first questionnaire started to converge with our theoretical framework, we wished to present the emergent ideas to the respondents and find out their views.

The second questionnaire was, therefore, split into three Parts (A-C).

PART A:

Respondents were given a table which presented their own, individual classifications of the 35 data types as *always*, *sometimes* and *never* personal data. A series of 4-5 questions were individually designed for each country. Each respondent was presented with a summary table of their own responses and given the opportunity to make alterations. Next, questions were designed to probe into interesting trends emerging from the original answers in the following ways:

- Where a respondent had distinguished between data types which appeared to be similar (such as 'Mother's maiden name' and 'Parents' names'), we asked the respondent to explain why the data types were placed in different categories. So, for example, a respondent would be asked to explain why 'Mother's maiden name' is only *sometimes* personal data, but 'Parents' names' are *always* personal data.
- Where a number of data types had been classed as *sometimes* personal data, the respondent was asked whether these data types would be classed as personal data in the same circumstances, or whether there were differences between them. Respondents were encouraged to use examples to help illustrate their responses.
- If a respondent had not used one of the three categories (*always*, *never* or *sometimes*), s/he was asked to give examples of data not listed which may fall into the empty category. So, for example, if a respondent had not listed anything in the *never* category, s/he was asked whether s/he could think of any kind of data which might not ever be capable of being personal data.

Following the first questionnaire, our understanding of the use of the term 'personal filing system' was limited. Respondents had not been given the opportunity to expand on their answers to question 6, at least in the way they had been for question 1. Therefore, in the second questionnaire respondents were given a table which presented their own individual classifications of filing systems as *always*, *sometimes* and never personal filing systems. Again, respondents were given the opportunity to make alterations to their classifications. Next, in a similar way as for the previous question, a series of 4-5 questions were designed to explore interesting trends to emerge from the classifications in Questionnaire 1.

- Where a number of filing systems had been classed as *sometimes* personal filing systems, the respondent was asked whether these would be classed as personal filing systems in the same circumstances, or whether there were differences

between them. Respondents were encouraged to use examples to help illustrate their responses.

- If a respondent had not used one of the three categories (*always*, *never* or *sometimes*), s/he was asked to give examples of filing systems not listed which may fall into the empty category. So, for example, if a respondent had not listed anything in the *never* category, s/he was asked whether s/he could think of any filing system which might not ever be capable of being a personal filing system.

PART B:

SECTION 1:

In the first section of Part B, a series of scenarios was designed to tease out specific elements of the complex interpretation of the term 'personal data'.

Scenario 1:

In this scenario, a character named 'Gordon Rocer' buys a grocery store and finds an old order book (containing customer's telephone orders, listed by the customers' surnames) and five CCTV tapes (containing clear pictures of individuals' faces). Respondents were asked to indicate whether the information stored in the order book and the video tapes constituted 'personal information' in two distinct circumstances:

a) Where the shop is situated in a densely populated area (where the customer surnames may be shared with a number of people in this area).
b) Where the shop is situated in a small village (where the surnames are more likely to be related to an individual).

The aim of this scenario, then, was to explore the importance of the capacity to identify an individual in two different contexts. In the first, it would be quite difficult to identify an individual from the list of names or the CCTV tapes as the area is densely populated and the information may relate to more than one individual. In the second scenario, the customers could more easily be identified from their surnames and/or CCTV pictures.

Scenario 2:

In this scenario, 'Music Maker' is an online service for young musicians. Musicians put their music on the website and visitors to the site post comments on the recordings. These comments are collected into an online profile and visitors to the site may access each profile. Each profile is listed by the musician's name (or the name of a band). A positive profile is of economic value to the musician as it boosts the reputation of the musician and may influence future recording contracts. Respondents were asked whether the information contained within the profiles constituted 'personal data' in three different circumstances:

a) Where a musician posts his music on the site using his own name
b) Where a musician posts his music on the site using an alias
c) Where a group/band post their music on the site using the band's name

The aim of this scenario was to explore two questions. Firstly, must the data be capable of identifying a specific individual? Can data which relates to a group or to an individual who hides his real identity through an alias be classed as 'personal data'? Secondly, what weight do respondents attach to the likely effect that data will have on an individual? In this example, a negative profile may be damaging to the reputation of the musician(s) and thus have economic implications. How important is the potential to affect, in cases where identification is and is not possible?

Scenario 3:

In this scenario, 'CD Success' is a record company which uses a website to gage public opinion about music. Visitors to the site are asked what music they like, what bands they listen to, etc. and this information is used to inform CD Success's decisions about which bands to sign or contracts to extend. Respondents were asked whether the information submitted by visitors to the site is the 'personal data' of:

a) The groups/bands already recording under the CD Success label.
b) The unsigned bands who hope to get a contract with CD Success.
c) The individual visitors to the site who have submitted their preferences/opinions.

This scenario was similar to Scenario 2, but focussed on two slightly different questions. Firstly, where an individual posts his/her opinions in a public forum, do those opinions remain the 'personal data' of that individual or do they become the 'personal data' of someone else? Secondly, if data which can affect an individual may be classed as 'personal data' how close must the relationship between the data and the subject be? In this example, is the relationship between the data and the unsigned bands too remote?

Scenario 4:

In this scenario, Albert and Brenda are named as owners of an antique restoration business. Albert leaves the business and takes all the records of the clients (listing names and addresses of clients). Brenda subsequently claims that the client records are her 'personal information' because they affect her ability to conduct her business. Respondents were asked whether they agreed with Brenda's argument. Next, respondents were asked to indicate whether their answer would change if the clients had become Brenda's friends and thus she argued that the loss of the information affected her social life.

The aim of this scenario was to build upon some of the ideas explored in Scenario 3. At the first stage of this scenario, we simply asked respondents whether the names and addresses remain the 'personal data' of the clients, or whether they become the 'personal data' of Albert and Brenda. Next, in order to explore the importance of the nature of the possible effects of the data, we asked respondents to distinguish between two sets of circumstances. First, where Brenda will suffer financially/economically and, second, where Brenda's social life will be affected.

Scenario 5:

In the final scenario, respondents were asked to consider the case of George, a social science researcher, who has conducted a survey of 200 students to investigate the reading habits of different students. Each student has provided his/her name and address. Respondents were asked to indicate whether the contents of the student questionnaires constitute personal data where:

a) The questionnaires are kept alphabetically, according to students' names
b) The questionnaires are stored according to the favourite book of students
c) The questionnaires are kept in a random pile in George's office, in no order.

The aim of the final scenario was to investigate the relevance of different filing systems to the concept of 'personal data'. In each of the three examples, the survey data is stored in a different way. Whilst it may be easy to identify an individual where the questionnaires are ordered or organised in a specific way, this may or may not determine whether the data should be classed as 'personal'. This scenario attempted to explore this idea further.

SECTION 2:

This second section was designed to explore the relevance of different contexts to the classification of personal data. Each respondent was presented with a table (see Q2 in Appendix 4) and asked to indicate whether 11 different data types are more or less likely to be classed as personal data within a series of 13 different contexts.

PART C:

In Part C, we presented some of the ideas which had emerged from our theoretical discussions to the respondents in order to test their views. Respondents were asked whether they agreed or disagreed with the following statements:

1. It is impossible to create a list of what is or is not 'personal data' as the concept is entirely dependent on the context in which the information is placed.

2. a) Information can only be personal data if it can identify an individual
   b) Information can only be personal data if it does not identify an individual but can affect an individual in a different way.
   c) Information can only be personal if it both identifies and affects an individual.

3. a) The effect of processing information about an individual can relate to his or her fundamental rights to private and family life in many ways. The definition of personal data should only reflect protection against significant harm to an individual.
   b) Information that produces any effect upon an individual must be defined as 'personal data' and it is then the rest of the law that determines its protection.

4. Individual identity goes far beyond identification but is protected within the definition of personal data.

The responses to both questionnaires are extensively referenced throughout this report. In order to make referencing as clear as possible, we have adopted the following style:

Following the country name, a series of references are given in brackets. The first refers to the questionnaire number (Q1 or Q2). The references which follow can be directly linked to the sections and specific questions within both questionnaires. For Questionnaire 1 (Q1), reference will simply be made to the question number. For

Questionnaire 2 (Q2), the reader will be guided to the relevant section of the questionnaire, and then the specific question number.

Examples: Country 27 (Q1, q1), Country 17 (Q2, A, Part 1, q.3(a))

PART B: 'PERSONAL DATA' IN PRACTICE

B1: Introduction

Part B of the report contains the results of the analysis of both Questionnaires 1 and 2. The discussion centres on the conceptualisation and practical operationalisation of the term 'personal data'.

In section B2, the formal definitions of the key terms within Directive 95/46/EC are presented, based on the responses given in Questionnaire 1. We will see that countries within the EU have generally approached the implementation of the Directive by simply transferring the key terms directly into state legislation. The EU countries report very few difficulties with the day-to-day interpretation of the key terms. In contrast, the non-EU jurisdictions seem to have recognised a few more problems, both with the development and the application of the concepts. In particular, the distinction between 'data' and 'information' and the definitions of the terms 'relating to' and 'identified' seem to have caused the most problems for Data Protection Authorities. In order to complement this discussion, a full summary of responses to these questions and references to further sources can be found in Appendices 1 and 2.

In section B3, we focus specifically on the question of 'what are personal data?' Specifically, we look at how the concept of 'personal data' is operationalised. Using the responses from question 1 (Questionnaire 1, as confirmed by Part A of Questionnaire 2), a discussion of the results reveals widespread inconsistency in the understanding of and application of the concept. Not only do we see differences in approach *between* different countries (both within and outside of the EU) but we also see interesting conceptual anomalies *within* many of the individual countries. Also discussed in this section are the results to question 6 (Questionnaire 1), which explored the interpretation of the term 'personal data filing system'. Again, we see significant differences in approach across EU and non-EU countries.

The quantitative analysis in section B3 leads us directly into the more detailed analysis of the qualitative responses provided in Questionnaire 2. In section B4 we ask 'what concepts of personal data are at work?' We explore the responses to the questions on a country-by-country basis and discover that there are three clear approaches to the understanding of the concept 'personal data'. The first is to accept that the defining feature of 'personal data' is its capacity to identify an individual. In contrast, the second approach is to include all data which 'relates to' or 'affects' an individual in some way. The third approach is to require data to have both 'identification' and 'affect' properties in order to qualify as 'personal data'.

Part B5 concludes by asking whether we should be concerned as to the apparent lack of clarity surrounding the concept of 'personal data'. It is argued that uncertainty does indeed create a number of serious practical problems. We suggest that, as long as Data Controllers remain confused about the foundations and mechanics of the concepts they employ, the possibilities for arbitrary decision-making are increased. Moreover,

fundamental differences in approach between EU jurisdictions may prevent the achievement of the original aims of the Directive (i.e. creating a harmonious system which protects personal data and fundamental rights and freedoms, most notably privacy, and simultaneously supports the single market).

B2: Formal definitions of the key terms within Directive 95/46/EC

This section will analyse how the key terms within Directive 95/46/EC have been implemented in EU jurisdictions and how these terms (or similar concepts) have been incorporated into the laws of jurisdictions outside of the EU. The analysis is based upon the responses given to questions 2-4 of Questionnaire 1.

The Directive defines personal data as:

" any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity"[136]

The key terms identified within this definition are: '**information**', '**data**', '**relating to**', '**identified or identifiable**', **'natural person**' and '**directly or indirectly**'. In the sections that follow, we present the interpretations of each term.

The jurisdictions surveyed were divided into three groups for analysis:

Group 1 - 8 EU jurisdictions
Group 2 - 7 jurisdictions outside the EU. These either wish to have 'adequate' data protection compatible with EU legislation for trade (under Article 25 (6) of the Directive), or be compliant with the legislation in order to join the EU in future.
Group 3 - 3 countries outside the EU with no requirement of compatibility.

Group 1:

Transposition into national legislation within the EU has produced wording that reasonably follows the Directive definition of personal data in order to comply with obligations of membership of the EU.

'personal data', 'data' and 'information'

- The term 'personal data', or 'personal information', is used by all EU jurisdictions in their legislation, usually as part of a list of definitions for use in the Data Protection legislation.
- Few jurisdictions have a formal definition of the terms 'data' or 'information'. [**Country 6, Country 7, Country 8, Country 1**]

[136] Article 2 (a)

- **Country 1** defined data in the following way: Information relating to data subjects who are identified or identifiable. An additional definition is provided for 'sensitive data': Data relating to natural persons concerning their racial or ethnic origin, political opinion, trade union membership, religious or philosophical beliefs, and data concerning health or sex life. No formal definition of 'information'.

  Country 7 defined information as**:** 'individual or single information' (translation). This is distinguished from information relating to groups of individuals, emphasising the need to refer to an individual.

- No jurisdictions reported any difficulty in defining or interpreting the terms 'data' or 'information'.

'relating to'

- Not used by all EU jurisdictions.
  - **Country 6** uses the term 'concerning'.
  - **Country 17** substitutes the phrase 'may be referable'.
- No formal definitions.
- Only two jurisdictions reported any difficulty in defining or interpreting this term. [**Country 1, Country 6**]

'identified or identifiable'

- All jurisdictions but one [**Country 17**] use these terms in their legislative definition of personal data.
  - **Country 1** uses 'identified or identifiable individual'
  - **Country 3** uses 'identified or identifiable natural person'
- Only **Country 1** reported difficulty in defining or interpreting these terms.

'natural person'

- No jurisdictions in this group have a formal definition of this term within their national legislation, although some have a Civil code covering the definition. [**Country 1, Country 12**]
- Only three jurisdictions use this term within their definition of personal data. [**Country 3, Country 17, Country 12**]
  - **Country 6** refers to an 'individual (the data subject)'

- **Country 1** refers to 'data subjects'

- For **Country 12** 'personal data' includes not only information about a 'natural person', but also information about 'a legal person, body or association'

- None of the jurisdictions reported any difficulty in defining or interpreting this term.

'directly or indirectly'

- No formal definitions.
- Three jurisdictions use the term 'directly or indirectly' within their definition of the term personal data. [**Country 17, Country 1, Country 12**]

    - **Country 1** makes the meaning as clear as possible in its use of this term: 'Data are only indirectly personal for a controller, a processor or a recipient of transmission when the Data relate to the subject in such a manner that the controller, processor or recipient of a transmission cannot establish the identity of the data subject by legal means.'

    - **Country 12** expands on the use of this term within its definition of personal data; information is personal data if it relates 'directly or indirectly by reference to any other information, including a personal identification number.'

    - It is clear that most of the terms highlighted from the Directive are not defined within the EU jurisdictions surveyed. The terms 'relating to', 'natural person', and 'directly or indirectly' are not formally defined by any participating Member States.

Group 2

There is similarity between the EU countries in Group 2 and the EU Member States in Group 1. This is explicable as the jurisdictions concerned are either aspiring to join the EU or wish to have trading relationships which feature the transfer of data - they wish to have adequate protection and therefore have complied with the terms of the Directive in their own legislation.

'personal data', 'data' and 'information'

- All non-EU jurisdictions surveyed have a definition of 'personal data' or 'personal information' in their national data protection legislation.
- Most jurisdictions have a formal definition of at least one of the terms 'data' or 'information'. [**Country 35, Country 27, Country 33, Country 34, Country 29**]

- **Country 27's** definition of information refers to 'public and private information at the disposal of the state and municipal institutions to which a person has access in the procedure laid down by this law.'

- **Country 35, Country 33 and Country 34** all refer to the method of processing in their definitions of data. For example, 'Information recorded in a form in which it can be processed by equipment operating automatically in response to instructions given for that purpose.' [**Country 34**]

- **Country 22** specified that it makes no distinction between 'data' and 'information'.

- All but two jurisdictions reported no difficulty in defining or interpreting the terms 'data' or 'information'. [**Country 35, Country 27**]

    - **The Country 35** reported difficulty due to another term - 'relevant filing system'.

'relating to'

- No formal definitions.
- Only one jurisdiction does not use this term in their definition of personal data.

    - **Country 20** substitutes the term 'linked to'.

- No jurisdictions reported any difficulty in defining or interpreting this term.

'identified or identifiable'

- Only one jurisdiction [**Country 20**] does not use at least one of these terms in its definition of 'personal data'.

    - **Country 22** specifically excludes 'consolidated data of a statistical nature, from which the data subject cannot be identified' from its definition of personal data.

    - **Country 27** has specifically used the definition from Article 2(a) of the Directive as a definition of this term: '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'.

    - **Country 35, Country 34 and Country 33** all use a very similar phrase, based upon The UK Data Protection Act, which determines the span of the term 'identified' in the following way:
      "data which relate to a living individual who can be identified - (a) from those data (b) from those data and other information which is in the possession of or is likely to come into the possession of, the data controller" [**Country 33**]

- No jurisdictions reported any difficulty with defining or interpreting this term.

'natural person'

- No jurisdiction has a formal definition of this term.
- Three jurisdictions use this term within their definition of personal data [Country 20, Country 27, Country 29]
    - **Country 22** uses the term 'living data subject' in its definition of personal data, but refers to 'natural person' in its definition of 'person': 'person means any natural person or any public or private body corporate body whether or not it has legal personality and includes the Government of the Republic.'
    - **Country 35, Country 34 and Country 33** all use 'living individual'.
- Two jurisdictions have a Civil Code that covers this term, [**Country 27, Country 29**]. **Country 27** was the only jurisdiction of both groups to report any difficulty in defining or interpreting this term.

'directly or indirectly'

- No formal definitions.
- Only one jurisdiction uses this term in their definition of 'personal data' [**Country 27**].

Few difficulties were reported in defining or interpreting any of these terms within the non-EU jurisdictions surveyed. Very few jurisdictions give formal definitions of the terms within their domestic legislation. The terms 'relating to', 'directly or indirectly' and 'natural person' were not defined by any of these jurisdictions.

Group 3

These non-European jurisdictions also demonstrated similarity both in their lack of definitions of the terms used, and their confidence in understanding them.

'personal data', 'data' and 'information'

- All jurisdictions surveyed have a definition of 'personal data' or 'personal information' in their national data protection legislation.
- Only one jurisdiction gave a formal definition of the terms 'data' or information'.
    - **Country 36** gave definitions of information and data in terms of the definition of 'personal information' and 'record':

      "*record* means: (a) a document; or (b) a database (however kept); or (c) a photograph or other pictorial representation of a person;"

- Two jurisdictions reported difficulty in defining or interpreting the terms 'data' or 'information'. [**Country 36, Country 40**]

'relating to'

- No formal definitions.
- None of the jurisdictions use this term in their definition of 'personal data'.
- 2 jurisdictions reported difficulty in defining or interpreting this term. [**Country 36** - whose discussion centred around primary and related purposes in its Privacy Principles rather than information being 'about' something - and **Country 40**]

'identified or identifiable'

- All three jurisdictions either use these terms, or refer to identity, in their definitions of 'personal data'.
    - **Country 36** uses the phrase 'whose identity is apparent'
- Only one jurisdiction reported any difficulty with defining or interpreting this term. [**Country 40**]

'natural person'

- Only one jurisdiction gave a formal definition of this term in their questionnaire answer.
- No jurisdictions use this term within their definition of personal data.
    - **Country 40** uses this term in its definition of 'individual', defined to mean "a natural person, other than a deceased natural person".
- No jurisdiction reported any difficulty in defining or interpreting this term.

'directly or indirectly'

- No formal definitions.
- None of the jurisdictions in this group use this term in their definitions of 'personal data'.

Within this group, few difficulties were reported in defining or interpreting any of these terms.
All three have defined personal information as being 'information' rather than 'data', which is 'about', rather than 'relating to', an individual. All three definitions also refer to the individual being 'identifiable', or their identity being 'apparent'. Where the same terms are used as appear in the EC Directive, these terms are not formally defined. There are no uses of the term 'directly or indirectly', and no references to a 'natural person' in the definitions of personal data.

Although there is independent use of some of the same terms between these jurisdictions, there is exclusion of different types of information from the definitions, and in **Country 36**'s case an entirely different set of criteria and different terms is used. The definitions appear to be less specific than those used by the Data Protection Authorities compliant with the Directive, possibly because there are no imposed stipulation to which these Authorities must independently adhere. Despite this, there still appears to be a great deal of overall similarity of wording between the definitions of personal information.

B2: Key findings
Between the jurisdictions surveyed, there is confidence in understanding the terms found in the Directive, demonstrated by a lack of need for definition or by a lack of difficulty in defining or interpreting the terms. There is a large degree of similarity in defining personal data, with some consistency in use of terminology.

B3: The concept of 'personal data' in practice

If we recall, question 1 of the first questionnaire was designed to test whether it is possible to draw up a list of data types which are always, sometimes or never personal data. Respondents were presented with a list of 35 data types (see Appendix 3) and asked to classify each type as *always*, *sometimes* or *never* personal data. In this section, we discuss the results to this question.

During the analysis of the first questionnaire, we observed a striking level of inconsistencies between countries. This is discussed in details below. However, it is important to stress that it was recognised that a possible explanation for an inconsistent classification of data types was simply a fundamental difference in interpretation of the question (especially where the native language of the respondent was not English). If we recall, question one was deliberately worded to leave open the question of whether the data types referred to an individual who had already been identified. Similarly, there were a few data types which had seemed to cause some confusion (Medical history of family members, Family portrait, Death details and Vehicle description).

Therefore, it was important to consider whether respondents had been interpreting the questions in different ways. One of the main aims of Part A of Questionnaire 2, therefore, was to address the issue of question interpretation. Respondents were presented with summary tables of their classification of data types and asked specific questions which attempted to clarify anomalies and confirm themes. By probing in this way, it was hoped that respondents would tell us how they had interpreted the original question and whether they wished to make changes to their data classifications.

Analysis of the second questionnaire revealed a very low frequency of response alterations, indicating that respondents were generally happy with their original answers. Where respondents indicated that they wished to change their classifications of data types, two issues emerged as having prompted reconsideration. Firstly, there were two respondents (**Country 29** and **Country 40**) who wondered whether they had become confused about the wording of the question and explained their own interpretation. **Country 40** had assumed that the data related to an identified individual in the first instance. Similarly, **Country 29** stated that '[i]t is difficult for us to decide what type of information is personal data, without specification if this information relates solely to the identifiable person' (Q2, A, Q1). Secondly, some changes were made where respondents had simply reconsidered the question in new contexts or perspectives, often after probing. So, for example, the Country 35 moved 'Football team Supported' from the *never* to the *sometimes* category, explaining that '…on reflection there are contexts when sometimes this can be personal data…'. Country 33 moved Sexual orientation, Religion and State benefits received from the *always* category to the *sometimes* category because on reflection it was felt that '…more data will be needed to identify an individual'.

The results presented in the sections which follow take into account the (few) alterations made in the second questionnaire.

An inconsistent classification of data types

Analysis of the responses to Question 1 (and confirmed in Questionnaire 2 follow-ups) suggest that there is significant disagreement as to the status of all data types listed as 'always', 'sometimes' or 'never' personal data (see Graph 1). For the majority of data types, there seems to be a consensus that the data *is capable* of being personal data but there is some dispute as to whether this occurs in all or only some circumstances. For just under a third of data types[137], there is an element of doubt that the data is even capable of being personal data, although the use of the 'never' category is restricted.

Although there is no complete consensus for any of the data types, we can distinguish between data types that are classed as 'always' or 'sometimes' by the majority of respondents. Table 1 shows the data types arranged according to the majority classification, i.e. where a classification of 'always' or 'sometimes' is made in 50% or more of cases.

The data types are listed in the table according to the number of 'always' classifications received, in descending order. So, for example, 'national registration number' received the highest proportion of 'always' classifications (78%) and those data types in the bottom cell (including shoe size and death details) received the lowest proportion of 'always' classifications (33% in each case). The data types marked with an asterisk (*) are those data types which received at least one 'never' classification. These data types appear towards the bottom of the table where there is more uncertainty as to whether the data will be personal.

The data types can be divided into four groups, ordered according to the likelihood of being personal data.

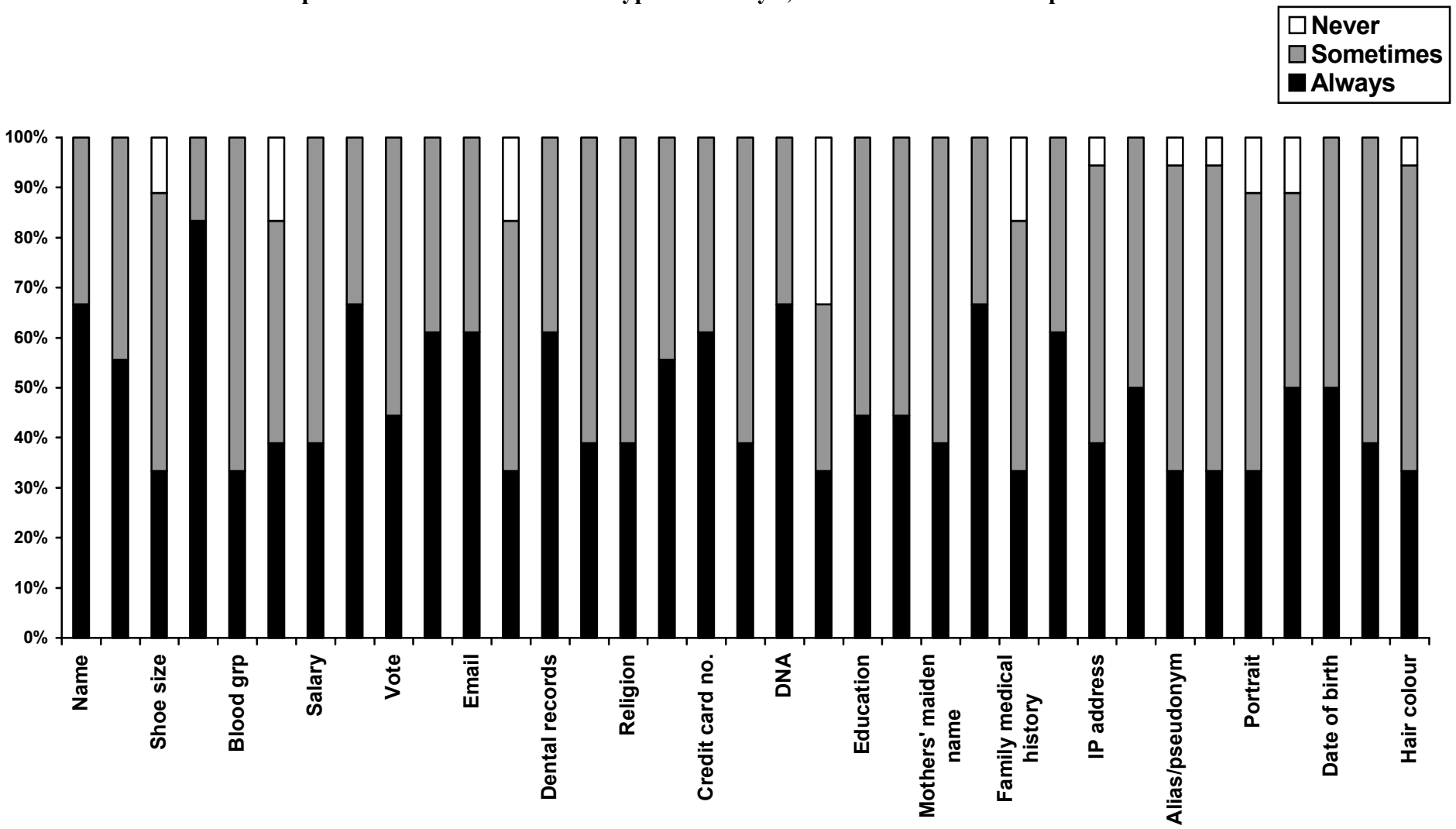
A: Strong likelihood of being personal data
B: More likely than not to be 'always' personal data
C: Will be personal data in some circumstances.
D: Will be personal data in limited circumstances, and possibly not at all.

[137] Shoe size, countries visited, TV viewing habits, death details, family medical history, computer IP address, chat room alias, football team supported, family portrait, vehicle description and hair colour

Table 1:

	Data Type	Percentage classifications 'always'
A	National registration number	75% +
	Name	66% - 74%
	Head and shoulders photo DNA profile Fingerprint	
B	Car registration Email address Dental records Credit card details Bank account details	51% - 65%
	Telephone number Parents' names	
	Date of birth State benefits received	50%
C	Vehicle description * Party voted in last election Education Ecommerce transactions	34% - 50%
	Salary details Sexual orientation Religion Addiction history Mother's maiden name CCTV image Countries visited in last 5 years * Computer IP address *	
D	Shoe size * Blood group TV viewing habits * Family medical history * Chat room alias * Football team supported * Family portrait * Natural hair colour * Death details *	33%

Graph 1: % classification of data types as 'always', 'sometimes' and 'never' personal data



Data classification by country

Despite the similar definitions of 'personal data' used by Data Protection Authorities around the world, in practice there are differences between what is and is not considered to be 'personal data' in each jurisdiction. The apparent harmony suggested by the broadly consistent definitions of 'personal data' is not manifested in practice.

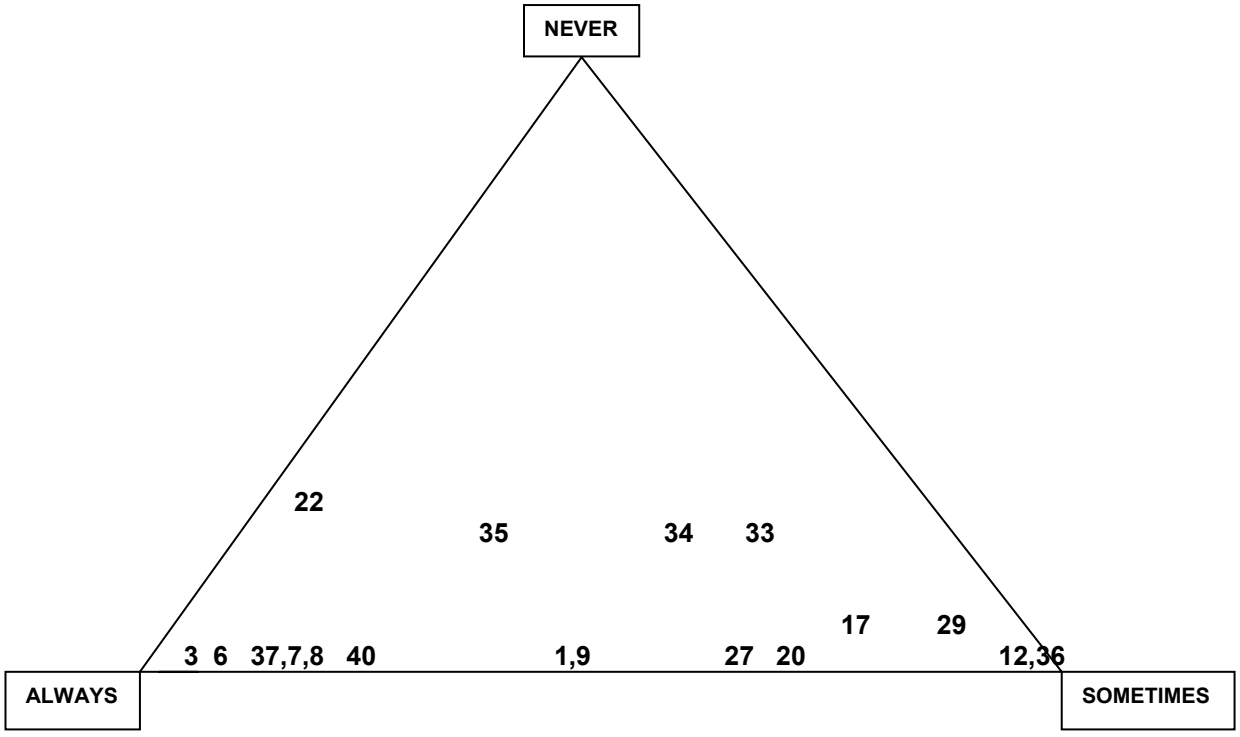
In Questionnaire 1, Data Protection Authorities were asked to indicate whether various pieces of data would be classed as personal data *always*, *sometimes* or *never*. The differences in the responses are illustrated in the diagram below.

UNDERSTANDING THE DIAGRAM

If an authority indicated that all the pieces of data were '*never'* personal data, then the country number would appear on the table at the top of the triangle next to the NEVER box. If they answered that all of the pieces of data were '*sometimes'* personal data, they would appear next to the SOMETIMES box. Similarly, '*always'* responses would appear next to the ALWAYS box.

If an authority indicated that a third of the data was '*always'* personal data, a third was '*sometimes'* personal data, and a third was '*never'* personal data, the country number would appear in the centre of the triangle, equidistant from each corner.

If an authority had indicated that half of the data was '*never'* personal data, and half of the data was '*sometimes'* personal data, the country number would appear halfway on the line between NEVER and SOMETIMES.



The diagram is based on responses to Questionnaire 1, Question 1. In the analysis that follows, any modifications suggested by the respondents in Questionnaire 2, Section A, are taken into account. Note that none of the modifications suggested in Questionnaire 2 would significantly change the position of any country in the diagram.

The diagram clearly demonstrates that there are divergences in approach:

- The only consensus was a reluctance to state that a piece of data can *never* be personal data.
- Some Data Protection Authorities suggested that all (or most of) the pieces of data are *always* personal data. **Country 3** was the most extreme examples of this.
- Some Data Protection Authorities suggested that all (or most of) the pieces of data are *sometimes* capable of being personal data. **Country 36** and **Country 12** were the most extreme examples of this.
- The other countries fell at varying points between these two extremes.
- **Country 22** felt that data are either *always* or *never* personal data - data cannot be personal data *sometimes*.
- Even those Data Protection Authorities that appear very close in the table did not necessarily give the same responses for each piece of data (for example **Country 20** and **Country 27**).

What the diagram clearly demonstrates is the wide range of responses to Questionnaire 1, question 1. This suggests that there is little consistency in approach: Data Protection Authorities appear to adopt inconsistent classificatory strategies when answering the question whether a particular data type will *always*, *never* or *sometimes* constitute 'personal data'.

Further discussion about the distribution of *always*, *sometimes* and *never*

classifications

The distribution of classifications varies considerably across jurisdictions. Looking at Graph 2, we see that for a number of countries there is a clear tendency to place all data types (or the vast majority) in the same category. **Country 3**, **Country 6**, **Country 7**, **Country 8** and **Country 37** have reported that more than 90% of data types are *always* personal data. Similarly, **Country 36**, **Country 12**, **Country 29** and **Country 33** have stated that more than 90% of data types are personal data in only some circumstances. The remaining countries demonstrate a more dispersed categorisation of data types, suggesting that there are clear differences between the different examples of data given. **Country 17**, **Country 22**, **Country 29**, **Country 33**, **Country 34**, **Country 35** and **Country 40** have all made use of the *never* category.

Table 2 enables us to distinguish between countries which have used either the *always* or *sometimes* classifications in the majority of cases (i.e. in more than 50% of cases). The table suggests that there is a lack of conceptual agreement between members of each of the three groups.

Within the EU members (Group 1), there is a clear lack of consensus. For **Country 3**, **Country 6**, **Country 7**, **Country 8** and **Country 1**, the data types are more likely to be classed as *always* personal data, although it should be noted that for **Country 1** this tendency is slight (the proportion of always vs. sometimes responses is 51:49). **Country 12**, **Country 9** and **Country 17**, on the other hand, state that the majority of data types will be personal data in some circumstances. It is interesting to note in particular the apparent disagreement between the Country 6 states (compare **Country 9** with **Country 7**, **Country 8** and **Country 6**).

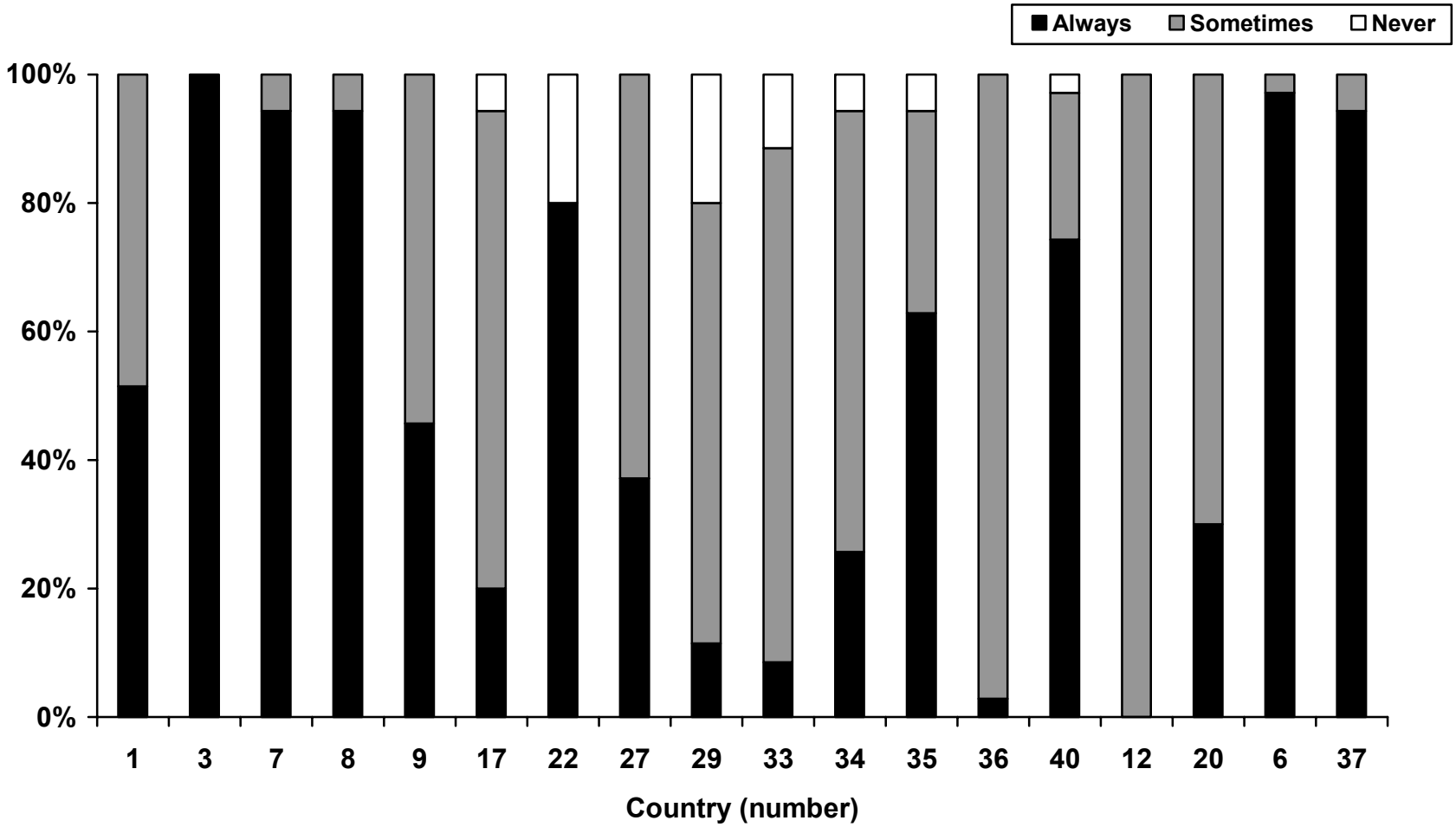
However, there does seem to be a little more consistency between the non-EU countries with requirement of capability (Group 2). **Country 27**, **Country 29**, **Country 34**, **Country 33** and **Country 20** have all stated that the majority of data types given will be personal data in some circumstances. **Country 22** and the **Country 35** are in disagreement, stating that the majority of data types given are 'always' personal data.

The non-EU countries with no requirement of compatibility with the directive (Group 3) seem to display the weakest consistency amongst themselves. **Country 37** has classified more than 90% of the data types as *always* personal data. In contrast, **Country 36** has stated that the majority of data types as personal data in only some circumstances. **Country 40** have classified the data types as *always* personal data in the majority of cases, but distinguishes between a few data types which are *sometimes* and *never* personal data.

Table 2:

Mostly *always* (country group)	**Mostly *sometimes* (country group)**
Country 3 (1)	Country 9 (1)
Country 6 (1)	Country 36 (3)
Country 7 (1)	Country 17 (1)
Country 8 (1)	Country 27 (2)
Country 37 (3)	Country 12 (1)
Country 22 (2)	Country 29 (2)
Country 40 (3)	Country 34 (2)
Country 35 (2)	Country 33 (2)
Country 1 (1)	Country 20 (2)

Graph 2: Classification of data types across all countries



A closer look at the classification of data types with groups 1, 2 and 3

Graphs 3, 4 and 5 show the classification of data types for each of the groups of jurisdictions. Examination of these graphs gives an extra dimension to our understanding of the inconsistencies between and within jurisdiction groups.

Group 1 (see Graph 3):

Analysis of the data classifications for EU countries shows that all data types were considered to be *always* personal data by at least half of the group (contrast this finding with group 2). The strongest consensus was achieved for the following data types: Name, Telephone number, National registration number, Head and shoulders photo, Car registration, Email address, Parents' names, Credit card number, DNA profile, Mother's maiden name, Fingerprint, Bank account details, Computer IP address, Alias/pseudonym, Date of birth and CCTV image. Only Death details and Family portrait were considered to be *never* personal data, and in both cases this classification was awarded by only one country.

Group 2 (see Graph 4):

The distribution of classifications across data types seems to be much more inconsistent for the non-EU countries with the requirement of compatibility with Directive 95/46/EC. There is some agreement that the following data types will only be personal data in some circumstances (i.e. they receive no *always* classifications): Shoe size, TV viewing habits, Mother's maiden name, Alias/pseudonym, Football team supported, CCTV image and Natural hair colour. In addition, the majority of countries within this group agree that National registration number, Blood group, Dental records, DNA profile, Fingerprint, and Bank account details will be more likely to be *always* personal data (having received more than 50% 'always' classifications). One of the most striking features of the responses of this group is the comparatively extensive use of the *never* classification. Data types which are said to be incapable of being 'personal data' by some members of this group include: Shoe size, Countries visited in the last 5 years, TV viewing habits, Death details, Family medical history, Computer IP address, Alias/pseudonym, Football team, Family portrait, Vehicle description and Natural hair colour.

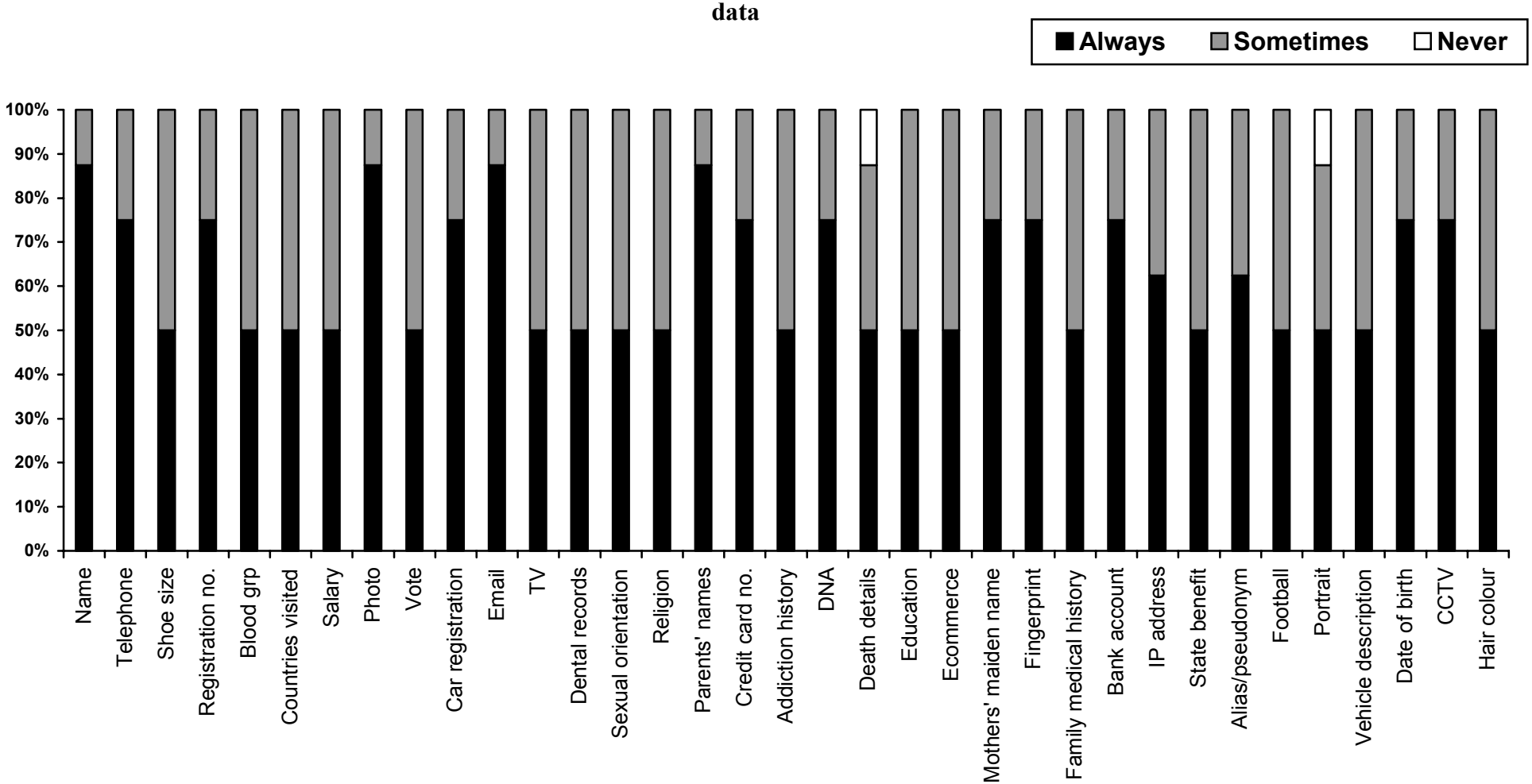
Group 3 (see Graph 5):

Since there are only three members of this group (non-EU members with no requirement of compatibility), one should exercise care when interpreting this final graph. One must keep in mind that **Country 37** and **Country 36** placed more than 90% of data types in opposing categories (*always* and *sometimes* respectively). **Country 40** falls between the two, with approximately 75% of data types being classified as *always*.

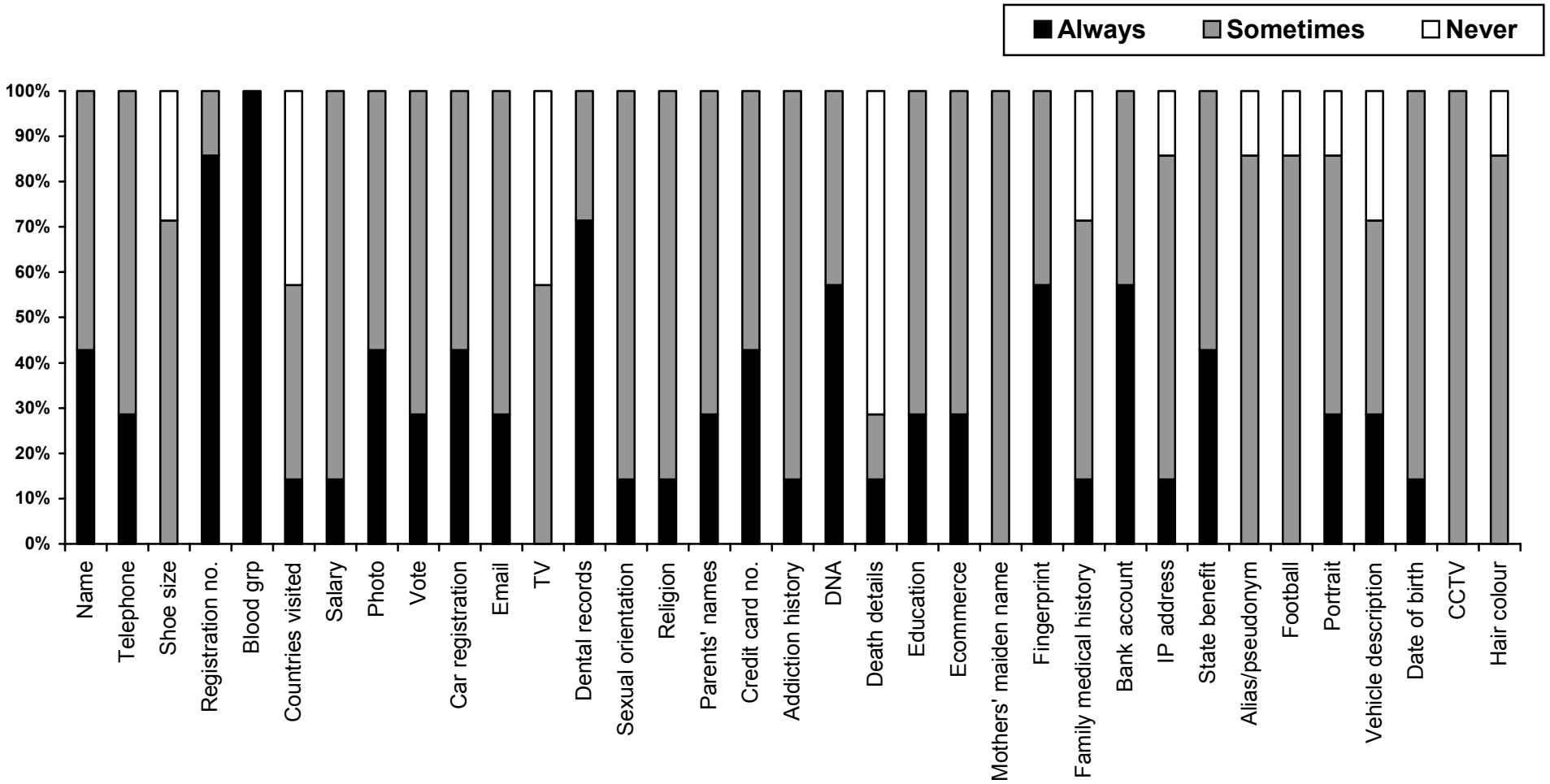
We can, however, make some interesting distinctions between the approaches of the three countries. All three countries agree that Family portrait and CCTV image will be personal data only *sometimes*, and that National registration number will *always* be personal data. **Country 37** and **Country 40** agree that all of the data types are always personal data except for: Telephone number, Parents' names, Death details,

Mother's maiden name, Computer IP address and Alias/pseudonym (**Country 40** says that these are *sometimes* personal data). All three countries disagree on the classification of Family medical history.

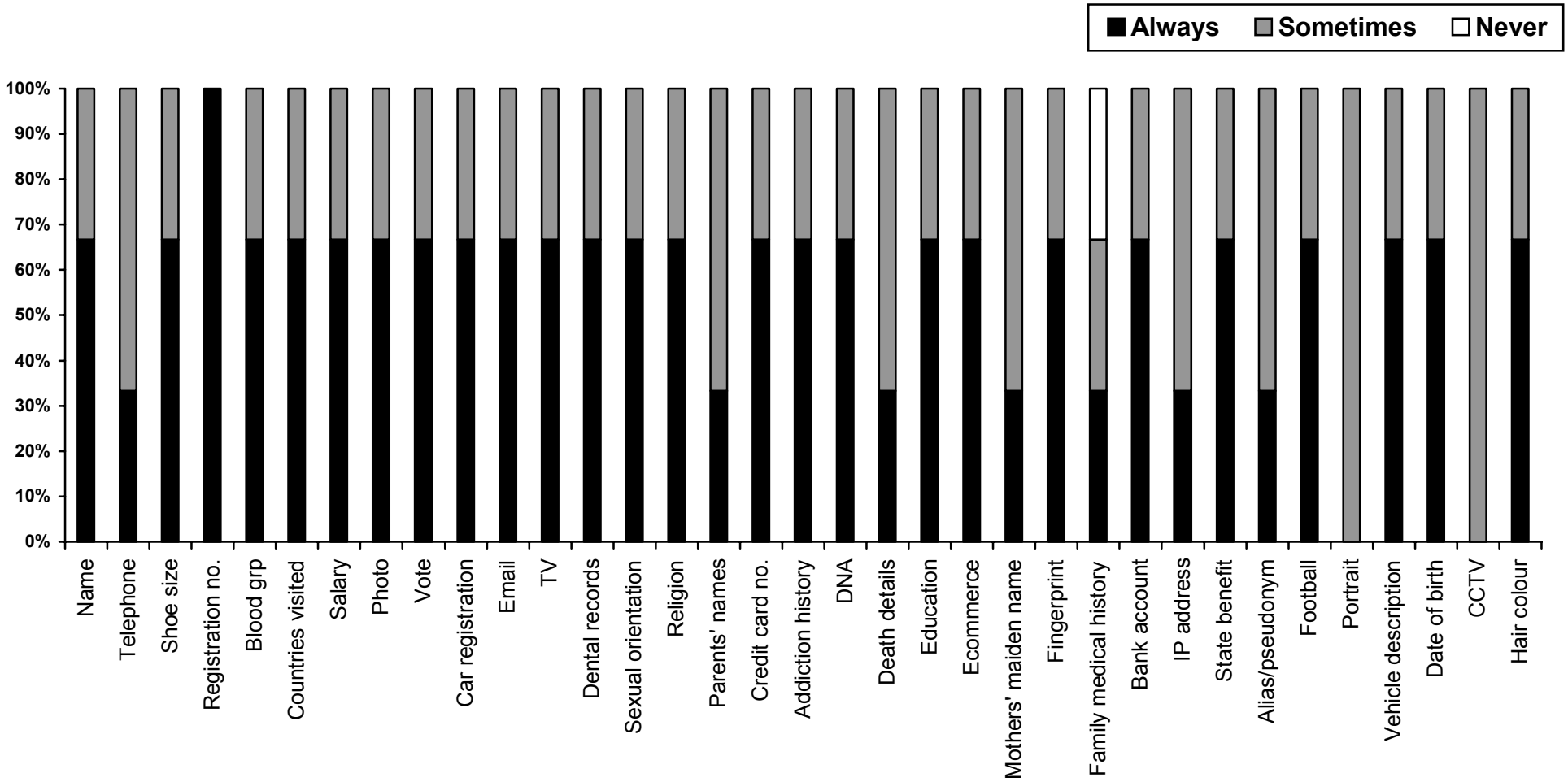
Graph 3: Group 1 (EC members) : % classification of data types as 'always', 'sometimes' and 'never' personal data



Graph 4: Group 2 (Non-EC members with requirement of compatibility with Directive 95/46/EC) : % classification of data types as 'always', 'sometimes' and 'never' personal data



Graph 5: Group 3 (Non EC members) :
% classification of data types as 'always', 'sometimes' and 'never' personal data



A further dimension: Personal data filing systems

In addition to the exploration of the concept of personal data in Question 1, Question 6 (also of Questionnaire 1) asked respondents to state whether each of 13 examples of filing systems would be considered to be 'personal data filing systems' 'always', 'never' or 'sometimes'. Responses were followed up in Questionnaire 2 in order to check responses and seek clarifications.

NB. Country 37 and Country 40 did not complete this question because they do not use the concept 'personal filing system' within their respective jurisdictions. Therefore, results are only presented for jurisdiction groups 1 and 2 in this section. Country 36's responses are included within the general observations.

Looking at Graph 6, which includes all countries, we can see that there is significant disagreement as to the classification of filing systems. Card indexes, Electronic databases, Electoral registers, Registers of births/deaths/marriages, Membership lists of voluntary organisations and Telephone directories were more likely to be classed as *always* 'personal filing systems'. Most respondents agreed that the following filing systems would be classed as 'personal filing systems' in only certain circumstances: Organisational filing systems, Photo albums, Diaries, Archived minutes of meetings, CCTV footage and Organisational websites. The majority of respondents agreed that a Newspaper is highly unlikely to be classed as a 'personal filing system'

We see, then, a wide dispersion of responses across the *always*, *sometimes* and *never* responses, suggesting a number of different interpretations and applications of the concept 'personal filing system'. Newspaper and Diaries are the only two examples to have no *always* classifications, suggesting that these will only be classed as personal filing systems in special contexts, and often not at all. In contrast, respondents viewed Card index as the only example to *always* be *capable* of being a personal filing system (having received no *never* classifications).

Although there is a general lack of consensus as to the classifications of the filing systems, it is possible to draw up a list of those filing systems which are *more likely* to be considered to be 'personal filing systems'. Table 3 lists filing systems according to the majority of classifications received (i.e. 50% or more responses). We see that Newspaper is the only filing system to receive a majority of *never* classifications, with more than 70% of respondents agreeing on this classification. Of those filing systems which were more likely to be classed as 'personal filing systems' (mostly *always*), three types of filing system emerged as *most* likely (receiving more than 75% *always* classifications): Electronic databases, Electoral registers and Registers of births/deaths/marriages.

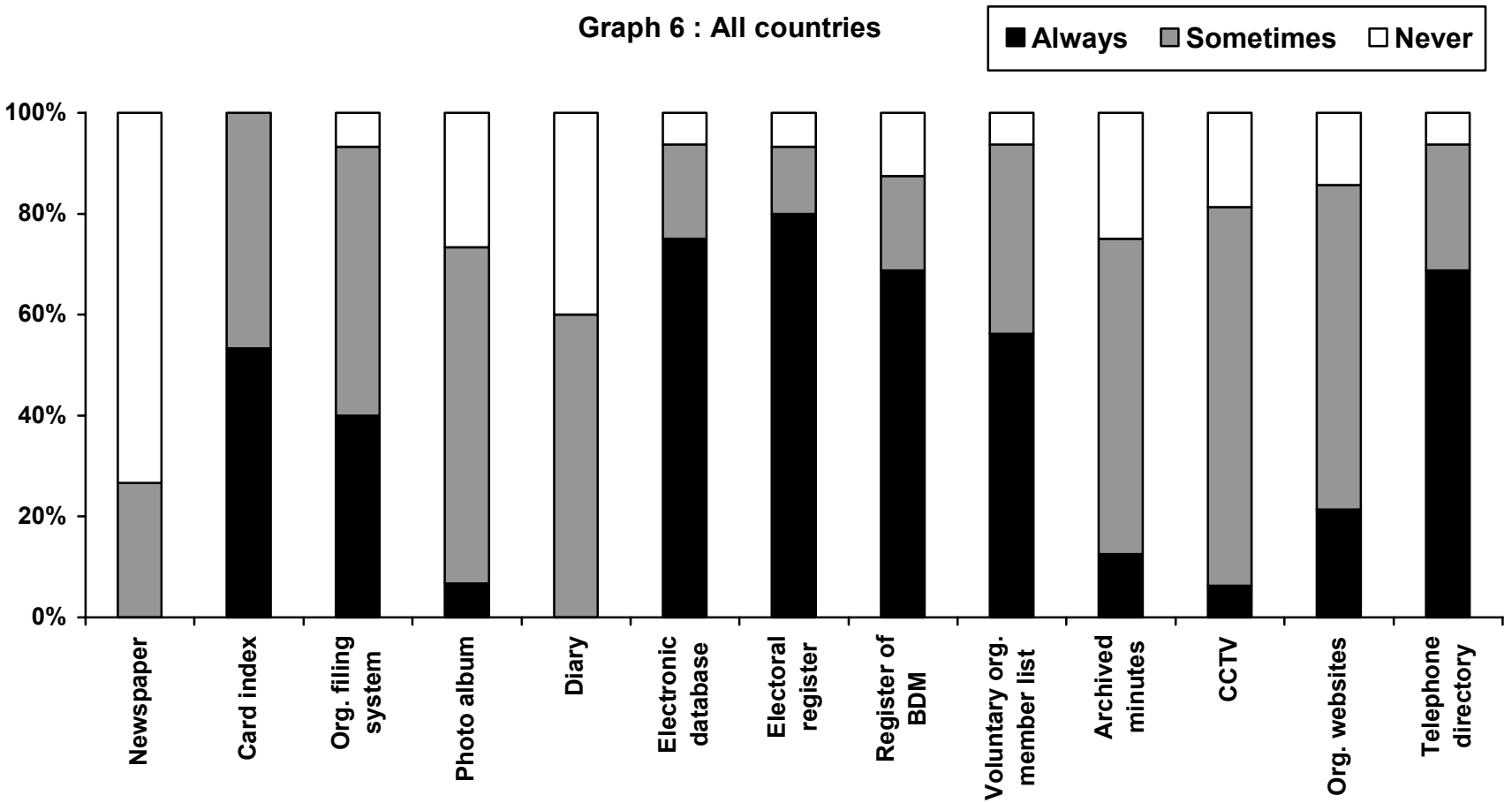
For the EU members (Group 1- see Graph 7), a strong majority agreed that Electronic databases, Electoral registers and Registers of births/deaths/marriages would be classed as 'personal filing systems' in all circumstances (i.e. at least 75% of countries said these would *always* be 'personal filing systems'). Countries in this group also reached a high level of agreement that CCTV footage would only be classed as a 'personal filing system' in some circumstances. In contrast, for the non-EU members (Group 2- see Graph 8), the classification of filing systems was more varied. A high level of agreement was reached for Organisational filing systems

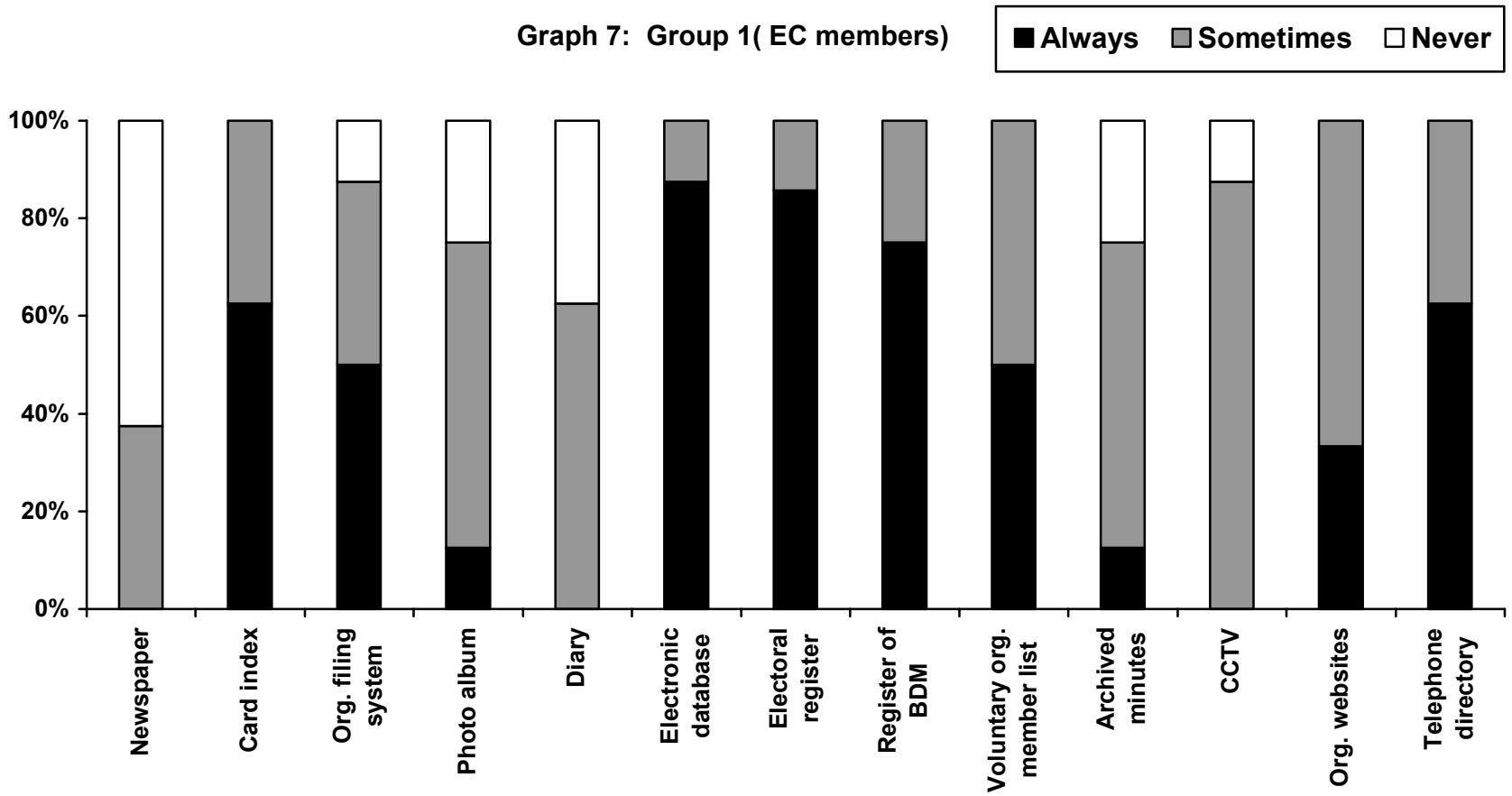
(*sometimes* being regarded as personal filing systems) and Newspapers (*never* personal filing systems), but there was little agreement as to the classification of the other filing systems. Indeed, the non-EU countries were more inclined to use the *never* classification, demonstrating a significant difference in approach between EU and non-EU countries.

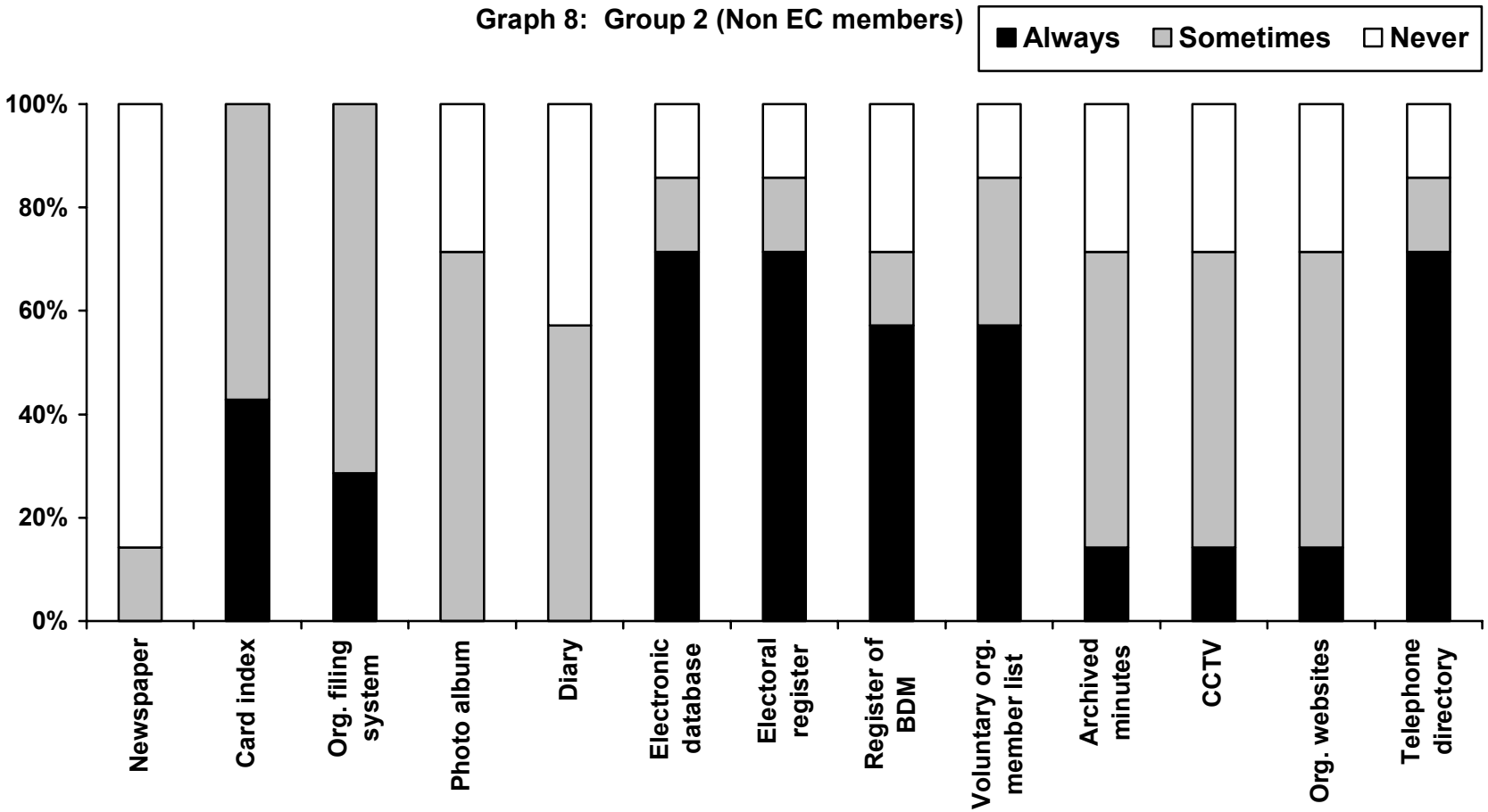
EU members take a different approach to non-EU members with regard to the operationalisation of the concept of a 'personal filing system'. EU members tend to take a more consistent approach, with most of the disagreement being attributable to the *always* vs. *sometimes* division. In contrast, non-EU members tend to take a more diverse approach and are more likely to operationalise the concept in different ways.

Table 3:

Mostly *Always*	**Mostly *Sometimes***	**Mostly *Never***
Card index	Organisational filing system	Newspaper
Electronic database	Photo album	
Electoral register	Diary	
Register of Births, Deaths, Marriages	Archived minutes of meetings	
Membership list of voluntary organisations	CCTV	
Telephone directory	Organisational website	


Graph 6 : All countries
Always
Sometimes
Never
100%
80%
60%
40%
20%
0%
Newspaper
Card index
Org. filing system
Photo album
Diary
Electronic database
Electoral register
Register of BDM
Voluntary org. member list
Archived minutes
CCTV
Org. websites
Telephone directory


Graph 7: Group 1(EC members)
Always
Sometimes
Never
100%
80%
60%
40%
20%
0%
Newspaper
Card index
Org. filing system
Photo album
Diary
Electronic database
Electoral register
Register of BDM
Voluntary org. member list
Archived minutes
CCTV
Org. websites
Telephone directory


Graph 8: Group 2 (Non EC members)
Always
Sometimes
Never
100%
80%
60%
40%
20%
0%
Newspaper
Card index
Org. filing system
Photo album
Diary
Electronic database
Electoral register
Register of BDM
Voluntary org. member list
Archived minutes
CCTV
Org. websites
Telephone directory

In Conclusion

Given the divergences in approach to the operationalisation of the concept 'personal data', it is interesting to consider why the jurisdictions reported that there were no problems in understanding the terms used in the formal Directive definition of 'personal data'. It is possible that, despite the differences *between* jurisdictions, there are clear concepts operating *within* jurisdictions.

Two questions arise from this analysis:

1. Which ideas/understandings guide interpretations within jurisdictions?
2. Is there a consistent application of a clear and express conceptual understanding of personal data (and related key terms) within jurisdictions?

We go on to address these questions in B4.

B3: Key Findings

Despite the apparent similarities between the 'on paper' formal definitions described in B2, the Data Protection Authorities demonstrate a lack of consistency at the level of operationalisation of the concept 'personal data'. These divergences in approach are found both within and outside of the EU.

B4: What Concepts of Personal Data are at Work?

The inconsistencies in the classification of different data types (illustrated in B3) indicate differing approaches to the interpretation of relatively similar formal definitions (see B2). While some of this inconsistency in approach could be explained by supposing different interpretations of the questions, the degree of inconsistency found seems unlikely to be attributable to this alone.

Terms identified as 'key' to an understanding of the definitions of '**personal data**' operative within the jurisdictions – '**relating to**', '**identified or identifiable**', '**person**' – are all terms which may be described as ambiguous in some crucial respects (see the literature review in A2) and none of which benefit from more detailed definition within either Directive or domestic data protection legislation. It may be that the inconsistency observed between responses is attributable to this lack of clarity; allowing different ideas about how these terms are to be interpreted to manifest within practice.

Simply observing ambiguity within the key terms, and inconsistency between jurisdictions in their classification of data types as 'personal data', might indeed be considered suggestive of some degree of *confusion* over the appropriate interpretation of those key terms that together define the concept of 'personal data'. These observations are however equally consistent with the thesis that, while interpretations vary between jurisdictions, each individual jurisdiction is operating with a clear and unambiguous concept of personal data. Rather than 'confusion' then, these observations would rather indicate some level of 'conflict' over the appropriate concept of personal data to be operationalised through interpretation of the formal definitions[1].

If individual countries are operating with clear and unambiguous concepts of personal data then we can at least observe that these appear to be distinct from their neighbours. This appears to be as true for those countries within Group 1 as it is between countries in Groups 1, 2 and 3. If we are seeking to understand how the term 'personal data' is understood by these various countries, and how their formal definitions are interpreted in practice, then it is appropriate to try and understand more about the concepts of 'personal data' that may be informing their day to day operations. It will be within a particular concept of 'personal data' that a particular position on the key terms already identified may be justified.

Which concepts of 'personal data' appear to be informing the different approaches toward data classification?

As well as asking countries to identify whether they considered specific data types to constitute personal data '*always*', *'never'*, or *'sometimes*', we also asked them to indicate the *circumstances* in which they would consider particular data types to constitute personal data. An understanding of how the countries *apply* their formal definitions (of personal data) may contribute to clarification of how they *understand* these respective definitions. Their understanding will be linked to the particular concept of personal data that they hold. Application of a definition inevitably involves a process of interpretation informed by a conceptual understanding of the subject

[1] 'Conflict' is not the only available explanation. Different jurisdictions may simply be operating with different interpretations of personal data to take account of the contingent circumstances of the jurisdiction.

defined. This process of elucidating operative concepts was additionally informed by considering the justifications that were offered for the decisions made on the information contained within a number of scenarios (See Q2, Section B).

Comparing the reasons that different countries offered for classificatory decisions certainly gives the impression that they understand the term 'personal data' to mean slightly different things. They may be said to be operating with (interpretations of formal definitions informed by) different concepts of personal data.

That countries are operating with different concepts of 'personal data' may be most succinctly illustrated through their responses to one question in particular. They were asked to give their opinion on the following three statements:

a) Information can only be personal data if it can identify an individual

b) Information can only be personal data if it does not identify an individual but can affect an individual in a different way

c) Information can only be personal data if it both identifies and affects an individual

It may be noted how these different statements would strike a different attitude to the nature of the relationship that must exist between data and individual before that data could be said to be 'personal'.

Broadly speaking, respondents tended to agree with only one of the above statements and disagreed with the alternates. While the statements did not enjoy equal approval each did receive some support and this observation may be used to divide respondents into three broad 'conceptual' camps.

1. Those that require personal data to be capable of identifying an individual.
2. Those that require personal data to relate to an individual in some way (that does not presuppose the possibility of their identification *from the information in question*)
3. Those that require personal data to *both* be capable of identifying an individual *and* to relate to them in some other way.

While this division is admittedly crude it does point to explanation of some of the inconsistencies in responses identified in B3. It does indeed suggest that some of the inconsistency may be indicative of conceptual conflict over a proper understanding of the term 'personal data'.

Unfortunately, this project does not allow an especially rigorous analysis of the different concepts that may be at work. However, comparing the responses of countries that appear to align themselves more closely with each of these three statements does allow us to explore the idea that they are operating with differing concepts a little further.

1. Identificatory potential as prerequisite

If a country has agreed with the statement that 'Information can only be personal data if it can identify an individual' one would expect it to be operating with a concept of personal data that *requires* a piece of information, if it is to qualify for description

as personal data, to be capable of identifying an individual. There are some indications that at least some countries are indeed *operating* with this concept of personal data informing their interpretation of the formal definitions of personal data in practice. These countries are isolating a particular type of relationship between information and data subject: a relationship of identification, and then using the presence or absence of this particular relationship to inform their classification of data as 'personal'.

Country 9 is one example of such an authority. The idea that the 'identificatory potential' of the information is crucial to the classification of the information as 'personal data' may be seen to be at work within its classification of data types listed within question one of the first questionnaire.

Always	Never	Sometimes
Name		Shoe size
Home telephone number		Blood group
National registration number		Countries visited in the last 5 years
Head and shoulders Photograph		Salary details
Car registration/licence plate number		Political party voted in the last election
Email username and password		TV viewing habits
Parents names		Dental record
Credit card number		Sexual orientation
DNA profile		Religion
Details of time, place and cause of death of data subject		History of addiction
Fingerprint		Education/qualifications
Medical history of family members		E-commerce transactions
Bank account details		Mother's maiden name
Computer IP address		State benefit received
Date of birth		Alias/pseudonym used in internet chat room
Still image taken from CCTV		Football team supported
		Family portrait (painting)
		Vehicle ownership (make, model and colour)
		Natural hair colour

The most obvious example of a data type whose categorisation might initially resist explanation on the grounds of its identificatory potential would be 'Dental Record'. When asked what they thought was the difference between 'DNA profile', 'Fingerprint' and 'Dental Record' **Country 9** responded,

> "[T]he difference depends on the possibility of anonymisation. I have taken the view that there is no way to anonymise a DNA profile or fingerprint. So more generally all the distinctions I made in the above table depend on this possibility in both ways: If there is a way to anonymise DNA data or fingerprint they lose the characteristics of personal data as do dental records which have been anonymised. On the other hand if dental records cannot be anonymised they are always personal data."

Country 9 is not alone in apparently privileging this relationship of 'identificatory potential' between data and data subject within their concept of 'personal data'. **Country 27** also agreed with the statement that 'information can only be personal data if it can identify an individual' (Q2,C,q2). It may be noted however that, despite

both agreeing with this statement they do *not* agree on what data types would *always* be personal data.

Country 27 classified the data types within the first questionnaire thus:

Always	Never	Sometimes
Name		Home telephone number
National registration number		Shoe size
Head and shoulders Photograph		Blood group
Dental record		Countries visited in the last 5 years
Parents names		Salary details
DNA profile		Political party voted in the last election
Details of time, place and cause of death of data subject		Car registration/licence plate number
Fingerprint		Email username and password
Medical history of family members		TV viewing habits
Bank account details		Sexual orientation
State benefit received		Religion
Family portrait (painting)		Credit card number
Vehicle ownership (make, model and colour)		History of addiction
		Education/qualifications
		E-commerce transactions
		Mother's maiden name
		Computer IP address
		Alias/pseudonym used in internet chat room
		Football team supported
		Date of birth
		Still image taken from CCTV
		Natural hair colour

Again the data listed within the *always* category tend be those that might more usually be used to *identify* an individual. There are however clear differences in approach between **Country 27** and **Country 9.** If the thesis that inconsistencies in classification are not the result of conceptual confusion, but rather conflict, then it should be possible to identify some distinction between the data types that they differ over.

Certainly **Country 27**, when asked to explain its approach to classifying specific data types chose to emphasise the *likelihood* of the information identifying an individual rather than the *possibility* of anonymisation. When asked to comment on its differing approach to the classification of 'Name' and 'alias/pseudonym used in a chatroom' **Country 27** explained,

> "Name of the individual is given to the person according [to] the law and is registered. Therefore we think that there are more possibilities to identify a person than by pseudonym." (Q2,A,Part1,q3(a))

Country 27 similarly justified classifying 'Parents' names' as *always* personal data but 'Mother's maiden name' as only *sometimes* personal data due to the fact that it does not consider it always possible to correctly identify an individual from their Mother's maiden name (see Q2, A, q3(b)).

One of the ways to distinguish between different types of information possessing the potential to identify an individual is to note that certain data types may possess the potential due to them being *uniquely related* to an individual, and always referable to an individual given sufficient resource, while others, while not necessarily unique *per se,* take the form of data deliberately attributed to an individual for the purposes of

identification (and may therefore, for practical purpose, often be *unique* in the context of a particular data controller but not more generally).

If we look at the data described by **Country 9** as *always* being personal data then we find *both* types of data (e.g. 'DNA profile' and 'Fingerprint' (unique identifiers), 'Name' and 'Email username and password' (attributed identifiers)). If we look at the data described by **Country 27** as *always* personal data then we find a relative paucity of 'attributed identifiers'. **Country 9** listed 'Home telephone number', 'Car registration', 'Email username and password', 'Credit card number', and 'Computer IP address' all within the *always* category and **Country 27** regarded *none* of these as *always* constituting personal data.

The property that **Country 9** appears to consider attributable to *both* types of identifier (and the property that justifies their classification as *always* personal data) is the impossibility of anonymising them in a way that would prevent *anybody* relating them back to a specific individual (in the case of an 'attributed identifier' it is *whoever* possesses a specific resource i.e. the database linking the individual to the identifier in question. In the case of a 'unique identifier' it really is *anybody* with sufficient resource, i.e. an appropriate database could always be constructed). A key element informing **Country 9's** concepts of personal data appears therefore to be the impossibility of (absolute) anonymisation of the data: the *possibility* of *somebody* identifying an individual *from* the information is sufficient to justify classifying the data as personal.

When asked *why* it regarded bank account details as *always* personal data but sexual orientation as only *sometimes* personal data **Country 9** explained,

> "I see a difference in that there is a possibility to anonymise information about a person's sexual orientation but not bank account details (because at least the bank can always relate the information to the data subject)." (Q2,A,q1)

The fact that **Country 27** chose not to select the same data types for description as *always* personal data suggests that it adopts a slightly different approach to classification and perhaps therefore also a slightly different concept of 'personal data'.

One way of trying to explain the differences in classification may be to suggest that **Country 27** only thinks information will be personal data if it is possible, in the occurrent context, to link the information back to the individual (i.e. it is *not uniquely related* to that individual in the circumstances). Whether 'attributed information' may *actually* be linked to an individual will depend upon whether access may be had *in fact* to the specific database (that could not be reconstructed independently regardless of resource) in the context. Access to this database by the individual possessing the information will determine whether it will, in the circumstances, constitute personal data.

This hypothesis, that the distinction drawn between **Country 27** and **Country 9** may revolve around the significance of context, may be supported by their responses to Q2, Part 2. **Country 27** identified certain data types as being 'more' or 'less' likely to be personal data in specific circumstances. For example, **Country 27** indicated that, in the context of a 'Police Force', information about criminal record is very likely (10 on a scale of 0 to 10 with 10 representing most likely) to be 'personal data'. In the context of a Sports club however it was regarded as very unlikely to constitute

'personal data' (0 on a scale of 0 to 10 with 10 representing most likely). It may be supposed that this is due to the average sports club lacking the wherewithal to identify an individual from details of their criminal record. **Country 9** in contrast declined to rank the data types as more or less likely to constitute personal data in the different context expressly stating,

> "I don't think the characterisation of a piece of information (as named here) as personal depends on the context …" (Q2, Part2)

It may therefore be tentatively suggested that **Country 27** are working with a concept of personal data that requires the data in question to enable the identification of an individual *in the circumstances* while **Country 9** are more likely to simply require recognition that the data *may* enable identification of the individual *in principle.*

It is readily acknowledged that this involves a considerable degree of conjecture. Ideally it would be possible to return to each of these countries, and the others that took part in the survey, and further test various hypotheses about the concepts of personal data that they work with. The data analysed so far does however clearly indicate that there *are* differences in classificatory approach. Theses are here constructed simply to help explore these differences and to suggest grounds upon which they *might* possibly begin to be explained. It is quite clear however that even if this suggested difference in approach between **Country 27** and **Country 9** were to be supported by the respective countries themselves, it would not necessarily explain all of the variation between them.

One example of the kind of anomaly that would remain is provided by 'State benefit received'. **Country 27** listed 'State benefit received' as *always* 'personal data' but 'Salary details' as only sometimes. Without further research it is impossible to know whether this is due to a perceived difference in the possibility of this data type enabling an individual's identification (which may itself be informed by contingent domestic circumstances) or due to information about 'State benefit' being perceived to bear some other kind of relationship to an individual.

It may also be worth observing at this point that **Country 27** also appear to be operating with a particular concept of what it means to be able to 'identify' an individual. When asked 'In what circumstances would Email address become personal data?' it replied 'If e-mail address contains name and surname of the person, this e-mail address becomes personal data' (Q2,A,q4). It is clear that Country 27 considers 'identification' to require more than simply being able to either *contact* the individual *via* email or *recognise* them by a chatroom alias (see above quote). One question that remains to be answered is whether the operative concept of 'identification' requires the possibility of a physical 'handshake' and/or contact with the 'real world' person rather than some imagined persona?

Country 20 also agreed that personal data must be capable of identifying an individual (Q2, Section C, q2). It suggested that it would amend the statement provided to read 'Information can only be personal data if it can *directly or indirectly* identify an individual' but the focus clearly remains upon the potential of the information to identify an individual: 'identificatory potential' again clearly appears a prerequisite of personal data. The list of data types that **Country 20** would consider *always* constitute 'personal data' again varies from those suggested by other

countries. In fact it suggested that fewer of those data types listed would *always* constitute 'personal data' than either of the previous two jurisdictions;

Always	Never	Sometimes
Name		Shoe size
Home telephone number		Blood group
National registration number		Countries visited in the last 5 years
Head and shoulders Photograph		Salary details
Car registration/licence plate number		Political party voted in the last election
Email username and password		TV viewing habits
Dental record		Sexual orientation
Parents names		Religion
Credit card number		History of addiction
DNA profile		Details of time, place and cause of death of data subject
Fingerprint		Education/qualifications
Computer IP address		E-commerce transactions
		Mother's maiden name
		Medical history of family members
		Bank account details
		State benefit received
		Alias/pseudonym used in internet chat room
		Football team supported
		Family portrait (painting)
		Vehicle ownership (make, model and colour)
		Date of birth
		Natural hair colour
		Still image taken from CCTV

Similar to **Country 27, Country 20** appears influenced by the perceived 'likelihood' of the information *actually* enabling identification in the occurrent context. When asked to explain why it would class 'Parents' names' as *always* personal data but 'Mother's maiden name' as only *sometimes* personal data **Country 20** responded,

> "Mother's maiden name is only one last name. It could possibly identify a person if the mother's maiden name is very rare, but the possibility of identifying a person on the background of two full names, both first name(s) and last name(s) (as I would understand parents' names) is much more realistic" (Q2,A,pt.1,q2)

Whether the information in question will enable identification, and therefore be termed 'personal data', appears to depend upon the context within which the information is located: 'How *realistic* is the prospect of identification?'

It may be worth explicitly noting at this point something about the apparent contradiction inherent within supporting *both* the idea that whether data will constitute personal data will *depend entirely upon context* and also that it is possible to draw up a list of data types that will *always* constitute 'personal data'. **Country 20, Country 17, Country 40, Country 9, Country 29,** and **Country 27** all agreed with the statement that,

> "It is impossible to create a list, since the concept is entirely dependent upon whether the context in which the information is placed."

They also *all* listed certain data types as *always* constituting personal data. While apparently contradictory, the occupation of both of these positions may perhaps be

explained by a realisation of the significance of context at a conceptual level meeting the need to provide more 'definitive' ('context independent') guidance in practice. **Country 36** said, in response to the statement,

> "Disagree – it is probably not possible to create a definitive and static list. It should be possible to list types of data that are more or less likely to be personal data and the factors/characteristics that affect this likelihood. Meaningful guidance of general scope ought not to be impossible"

When assessing whether the context is one such as to justify description of information as 'personal data' **Country 20** indicated that it was not the physical context but the informational context that it considered to be significant. As indicated earlier, respondents were asked to comment on the likelihood of particular data types constituting personal data in particular environments (e.g. Doctor's surgery, Small business, Police Force). When asked this question **Country 20** said,

> "The scenarios suggested have very little significance for the possibility of linking information to a certain person, in my opinion, unless I make a lot of presumptions, and invent more information than there is. The contextual situation of interest, in my opinion, would be which of these bits of information [i.e. Name, Internet Chatroom Alias, Criminal Record etc.] you found 'at the same time', not necessarily locus." (Q2,pt2)

Despite the differences that may be observed between the responses offered by **Country 9**, **Country 27,** and **Country 20,** they do seem to be clearly orientated around the idea that *'personal data' must identify an individual*. Differences between their approaches may be explained to some extent by differences over whether the clearest cases of 'personal data' are those that cannot be anonymised due to their unique nature or those that are most likely, in reality, to actually enable an individual's identification.

It may be significant that none of these countries considered any of the data types to *never* be capable of constituting 'personal data'. It would seem that *whether* data that *sometimes* constitutes 'personal data' actually constitutes personal data in any given case depends upon the presence or absence of sufficient *other* information to enable the individual's identification. This of course raises the possibility that it is in fact this *other* information that is actually identifying the individual in the circumstances at hand and the data in question is simply linked to the individual through association with this identifying data. One of the questions that arise is whether information that is associated with identifying data may be 'personal data' even if the information itself is not capable of identifying the individual (in any context). May 'personal data' relate to an individual in a relevant way without necessarily enabling their identification?

2. A relationship other than one of identificatory potential as prerequisite?

Some countries, while emphasising the significance of 'personal data' being related to an 'identifiable individual' provided responses that seem to question whether 'personal data' must *itself* be capable of identifying the individual or whether it would be sufficient for it to be related to an identifiable individual in some other way. Examples of this kind of response were provided most clearly by **Country 29, Country 17,** and **Country 6**, **Country 7.**

It was **Country 29** that perhaps most explicitly implied that it held this view when it disagreed with all of the statements made (see above) and responded,

> "The information can be regarded as personal data both in case where it can identify a particular individual and when it can affect an individual in a different way. In consequence the information can be regarded as personal data also in a case when it both identifies a particular individual and affects this individual in a different way." (Q2,C,q2)

The suggestion is clearly that information may constitute personal data if it *identifies* an individual but it may also constitute personal data if it *affects* the individual in some way other than identifying them.

The possibility that information may be considered 'personal data' even if it cannot itself contribute toward the identification of a data subject is however something that doesn't appear to be further supported by the responses received. Certainly **Country 29's** overall position appears somewhat modified if this response isn't taken in isolation and their responses to other questions are taken into account. Responses to other questions clearly suggest that *more* must be required than an affect *per se* before information will be considered to be the personal data of an individual (who may be identifiable through some other means).

In response to the second scenario within Questionnaire 2 **Country 29** stated that comments posted on a website about a band's music,

> "[C]annot be regarded as personal data. The information of band or group name is not sufficient to identify an individual." (Q2,B,Pt.1,qb)

This is despite the fact that it is indicated that within the context of the scenario the comments may be of economic value to the musicians ("as it increases the musician's chances of selling their music online to visitors of the website"). Comments made about the music of a group to which an individual belongs may clearly *affect* an individual in various different ways. That specific individuals potentially affected by such comments may also be identified as members of the band (at least by the other band members) by *other information* also appears reasonably clear. It seems to be the difficulty of identifying an individual *from the comments* made about a band that is pointed to by **Country 29** as explanation of why it would not consider these comments to constitute personal data. It is possible that the reluctance to characterise such comments as 'personal data' is due to doubt that any individual to whom the comments related *would/could* in fact be identified by other information. This suggested explanation for a reluctance to attribute information capable of affecting an individual (potentially identifiable through other means) with the status of 'personal data' is however undermined by other comments.

The position that *more* than simple affect is required seems supported by **Country 29's** response to another scenario based question. In the third scenario in the second questionnaire respondents were asked whether the records of a business might constitute personal data. Two situations are described, in the first situation one partner in a small business leaves the business taking all records with them and this affects the other partner's ability to conduct a business. In the second situation one partner taking the business records affects the other's social life (due to the

clients of the business having been personal friends). In neither situation did **Country 29** consider the information to be 'personal data'. In the first it explained,

> "It isn't information that can identify individual [*sic.*], but only refer to Brenda's business" (Q2,B,Pt.1,Scenario 4)

In the second it explained,

> "The fact that the information about clients can affect her social life is irrelevant for solving this problem" (Q2,B,Pt.1,Scenario 4)

To some extent this is may be considered an entirely sensible reaction to recognition of the consequences of holding that any information capable of *affecting* an individual identified *via* other means would be considered 'personal data'. If the capacity to *affect* an individual were alone sufficient to qualify information as 'personal data' (of an individual identified *independent* of the information in question) then a huge amount of information would be considered 'personal data'. As **Country 9** put it,

> "The rapid growth of world population may affect individuals sometimes in the future however this does not render this information to be personal" (Q2,C,q3)

The difficulty is reconciling this with the claim that information may be personal data if it may *affect* an (identifiable) individual (without itself contributing toward their identification).

Country 17 stated the position that to qualify as 'personal data',

> "the information does not in itself have to identify an individual, even data that can be linked to other information that identifies an individual may be personal data." (Q2,C,q2)

This notion of 'linkage' may represent a way of understanding the term 'referable' within Country 17's national legislation. Again, this term appears to capture more 'information' than that which is, itself, capable of identifying an individual,

> "If combined with other identifying information, almost any information could be personal data according to the Country 17 Law" (Q2,A,Pt.1,q1)

The question that remains however is how must information be 'combined' or 'linked' to identifying information for it to constitute 'personal data'? When we look at Scenario Four and the hypothetical case of the broken business partnership we find that **Country 17** also appear to resist regarding information as personal data if it is not 'about' the individual identified. When asked if the client records referred in the scenario could constitute the personal data of a partner (named 'Brenda') in the business **Country 17** replied,

> "No the information refers to her clients" (Q2,B,Pt1,q1)

They also resisted classifying the information as her personal data if the clients had become personal friends. While it appears entirely sensible to deny that information that does not 'refer' to the data subject can constitute their personal data, to do so adopts an undeniably narrow concept of the possible ways in which data may be 'linked' or 'combined' with other *identifying* information. Any information contained within the records of a company may presumably be 'linked' to other information capable of identifying the proprietors of the company. While it is again unsurprising to find a country unwilling to assume that *any* data that could be linked in *any* way to *any* information capable of identifying an individual is the 'personal data' of that

individual, it is not entirely clear how it is established whether a 'relevant' link has been achieved. This seems to raise the question of whether information may ever be 'referable to an individual' in a relevant way without that individual also being 'identifiable' from the information in question (either directly or in conjunction with other information).

When considering the question of what constitutes personal data some countries choose to emphasise the significance of context. Indeed, **Country 29** stated that the concept of personal data is entirely dependent on the context in which the information is placed, emphasising that

> '…the concept of personal data is of a relative nature' (Q2, C, q1).

Thus, they are of the view that it is impossible to create a list of what is or is not 'personal data'. **Country 17** concurred that the concept of personal data is entirely context dependent,

> "…the concept is dependent upon whether the data may be referable to an individual or not" (Q2,C,q1).

Emphasising the significance of context does not however, by itself, determine the nature of the relevant context i.e. it doesn't answer the question of *how* information must affect or 'be linked' to an individual to constitute their personal data. **Country 6 - Country 7**, adopted a different approach. They responded to the three statements detailed above by saying,

> "Information is personal data if the individual that is concerned may be identified"
>
> (Q2, C, 2)

In establishing whether the information concerns an individual in a relevant way however they explicitly rejected the over-riding importance of context. **Country 7**, agreed that it is not possible to construct a definitive 'list of personal data' but for them it is not the issue of context that prevents the construction of such a list. Rather,

> "It is the mass of information you can imagine to gather about one single person that makes it impossible to create such a list" (Q2, C, q1).

The shift of focus from whether information may 'identify' an individual to whether it 'concerns' an identified individual again raises the possibility that, while the individual must be identified, it may not be necessary for the information in question to identify them (in fact or theory) as long as it maintains some other kind of relationship with the data subject (i.e., in this case, it 'concerns' them).

Country 7's response is clearly informed by the formal definition of personal data operative within its jurisdiction;

> 'Personal data means any information concerning the personal or material circumstances of an identified or identifiable individual (the data subject).' (Section 3, Part 1, BDSG cited in response to Q2, B)

The broad scope of this particular definition appears confirmed by its classification of the data types contained within the first questionnaire,

Always	Never	Sometimes
Name		Details of time, place and cause of death of data subject
Home telephone number		Computer IP address
Shoe size		
National registration number		
Blood group		
Countries visited in the last 5 years		
Salary details		
Head and shoulders Photograph		
Political party voted in the last election		
Car registration/licence plate number		
Email username and password		
TV viewing habits		
Dental record		
Sexual orientation		
Religion		
Parents names		
Credit card number		
History of addiction		
DNA profile		
Education/qualifications		
E-commerce transactions		
Mother's maiden name		
Fingerprint		
Medical history of family members		
Bank account details		
State benefit received		
Alias/pseudonym used in internet chat room		
Football team supported		
Family portrait (painting)		
Vehicle ownership (make, model and colour)		
Date of birth		
Still image taken from CCTV		
Natural hair colour		

It might initially be supposed that this represents quite a different concept of 'personal data' to that considered in the previous section. If you remove the requirement that the data itself must be capable of identifying an individual, but insist only that it 'concern' an identified individual then you appear to potentially capture a lot more information: the notion of 'concern' appears more inclusive that the notion of 'identifies'.

A similar question arises however to those raised with reference to both **Country 29** and **Country 17**. How might this concept be operationalised? How does **Country 7** establish how far the term 'concern' is to extend in practice?

Once more it is very difficult to answer this question with any real confidence on the basis of two relatively brief questionnaires, but again there is some indication that at least one way of establishing whether information does concern the 'personal or material circumstances' of another may be to ask whether an individual might be

identified through the information. This is illustrated through an example provided by **Country 7** itself,

> "The material that has been used for statistical purposes can be personal data but the statistic itself normally is not. But there might be cases in which even a statistic contains personal data. If there is, for instance, a statistic on the results of [university exams then] in most cases you can not know from the statistic which person got which results. But if there was in the year of the statistic only one person examined in a very specific field you can read the exact result for this person from the statistic. In this case the statistic contains personal data. … As you can see from this case … the only doubt we sometimes have on personal data results from the question [of] whether an identification of a person is still possible or not. This depends on the concrete circumstances."

While **Country 7** clearly listed more types of data as *always* personal data than other countries, might this simply be due to recognition that these data types are *always capable* of identifying an individual (given appropriate circumstances)?

While the term 'concerns' appears to cover a broader class of information than 'identifies' in theory, is there some reason to consider whether information that does 'concern the personal or material circumstances' of an individual might not be conceptually defined through *its potential* to contribute toward their identification? If it is not possible to imagine a circumstance within which the data may function as an identifier, it may be difficult to sustain the claim that it is 'personal' data. Indeed, this notion would seem immanent within the description of data as concerning the 'personal or material circumstances' of an individual: may we not identify an individual through a description of their 'personal and material circumstances'? If the description does not allow us to know something about them then how may we distinguish it from any other description of 'circumstances'?

An alternative concept?

Rather than attempt to expand the concept of 'personal data' *beyond* that which might potentially identify an individual to include that which might 'relate to' them in some *other* way, an alternative may appear to be to require some kind of 'effect' *in addition to* 'identification': to affect them as well as identify them. It should be noted however that this particular approach will only in fact represent an *alternative* concept if a *particular kind of effect* is required.

An explanation of this observation may be found within the responses made by **Country 6 – Country 8**. This jurisdiction suggested that 'If the information contains no hint at the identity (the individual can not be identified and is not identifiable) the information will not be regarded as personal data' (Q2,C,q2) but also observe that 'I can not imagine information that identify an individual (or make them identifiable) [but] can not affect this individual' (Q2,C,q2).

While this comment would seem to capture the idea that *any* information capable of identifying an individual is *also* thereby capable of affecting them *in some way* it should be emphasised that this idea does not necessarily collapse the distinction

between this category and the next. Some countries appear to hold that, before information may be classed as personal data it must be capable of affecting them in some *specific* way. While **Country 8** noted that identification might have *some* effect, it did not limit its classification of personal data according to the nature of this effect and, accordingly, its emphasis appears to remain upon the potential of the data to *identify* an individual (see response to Q2, Section C, q1).

Assuming that an individual may be identified **Country 8** holds that '[t]here are no unimportant or irrelevant types of information. Every [piece of] information that says something about personal and objective circumstances in the life of an individual is to be protected as personal data.' (Response to Q2, Section A, q1). This response seems consistent with **Country 8's** tendency to classify all data types as *always* personal data in answer to Q1, q1.

Always	Never	Sometimes
Name		Details of time, place and cause of death of data subject
Home telephone number		Medical history of family members
Shoe size		
National registration number		
Blood group		
Countries visited in the last 5 years		
Salary details		
Head and shoulders Photograph		
Political party voted in the last election		
Car registration/licence plate number		
Email username and password		
TV viewing habits		
Dental record		
Sexual orientation		
Religion		
Parents names		
Credit card number		
History of addiction		
DNA profile		
Education/qualifications		
E-commerce transactions		
Mother's maiden name		
Fingerprint		
Bank account details		
Computer IP address		
State benefit received		
Alias/pseudonym used in internet chat room		
Football team supported		
Family portrait (painting)		
Vehicle ownership (make, model and colour)		
Date of birth		
Still image taken from CCTV		
Natural hair colour		

This is consistent with the responses of the other **Country 6** Authorities and may perhaps be explained in the same way: recognition that ALL of the data types listed are *capable*, in appropriate circumstances, of identifying an individual: personal data must always contain some 'hint at the identity' of the individual even if, in the circumstances, it is not operative within their identification.

Country 36 is another country that emphasised the significance of context and more explicitly emphasised the significance of identification. Interestingly, this same emphasis led **Country 36** to adopt a strategy inverse to that of the **Country 6** Authorities and classify almost all of the listed data types as *sometimes* personal data.

Always	Never	Sometimes
National registration number		Name
		Home telephone number
		Shoe size
		Blood group
		Countries visited in the last 5 years
		Salary details
		Head and shoulders Photograph
		Political party voted in the last election
		Car registration/licence plate number
		Email username and password
		TV viewing habits
		Dental record
		Sexual orientation
		Religion
		Parents names
		Credit card number
		History of addiction
		DNA profile
		Details of time, place and cause of death of data subject
		Education/qualifications
		E-commerce transactions
		Mother's maiden name
		Fingerprint
		Medical history of family members
		Bank account details
		Computer IP address
		State benefit received
		Alias/pseudonym used in internet chat room
		Football team supported
		Family portrait (painting)
		Vehicle ownership (make, model and colour)
		Date of birth
		Still image taken from CCTV
		Natural hair colour

While the emphasis appears to remain upon the *possibility* of identification from the information in question, it is the circumstances at hand that are considered of prime importance when attempting to answer that question,

> "I come back to the issue of context – perhaps most, if not all, of these data sets could be personal data in the right circumstances (i.e. In the right combination of other data and/or collected by the kind of entity/person that has the facility/means to use it to identify the individual). Perhaps it is more a matter of degree of likelihood of such data being identifiable."

3. Identification and affect as prerequisites

Some countries indicated that they understood the term 'personal data' to only apply to information that is capable of both identifying an individual *and also* of affecting them in some additional way.

Country 35 agreed that it is not possible to create a definitive list, but suggested that the creation of some lists should be possible. It stated that information could only be

classed as personal data if it is capable of affecting an individual's privacy *as well as* identifying them; for Country 35 then, 'affect', requires something that affects *privacy*.

Despite this general position, when asked to classify pieces of data (Q1, 1), **Country 35** classed 'Medical history of family members' as *never* capable of being personal data, and 'Parent's names' as *sometimes* capable of being personal data. It is difficult to explain how information about parent's names is any more likely to identify *and affect* an individual than information about their parent's medical history.

Overall, **Country 35** listed 19 data types as *always* constituting 'personal data'. Information that is more likely to affect an individual if revealed appears to have been selected regardless of whether it is, of itself, likely to *identify* that individual (e.g. sexual orientation, religion). In Scenario 1 (Q2, B) **Country 35** stated that data that can identify an individual but "cannot… affect the individual's privacy" is not to be classed as personal data. This is consistent with its other statements that suggest that to 'affect' requires the individual's *privacy* to be affected.

Unlike other data protection agencies, **Country 35** suggested that 'Parent's names' are not personal data that relates to and identifies the subject unless specifically being used by the data subject for that purpose (e.g. as a security question for a bank) (Q2, A, q1). **Country 35** also qualified categorisation of CCTV footage in a similar fashion: "A CCTV image would only become personal data if it is being used to identify the data subject." Clearly, considerable account is being taken of the use to which the information in question is being put. This reaffirms **Country 35's** assertion that it is not possible to create a definitive list of personal data. The classification of data as personal is contingent upon circumstance and the context of use.

Country 33 also claimed that for data to be classed as personal it must both identify and affect an individual, "especially after the Durant judgment" (Q2, C, q2). Again, it also holds that 'affect' means to affect privacy. However, the idea that the information must be capable of identifying an individual seemed inconsistent with **Country 33's** responses to the first questionnaire. In Q1,q1 they classified 'Sexual orientation' and 'Religion' as *always* personal data but 'Credit card number' and 'DNA profile' as only *sometimes*.

When quizzed on these responses (Q2, A, q4) it moved 'sexual orientation' and 'religion' to *sometimes*, as "more data will be needed to identify an individual". However, it maintained that they are still "very confidential and important pieces of information". This re-classification ensures that **Country 33's** position remains broadly internally consistent. **Country 33** is clearly keen to class as personal that data which 'affects' an individual, but as its re-classification demonstrates, it also stresses the need for identification too. This is confirmed by its response to Scenario 4 (Q2, B, q4), where data that affected the individual's social life was not considered to be personal data as "the data do not identify [the individual] in any way". However, it is still difficult to explain why information on 'Countries visited in the last 5 years' was held to be *never* personal data. One would normally envisage that such information has the *potential* to both identify and affect the individual's privacy.

Country 34 agreed that it is not possible to draw up a list of personal data because the concept is entirely dependent upon the context in which the information is placed (Q2, C, q1). However, it listed 9 different types of data as *always* personal data (Q1,

q1). Of course, this may be because **Country 34** *assumed* a particular context. If **Country 34** did indeed assume a specific stable context, then the fact that it classified data types differently must be explained by the idea that there remains a property that distinguishes those data types identified as *always* personal data from those identified as only *sometimes* personal data.

When searching for such a distinguishing property that **Country 34** might consider significant, it is relevant to note that it agreed with the proposition that information "can only be personal if it both affects and identifies an individual" (Q2, C, q2). However, there is no obvious reason for distinguishing between those things Country 34 would register as *always* or *sometimes* personal data on the grounds of either identificatory potential (e.g. 'Car registration' was held to be *always* personal data, while 'DNA Profile' only *sometimes*) or effect (e.g. E-commerce transactions were held to be *always* personal data, while 'sexual orientation' only *sometimes*). Even with those countries then, such as **Country 34**, that appear to possess a clear concept of personal data, it is not always easy to see how that concept is being articulated in practice through their decisions.

The trio of **Country 35**, **Country 33** and **Country 34**, who all agreed with the statement 'information can only be personal data if it both identifies and affects an individual', are all following the law as it was laid down in the case of *Durant*. This helps to explain their position. In *Durant*, the individual had already been identified, but this alone was not enough to render the data sought 'personal'. To be personal, the data also had to be appropriately located upon a 'continuum of relevance and proximity to the data subject' (Auld LJ at para.28). This suggests that something more than simple 'identification' is required. The Court also felt that personal data "is information that affects [a person's] privacy" (Auld LJ at para.28).

All three jurisdictions believe that something more than identification is required for a piece of data to be classed as personal. The data must also 'affect' the individual in some way. This is consistent with the judgment in *Durant*. **Country 35** and **Country 33** in particular appear to have taken privacy to be the key component of 'affect'.

B5: What do we learn from the results of the survey?

Having presented the findings from all three elements of this study (the literature review, the survey of formal definitions/domestic legislation and the exploration of the practical applications of the concept 'personal data'), we can see that results seem to converge. We have demonstrated that there is a lack of clarity with regard to the concept 'personal data', both within and outside of the EU. Interestingly, these conceptual confusions have failed to raise concern and continue to be alluded to as unproblematic at the level of operationalisation.

Whilst it may not be either surprising or worrying to find inconsistencies in approach *between* different countries, it has become clear that not all of the divergences can be explained simply by variations in domestic data protection policies. We have, indeed, seen evidence of inconsistencies *between countries that adopt similar policies*, and we have even seen evidence of inconsistent applications *within individual countries*. Undoubtedly, what constitutes 'personal data' is shaped by formal definitions, but the amorphous nature of those formal definitions allow for

alternative interpretations in practice. Thus, how the formal definitions are interpreted will depend upon an underlying understanding of the concept of personal data.

What this study has shown is that this 'underlying understanding' varies considerably across countries. We have learned that that these varying interpretations are often difficult to extract. Indeed, the analysis of our questionnaires show that not only is it not always immediately apparent what concept of personal data that a country is working with, but a particular concept is not always unambiguously pointed to by their responses to specific questions. We recognise, of course, that further research would help us to unravel these ideas. However, we have been able to identify three underlying concepts which are frequently being applied to the question of 'what are personal data?':

1. The capacity of the data to identify an individual
2. The capacity of the data to affect an individual
3. The capacity of the data to identify AND affect an individual

Our analysis demonstrates that these concepts are both under-developed and applied interchangeably. The question we need to ask is 'does it matter?' Why is the lack of clarity a problem?

This question is best addressed by reminding ourselves of the original aims of the Directive. The Directive was intended to create a harmonious European wide system of data protection whilst simultaneously supporting the single market. Our results suggest that the vision of a harmonious European system of data protection is seriously threatened by inconsistent applications of the concept of 'personal data'. As long as these inconsistent approaches continue, the uncertainty faced by both data subjects and data controllers will increasingly present us with significant challenges. Perhaps more worryingly, if data controllers are uncertain as to what kinds of data may be classed as 'personal data', we must consider the possibility of arbitrary decision making. Taken beyond the context of the EU, these problems are amplified on a global scale by the worldwide lack of consistency. Global trading relationships could be damaged by the divergences in approach to personal data.

We are forced to conclude that there is only one solution to this growing problem: to develop an express, robust, theoretical framework within which we can begin to develop a clear understanding of 'personal data'. Until such a framework is developed, any attempt to re-define the individual key terms of the Directive is likely to be ineffective. In the next section (Part C), we attempt to inform the construction of such a robust theoretical framework through a critique of the existing approaches.

Part C – "Ideal Types" and "Decision-making Models": Developing a theoretical framework to inform an understanding of the term 'personal data'

C1: Introduction: Can alternate models be distilled from the practice of Data Protection Authorities?

Parts A and B were primarily concerned with three levels of analysis. Some *interim* conclusions from this analysis may now be restated in terms of their relevance to the development a robust theoretical framework capable of justifying decisions about the classification and definition of personal data.

- Firstly, the literature concerned with the meaning of 'personal data' (and other key terms from the Directive definition), within the perspectives of law, psychology and sociology was reviewed. It is clear that within this literature the key terms (identified in A3) have contested meanings.

  A review of relevant literature did not therefore, by itself, provide a clear model or *concept* of 'personal data'. It is not possible to draw straightforwardly upon the literature to develop a robust theoretical framework capable of justifying classificatory decisions about personal data in practice.

  The controversies within the literature did however point toward some of the difficulties widely recognised to be associated with the development of any concept of 'personal data' and which we might expect to have to be tackled by countries regulating personal data. These included the difficulties associated with privacy and identity being concepts that are given definition by context. Any understanding of personal data that is reliant upon them must therefore be responsive to changing circumstances; a potential challenge if attempting to regulate in a consistent and predictable fashion.

- Survey and examination of the formal definitions in the national legislation of participating countries yielded an overall impression of consistent use of terminology, and a degree of apparent similarity in legislative approach. This was particularly true if comparison was made within the three jurisdiction groups (see B2).

  There were very few reported problems associated with understanding any of the key terms identified. It appeared that the difficulties anticipated by the literature in defining the conceptual limits of 'personal data' were not widely acknowledged by countries as encountered within their everyday interpretation and application of the terms.

- Further analysis involving questionnaire responses from the participating countries on their approach to data protection indicated some considerable inconsistencies in data classification. Lack of agreement existed both between, and within, the three groups identified. The inconsistencies

appeared to be consistent with different classificatory *strategies* being employed by the countries.

Countries strike a particular attitude toward what information they will classify as personal data. Their classificatory attitude is clearly shaped by the 'formal definitions' operative within their own jurisdiction and some differences in classificatory strategy may be partially explained by variation between domestic data protection legislation. Differences continue to exist however between the strategies adopted by countries with relatively similar formal definitions. Each of these formal definitions has a sufficiently 'open texture' to permit various alternative interpretations. The adoption of a particular interpretation may be guided by the operation within these jurisdictions of an underlying concept of 'personal data'. Through clustering the responses of the countries we can develop certain themes which may be said to be indicative of such underlying concepts (see B4).

Despite indications that there may be different concepts being employed by different countries,[2] it was not absolutely clear what the true conceptual differences were or why they occurred. Certain consistent themes could however be identified and responses clustered. In this way, overlap and consistency could be emphasised and differences in conceptual understanding hypothesised.

The bases of the perceived differences indicated by the empirical data could however only be summarised by hypotheses. The absolute concepts which were in fact informing the responses given by each data protection authority, even following the second questionnaire, were not immediately obvious. Developing apparent themes does allow the construction of 'ideal types' which may illustrate alternative approaches toward understanding the term 'personal data'. These 'ideal types' may then be used to explore alternative 'theoretical frameworks' capable of justifying decisions about the classification and definition of the 'key terms'.

It may be observed at this point that the development of any 'robust theoretical framework' capable of justifying decisions about the boundaries and definition of 'personal data' cannot take place within a vacuum. (see Box 1). By selecting these ideal-types as an appropriate point of departure we ensure that comments remain grounded within the realities of practice.

While each of the 'ideal types' may represent a caricature of the conceptual understanding of 'personal data' actually adopted by any country they may help to elucidate and illustrate the elements that do contribute toward the composition of operative concepts. They also help demonstrate the difficulties that may accompany building a decision making strategy upon a conceptual understanding of the term 'personal data' incorporating these different elements in different ways.

By using the artificial simplicity of these 'ideal types' to guide the construction of model classificatory strategies we may thus explore the benefits and dis-benefits of taking differing positions on the meaning of personal data and perhaps also gain further understanding of the actions of countries in practice.

[2] See section A4

As different countries appear to be employing slightly different concepts of data protection we might expect their vulnerability to the conceptual difficulties articulated to vary in similar fashion.

Box 1: The Development of a Theoretical Framework:

The development of a theoretical framework not only presupposes a 'pre-theoretical' point of departure, but also, a particular method of progressing from that point. The integrity of any theoretical framework may then be seen to depend upon at least three different things: the selection of an appropriate starting point, the selection of an appropriate method of progressing from that point, and the error free application of that method in the construction of a theoretical framework.[3]

If one seeks to develop a robust theoretical framework capable of informing an understanding of the term 'personal data' then one might, for example, choose the plain and ordinary meaning of the terms 'data' and 'personal' as a point of departure. One might then proceed to see what, through a straightforward application of logical principles, followed from a qualification of the one term by the other. This approach would however only adopt an appropriate 'pre-theoretical' starting point if the terms that were being explored were themselves defined by common Country 42ge.

Similarly, one might select technical definitions of certain terms as proposed by particular literatures, and attempt to construct a conceptual framework from these definitions. This would however only be an appropriate 'pre-theoretical' starting point if one had good reason to isolate particular literatures and privilege them over the alternatives.

If one is seeking to inform an understanding of the terms used by countries it would be inappropriate to privilege either common Country 42ge or a specific disciplinary perspective as a pre-theoretical starting point if neither is accepted by the countries themselves as providing a relevant standard. It would not be inappropriate due to the starting point being contingent however, but rather due to its relationship with the definition of 'personal data' adopted by the countries being presumed.

If the aim is to inform an understanding of the term 'personal data' as used in the Data Protection Act 1998, the Data Protection Directive 95/46/EC and other domestic data protection legislation, then the technical definition of the term *as provided by these instruments* may more appropriately provide a suitable point of departure. A conceptual understanding of 'personal data' capable of grounding a theoretical framework may be inferred from the understanding, the interpretation and the application of the definitions by countries within their practice.

[3] If a person seeks to build a sound house they not only need to have suitable foundations, and appropriate plans, they need also to build the house according to those plans.

C2: Mapping a series of 'ideal types' inspired by practice

In this section, a number of 'ideal types' are developed to explore differences in approach toward data classification. These 'ideal types' may prove useful for exploring the strengths, and weaknesses, associated with operating particular concepts of personal data. More sophisticated decision making strategies, which may more closely represent those actually adopted in practice, can be developed by combining elements associated with these different 'ideal types'.

The 'ideal types' that we have developed attempt to isolate those differences in approach that were appreciable within the responses received from countries; they do not draw distinctions that were not evident within those responses. For example, a distinction is not drawn between 'data' and 'information'. Nor are the issues of 'direct' or 'indirect' identification, or 'anonymisation' directly addressed. These issues would however have to be addressed were these concepts to be operated in practice. The significance of these ideal types to these issues is however addressed later (in C4).

The first thing that may be noted is that the countries appear to be operating with two general conceptualisations of personal data:

1. The 'context *independent'* concept. Countries that take this approach suggest, either explicitly or implicitly, that a list can be drawn up of those data types that are *always* (and/or *never*) personal data. Context is not seen as a crucial factor in determining whether data should be classified as 'personal'. [4]

2. The 'context *dependent'* concept. Countries that take this approach class (almost) all pieces of data as 'sometimes' capable of being personal data. *All* data *could* be personal data "in the right circumstances".[5] Accordingly, these countries hold that it is *not* possible to draw up a *definitive* list of data that will *always* (or *never*) constitute 'personal data'. It may be possible to draw up a list of data that will (almost) always or (almost) never be personal data due to the relevant context (almost) always or never being present.

It should be said that before attempting to align individual countries with either one of these two positions a particular note of caution must be sounded. Some countries made a point when answering the first question in the first questionnaire (i.e. the 'data classification question') that has some significance here. **Country 37** may serve as an example of how it may affect how they should be understood within the typology described.

In answering Questionnaire 1, Question 1, **Country 37** assumes that the data relates to an individual that *is already identified*, and classifies the pieces of data as personal or otherwise on that basis. This helps to explain why the vast majority of

[4] It needs to be emphasised that in developing this ideal type we are drawing upon those countries whose operative 'concept' of personal data appears consistent with this feature. Many countries indicated that they would consider certain data types to be 'always' personal data in practice. That does not mean that they would favour a 'context independent' conceptualisation of 'personal data', it may simply indicate that they consider that, for those data types, the relevant context would be sufficiently prevalent in practice to justify adopting a default classification of 'personal data'.

[5] See Country 36, Questionnaire 2, Section A.

the data is considered by **Country 37** to be personal data. It is possible that the other countries that are close to the **Canadian** position (especially **Country 3**, who declared that *all* the pieces of data are *always* personal data) made a similar assumption.

What these countries appear to be doing is assuming a context - a context within which the individual has already been identified - and basing their classificatory decisions upon this assumption. If this is the correct interpretation of their approach then, the position of **Country 37** *et al* is actually much closer to the **Country 36** and **Country 29** 'context dependent' conceptualisation than appears from their responses to the 'data classification question'. [6]

An important point however is that this does not undermine the dual-conceptualisation approach suggested. Some countries do appear to favour a 'context dependent' test, whilst some do not accept, or do not appear concerned with, the effects of context.

Towards the "ideal types":

1. Context Independent

Within the context independent conceptualisation, two further variations may be identified and developed:

- The Context Independent 'Identification' model: The 'Unique Identifier'
- The Context Independent 'Affects' model

Operating with the 'unique identifier' concept, countries class as personal data that data which they consider to be a unique identifier of the individual. Two of the most obvious examples of this are 'Dental Records' and 'National Insurance number'. **Country 27** and **Country 20** appear perhaps to be the best exemplars of this approach in practice. Although they do not consider exactly the same pieces of data to be personal data, and they include items such as 'name' that are not (usually) unique, there does appear to be a tendency towards the 'unique identifier' model. **Country 9** and **Country 1** also provide responses that are consistent with the adoption of this 'unique identifier' model.

Under the 'affects' model, countries class as personal data that data which they consider capable of 'affecting' an individual. 'Affect' can potentially cover a vast spectrum of things. The most likely thing to be considered significant (for the purposes of measuring effect) under this model appears to be 'privacy'. While not as commonly held a position as the unique identifier concept (perhaps a consequence of the potentially enormous scope of this model), it is nevertheless a position that it is possible to associate with, *inter alia,* **Country 7**.[7]

[6] It remains unclear however whether the context assumed is one in which the data itself has identified (or is capable of identifying) the individual or whether it is thought that the information may be linked to an individual identified in some other way.

[7] That they may align themselves more comfortably with this position than many authorities may be (at least partly) explained by their formal definition of personal data, which states that "personal data means any information *concerning* the person…" (section 3, part 1 BSDG (emphasis added,), cited in response to Q2, B, and see also A4).

2. Context Dependent

Similarly, within the 'context dependent' conceptualisation, two more variations may be developed.

- The Context Dependent 'Identification' model
- The Context Dependent 'Affect' model

Under the first model, *all* data is *sometimes* capable of being personal data, as any data is capable of *identifying* an individual's privacy *in the right circumstances*. **Country 36** and **Country 29** in particular provided responses that might be at least partially explained through the operation of this concept.

Under the second model, all data is *sometimes* capable of being personal data as *any* data is capable of *affecting* an individual *in the right circumstances*. **Country 17's** responses come closest to consistency with this position.

These possible variations, then, allow us to develop four different 'ideal types' to illustrate the advantages and disadvantages that may be associated with operating with concepts of 'personal data' that draw upon the notions of identification or effect in either a context dependent, or a context independent, manner:

1. 'Context independent Unique Identifier' Ideal Type
2. 'Context independent Affects' Ideal Type
3. 'Context dependent Identifier' Ideal Type
4. 'Context dependent Affect' Ideal Type

C2: Summary - Four Ideal-Types

'Unique Identifier' Model
Personal Data is data which may be uniquely related to an individual. Due to the uniqueness of the data, it is impossible for it to be anonymised in such a way as to render it impossible for it to continue to be related to an identifiable person. Context is irrelevant.

'Affects' Model
Personal Data is data which is capable of affecting an individual in a relevant way. It is possible to anticipate whether data will affect an individual in a relevant way without taking account of context.

'Context Dependent Identifier' Model
Personal Data is data which may identify an individual. All data is capable of being personal data, as any data is capable of identifying an individual in the right circumstances.

'Context Dependent Affects' Model
Personal Data is data which may affect an individual in a relevant way. All data is capable of being personal data, as any data is capable of affecting an individual in a relevant way in the right circumstances.

C3: Discussion of the Ideal-Types

A closer look at each of these ideal types reveals some of the disadvantages, and advantages, that may be associated with each.

IDEAL TYPE 1: The 'Unique Identifier' model

The impossibility/improbability of anonymisation

Data may be described as 'personal' due to the perception that the data may be linked directly to an individual without reference to any other information: the data represents a unique identifier. In the paradigm case, while the process of linkage may require the generation of additional information, all the steps necessary to link the data to an individual may be completed entirely *ab initio* without reference to any pre-existing database of knowledge. It is, to that extent, data capable of yielding identifying information in a context independent fashion.

The first thing to note about this particular model is how few types of data may actually fit the paradigm of a 'unique identifier'. One rare example might however be provided by a DNA sample. This example may be explored in order to test the integrity of the 'unique identifier' 'ideal type'.

An individual's complete DNA sequence is unique to them.[8] A DNA 'profile' that is *practically* unique to an individual may be constructed through the analysis of a DNA sample.[9] If a sample of DNA is obtained, (assuming that the sample is of sufficient quality to construct a profile,) it will always be *theoretically* possible to match that sample with the individual from whom it originated (assuming also other sources of DNA continue to be associated with that individual). It is not *possible* for a DNA sample to be anonymised in such a way as to make it impossible to match the sample to a living individual without effectively destroying it.

One of the reasons that the possibility of a match endures, and an appropriate 'database' enabling 'linkage' may be developed *ab initio*, is that *other* genetic material (i.e. 'a comparison sample') from which a corresponding profile may be derived will (usually) continue to be available. If an individual is still alive, or only recently deceased, it will not usually be a problem to obtain a sample for the purposes of comparison: a complete DNA sequence resides in almost every cell of a human body. No other type of data is capable of independently providing such a

[8] Unless they have an identical (monzygotic) twin
[9] The National DNA database, used by UK Police Forces, uses a system of profiling known as SGM plus that constructs a profile with a chance match probability of less than a 1 in 1 billion (thoCountry 42nd million) (See Linacre 'The UK National DNA Database' *The Lancet* (2003) 361: 9372)

uniquely identifying 'profile' susceptible of matching to such an unchanging (and unchangeable) feature of an individual. It is these highly unusual characteristics that perhaps entitle a sample of DNA to description as the paradigmatic example of a 'unique identifier'.

Even with the case of DNA however, to present it as independently capable of the 'direct' identification of an individual without reference to any additional (extrinsic) information, is somewhat misleading. Even if the wherewithal were to exist to generate an appropriate interpretative framework *ab initio* (which it must be said is extremely unlikely given the immense resource implications) it would still be necessary for such a framework to exist before identification could take place i.e. while the potential for identification may reside within the DNA sample, this potential may only be realised if *additional* information helps provide a context that gives the sample meaning.

Minimally this 'interpretive framework' would include information capable of reliably and consistently identifying both the 'target sample' and the 'comparison sample' across a period of time, and, perhaps most importantly, of linking the 'comparison sample' to an individual. The formation of this 'interpretive framework' must necessarily draw upon data sources extrinsic to the sample itself if the 'comparison sample' is to be matched to an individual.[10]

Even DNA, perhaps the 'paradigm case' of a unique identifier, is not then capable of identifying an individual in a *wholly* context independent way. What may be considered significant however is that the relevant context may be *substantially* provided by the data itself. Notwithstanding the difficulties in achieving identification absent a 'pre-existing' database of information, it is technically possible for *all* of the data substantially needed to enable identification (beyond those formally associated with the *procedure* of identification *to* an individual) to be mined directly from the DNA sample itself.

While other data types may not quite so perfectly fit this ideal they might nonetheless be described according to their 'degree of fit'. One particular model of 'personal data' may then be constructed around how closely data types do fit this particular 'ideal type' of 'personal data': data which may be characterised as 'personal' according to the relative difficulty (or impossibility) of effective anonymisation.

There are certain types of data that might be immediately recognised as more likely to resist 'anonymisation' and, accordingly, more likely to be regarded as 'personal data' according to this conceptual understanding of the term. 'Profiles' generated from an individual's fingerprints, iris patterns and other biometric features may be idiosyncratic to an individual in comparable ways to their DNA profile.[11]

---

[10] A counter example may be offered to this if the individuals suspected of matching the sample were physically present. If the comparison sample is not to be drawn directly from a present physical person then some additional database will most obviously be required i.e. one that links the comparison sample with additional identifiers. It could be argued however that, even if the individuals were present, some framework of understanding would have to be developed sufficient to continue to associate the sample drawn from the individual for comparison that incorporated information not present within the sample itself (e.g. visual recognition and record of the physical features of the 'suspect').

[11] See Daugman JG 'How Iris Recognition Works' IEEE Transactions on Circuits and Systems for Video Technology, (2004) 14(1): 21 -30

Novel databases might be generated *ab initio* for these biological features (as with DNA), due to their relative permanence and their ability to consistently provide 'comparison samples' associated with specific individuals. While they rely upon features for the generation of these 'comparison samples' that are not quite as unchanging (or unchangeable) as an individual's DNA, the features may be considered sufficiently 'permanent' to render it unlikely that a 'pre-existing' database would be necessary. The identification of an individual from information of this kind would however still require an appropriate 'interpretive framework' and the generation of a suitable 'database' if one were not to already exist. While recognising the necessity of an 'interpretive framework' does undermine to some extent the notion of 'direct' identification through a 'unique identifier', this notion is, we believe, almost entirely eroded if a 'pre-existing' database is *necessarily* relied upon for identificatory purposes.

In practice of course, as already indicated, databases will not usually be constructed *ab initio*. The matching of DNA, fingerprint or iris pattern profiles will invariably rely upon matching pre-existing profiles contained within databases that contain their own 'internal' identifiers linking the 'comparison sample' with an individual. As soon as a 'pre-existing' database is not however a convenience but rather a *pre-requisite* for identification, then not only does the anonymisation of data become more feasible, but it also becomes much more difficult to sustain the claim that data of a particular type will necessarily be sufficiently unique to enable the ('direct') identification of an individual in a context independent manner.

An identifier may be unique within the context of a specific database (e.g. credit card number) but that identifier may bear no relation to an individual other than the relation it is given *by* that specific database. The relationship, once established, may be confirmed in a number of other transactions (here again a credit card number may provide a good example) but if that original database (and all subsequently generated ones) is lost then it may not be possible for it (them) to be reconstructed and the relationship between an individual and the data established.

The identifier is *attributed* to an individual and the record of this attribution provides the relevant database. It will be access to a relevant database that determines whether it is possible in fact to link the identifier to the individual. The 'more' unique an identifier (i.e. the more contexts within which it retains its status as unique) then the more difficult it may be in practice to disguise its potential from one who recognises its character, but, ultimately, the reliance upon an artificially generated database (i.e. a contingent attribution) *must* leave that assumption of significance vulnerable.

To give an example; the personal identification number (PIN) (used by individuals in conjunction with a bank card to withdraw money from an automated teller machine) is far from unique. It characteristically contains only 4 numbers. With approximately 146 million plastic cards in The UK capable of withdrawing money when associated with a PIN[12], even if PINs were evenly distributed across all of the cards in use then any given PIN could be successfully associated with over 14000 cards.[13]

Due to the prevalence of any given PIN, access to a PIN *alone* would not enable identification of an individual. This is true even if access were had to *all* of those

[12] Source: www.apacs.org.uk
[13] Success here being measured by the ability to withdraw money from an ATM.

databases within which the relationships between PINs and individuals were recorded. The problem would not however only be a result of the number of individuals that may be associated with any given PIN. If a PIN were found alone then one could not even be sure that it was a bank PIN. Four numbers could be used in a variety of contexts to identify an individual.

In contrast, a British driving licence number is relatively distinct in form. The first five characters are taken from the driver's surname (if the surname is less than 5 characters long the number 9 will be added to the gaps), the following 6 characters are all numeric with the first and last representing the year of birth, the second and third the month and the fourth and fifth the date. (If the licence holder if female then '5' is added to the second digit – it will always therefore be either 0 or 1 (if you are male) or 5 or 6 (if you are female)). The next two characters are taken from an individual's forename(s) (if they only have one then the number 9 is added) and the last three characters are computer check digits.

Given the relatively distinctive character of the identifier, if anyone with access to an appropriate database were to find a Driving licence Number (in isolation from any other data) they would (probably) be able to identify an individual. The relatively 'unique' nature of the identifier makes an individual's identification more likely; not only because the identifier is unique *within* a specific context (in this case a database), but also because the identification *of* the relevant context (capable of facilitating identification) is easier than with other identifiers e.g. PINs (i.e. it is more unique across databases/contexts).

The *more* unique an identifier, the more difficult it will be to effectively anonymise it. For so long as access may be had to the relevant database, the only way of effectively anonymising a truly unique identifier may be to destroy it. It is also true however that the 'uniqueness' *of an attributed identifier* is an entirely contingent quality: the uniqueness of an attributed identifier (isolated from context) is vulnerable to the proliferation of identifiers of similar natures. In other words, one way of destroying a 'unique identifier' may be to remove its quality as unique.

If plastic card companies decided to adopt a system of 'numbering cards' that incorporated five characters, six numbers, two characters and then three numbers, the number of databases within an alphanumeric sequence of this nature had relevance could multiply overnight. This could result in the same sequence of numbers and characters as represents an individual's driving licence number being linked to many *different* individuals. An individual's driving licence number would not be any less unique within the context of the DVLA database but the number of identifiers possessing this structure and content would multiply at the same time as would the number of databases within which they were recognised as possessing meaning and the number of individuals linked to them. Selecting *the* context within which an 'identifier' may identify *an* individual is reliant upon *additional* information.

Those things that operate as 'unique identifiers', in practice, are unlikely to consist of single pieces of data. They are rather more likely to constitute 'portfolios' of a number of pieces of data. These pieces of data may take the form of identifiers (which may, independently, be more or less unique in nature). It may be the more unique *a* specific identifier is then the fewer the number of additional identifiers it must be associated with, but, invariably, it will have to be joined with *some* additional information.

The *possibility* of a specific piece of data enabling identification will be linked to the existence of a context within which it may constitute a unique identifier. The *probability* of identification will be linked to both the availability, and the awareness, of that context in the circumstances.

In summary the Unique Identifier model:

Has the advantages that it:

- Allows individuals to reliably categorise information as 'personal data' in a (relatively) context independent fashion (i.e. you can draw up a list)
- The likely presence of the context that is acknowledged significant (i.e. the interpretative framework capable of recognising and using the identifier in question) may be estimated with a degree of accuracy and reliability that may not be associated with the other decision making models.

Has the disadvantages that it:

- Overestimates the extent to which any given piece of data may be unique independent of context and therefore,
- Underestimates the significance of context. Most data functioning as an identifier, and certainly *all attributed identifiers,* will only be unique *in a given context*.
- Even within a context in which an identifier possesses the status of unique, it may only enable the unequivocal identification of an individual when associated with other information.

IDEAL TYPE 2: The 'Context Dependent Identifier' model

The possibility/probability of identification

According to this ideal type, 'data' is judged to be 'personal' if it may be used to identify an individual. *Any* information that can be used, in appropriate circumstances, to identify an individual may be considered personal data.

Perhaps the prime exemplars of this kind of data are 'name' and 'address'. The fact that these data types are so often referred to in association emphasises not only their combined efficacy at identifying individuals but also their independent reliance upon a broader informational context to realise their identificatory potential (in this case the context is substantially provided reciprocally by the two data types).

It is clear from the examples chosen that data recognised as 'personal' by this concept is unlikely to be 'unique' (in isolation). Indeed, those attributes that might most commonly be used to identify individuals are often widely shared throughout a population (e.g. first name, height, gender). They can only effectively operate as identifiers however if, within the context, they point to a single individual.

As indicated above, common identifiers may be associated to generate a unique 'portfolio' of identifiers. Such a 'portfolio' may actually be used more often for identificatory purposes than a single (relatively) 'unique' identifier. Even the most 'common identifiers' might then be used as the 'key' to identification within an appropriate context (see example of PIN above).

According to this concept then, it is not the uniqueness of the data *per se,* that is significant but the availability of a context (possibly informed by other identifiers) *within which* that data may function as a unique identifier. The availability of an interpretative framework capable of enabling an individual's identification from a relatively common identifier is clearly dependent upon access to this broader informational context: description of data as (potentially) 'personal' according to this concept relies upon the '*possibility'* of accessing such additional information/relevant context.

One of the issues that must be addressed by this 'ideal type' is the significance, if any, that is to be attached to the '*probability'* of access being gained: how significant is the likelihood of the relevant context arising? Is mere possibility enough, no matter how distant, or must identification *actually,* in the instant case, be a *realistic* possibility?

It was established earlier that it is almost always (theoretically) possible for a DNA profile to be placed within a context that enables the identification of an individual. While a DNA profile may be far more 'unique' than a telephone number, for so long as telephone directories are more readily available than DNA databases (and contain more people within them), the likelihood of being able to identify any given individual through their telephone number will probably remain greater than through a sample of their DNA. According to this concept, what would be the significance of the fact that a specific individual's DNA profile was not *actually* held on a searchable database? Would the fact they could not *realistically* be identified from a target sample of (their) DNA *in the instant case*, affect the classification of their DNA profile (drawn from 'target' sample) as 'personal data'?

It is clear that this concept would have to be refined to account for such possible distinctions. An ideal type that recognises the significance of context must take a position on whether the *actual* presence of a context capable of enabling identification is required, or whether simply the *possibility* of the data in question being placed within such a context is enough. Both approaches have to face difficulties.

If mere (theoretical) *possibility* were to be enough then this would have serious repercussions for the amount of information that such a concept would class 'personal data'. This may be illustrated by considering the different ways in which data may 'relate to' an individual and still *possibly* enable their identification.

Information may 'relate to' a person in a wide variety of ways, including (but not limited to);

I. Information provided by the person i.e. they may be a/the source of the information (e.g. gossip or advice)
II. Information derived from the physical person (e.g. blood pressure)

III. Information describing the person (either their physique, their personality, or their beliefs)

IV. Information describing the person's history (e.g. what they've done; where they've been; who they've known)

V. Information generated in order to identify the person (e.g. Name/ National Insurance Number)

VI. Information capable of affecting the person (i.e. changing something about their material circumstance e.g. a drop in salary or the health of a loved one)

The clearest examples of information capable of enabling the identification of an individual might be drawn from categories (ii) to (especially) (v). Information derived from an individual (i.e. ii) will invariably tell you something about that individual (i.e. iii). Something that describes an individual may clearly enable their identification in appropriate circumstances. Similarly, a description of an individual, or their past, (i.e. iii or iv) may enable the identification of that individual (v).

Information falling within categories (ii) to (v) may then be said to possess the common characteristic that they could (in appropriate circumstances) *identify* an individual. Information that falls within these categories may also share the characteristic that they may more intuitively be described as being 'about' an individual. They more obviously *inform* you of something about an individual and, perhaps on that basis alone, appear more deserving of the title personal information.

The question remains however whether information falling outside these categories is *incapable* of *possibly* identifying an individual. Can information that might not ordinarily be said to be 'about' an individual, nevertheless be capable of identifying them in appropriate circumstances?

While there is a *relationship* between the individual and the information in categories (i) and (vi), it appears true to say that the information in these categories might not inform you of anything *about* an identifiable individual (and may not intuitively be described as 'personal information'). It however appears equally true to say that they *could* contribute to the informational context (interpretive framework) that *enables* you to identify an individual *in the circumstances*. So, for example, if I know that a British vehicle registration plate has a particular format (and that it differs from that adopted by other countries), and I know that Mr Smith is the only English person registered as staying at a French campsite, I may locate Mr Smith by walking around the campsite looking for a vehicle with a British registration plate.

A piece of information may *enable* the identification of an individual without being *about* the person. It may be sufficient for the information to 'relate to' the individual in some more indirect way so long as, when placed within a broader informational context, that piece of information enables the individual's identification. The significance of this in the context of this concept is that *any* information may be said to relate to an individual in *some* way. *Any* information may then *potentially* enable the identification of an individual in appropriate circumstances.

A 'context dependent identification' concept which recognised the bare *possibility* of identification to be sufficient to justify defining data as 'personal' would clearly class a tremendous amount of information as 'personal data'. The sheer amount of 'personal

data', (but not necessarily the breadth) may be limited by emphasising that it is not the bare possibility in theory but the *actual* possibility, in fact, either in the *instant case* or generally, that is significant.

This concept recognises that while ANY information that relates to an individual may perhaps be theoretically capable of 'identifying' the individual given an appropriate context, that such contexts are not actually always (often) present. Also, such a concept may also allow for recognition that certain contexts are in fact much more likely to arise than others. To refer back to the example given above, it is probably more likely that, for most people, their telephone number will be placed within a context that enables their identification than a sample of their DNA. This example does however also point to the extreme difficult in accurately *prejudging* whether any particular piece of information will *actually* enable the identification of any given individual in a prospective case i.e. whether the relevant context will present in the future for *that* person. The *relative* availability of the relevant interpretative contexts, capable of 'making the link' between a particular data type and specific individuals, may however be estimated (albeit in a general way).

If one attempts to assess the *actual* possibility *in fact* of a particular piece of data enabling an individual's identification then one may focus upon either the possibility of *anyone* identifying the individual, or one may focus upon the possibility of *someone* in particular identifying the individual. If one attempts to do the former, one encounters the difficulty of knowing what interpretive frameworks are possessed by *everyone* i.e. there is an insufficient resource to ascertain whether absolutely anyone could identify an individual from a specific piece of data. If one attempts to do the latter, one encounters the difficulty that it is not possible to know with certainty what interpretive framework that an individual possesses *prospectively*.

The only hope would appear to be to assess whether the relevant context is *available* currently (and/or prospectively) (and/or retrospectively) to either everyone or some particular individual (depending on what has been adopted as the relevant yardstick). We may, thus attempt to assess, with relative accuracy, the *likelihood* of anyone identifying an individual from a particular piece of data i.e. by assessing the general availability of relevant context. As was indicated earlier, judgements may be made about the relative availability of telephone directories and DNA databases. The generality of such judgments, and their fallibility when applied to the instant case, is however clearly apparent.

In summary the Context Dependent Identifier model:

Has these advantages:

- It recognises the significance of the informational context.
- Allows determination of what is more, or less, likely to enable identification (in a very general way) by taking account of relative availability of relevant contexts (interpretive frameworks)

Has these disadvantages:

- It does not allow you to draw up a *definitive* list of data types that will 'always' or 'never' be personal data.

- It is not possible to limit the ways in which the data may 'relate to' an individual beyond those limits imposed by the possibility of identification. It is significant to note in this context that *any* information might contribute toward the identification of an individual if appropriate contexts were available.
- If a predictive judgment is to be made on whether a particular piece of data *will* be considered 'personal data' then fallible prediction must be made of whether the relevant context will be present.

IDEAL TYPE 3: The Context Independent 'Affects' model

The possibility (probability) of relevant effect

This ideal type is founded upon a concept that holds data will only be 'personal' if it is capable of 'affecting' an 'identifiable person' in a material way. It is clear that the notion of what constitutes a 'relevant' affect permits various interpretations.

One version of this concept, and the one that will be adopted here due to being favoured within the responses received, holds that only that information which affects an identifiable person's 'privacy' is personal data and, according to this model, whether data is capable of affecting an individual's privacy can be assessed in a context independent fashion.

It should be emphasised that those responses that seemed suggestive of this particular concept of personal data did not undermine the requirement that the data in question be linked to an identified, or identifiable, individual. The significant quality was however the character imparted to the information due to the quality of the link between it and that identifiable person. 'Personal data', according to this model, doesn't have to be able to identify the individual *itself*; it will be sufficient (but also *necessary*) for the information in question to be capable of affecting the 'privacy' of an individual that has been/may be identified *via* other means.

The first and, perhaps, most significant problem that seems to be faced by this particular ideal type is its reliance upon an apparently untenable concept of 'privacy'. In order to sustain a 'context independent' classification of personal data it would appear necessary to invoke a concept of 'privacy' capable of supporting the idea that specific types of information may affect dissimilar individuals' 'privacy' in similar ways.

The difficulties associated with such a concept of privacy have already been raised in Part A. The domestic law in The UK, and indeed elsewhere, does not provide a single definition of the term privacy and therefore it is unclear what might be included in a 'right to privacy' even in a legal context.[14] Moreover, it has been recognised that an explanation for the absence of a single definition may be found in the fact that any definition is unavoidably subjective. This gives rise to a particular problem: if it is undesirable to attempt to specify what needs to be included within a definition of privacy for it to be protected (due to the significance of context) then how may one hope to reliably, and consistently, identify what data might *affect* an individual's privacy (without also taking context into account)?

[14] Privacy and data-sharing: The way forward for public service, (The Performance and Innovation Unit) www.number-10.gov.uk

The deficiencies of such a concept if presented in this way are so manifest that it may be appropriate to offer a more generous interpretation. It may be possible to take account of the context *dependent* nature of privacy without entirely collapsing the distinction between this ideal type and the next. If what constitutes 'privacy' may be *determined* (to at least a certain extent) by *social context* then it may be possible to determine whether a particular data type will affect an individual's privacy independent of the contingencies of the individual (beyond the contingency of their residence within a particular society).

The suggestion that it may be possible to identify certain matters that may *generally* be considered 'private', and 'personal', seems consistent with this idea. It might be possible to identify certain types of data which are attributed with such significance *by society* that they will likely (perhaps even inevitably) have an effect upon the privacy of an individual. In this way, while the significance of context is not wholly avoided, its proper place may be realised within a perspective that reduces its relevance on a day to day basis.

It should be expressly recognised however that *even if it is correct to suggest that certain 'affects' to an individual's 'privacy' may be reliably predicted by taking account of a* society's *concerns with a particular type of data*, this only saves the concept of 'personal data' from 'context dependency' at a particular level of generality.

Also, the difficulties in establishing in practice, not only which data would be attributed with such significance by society, but the precise nature of an individual's relationship with wider 'society', and even *which* 'society' an individual 'relates' to, make it extremely difficult to confidently predict with accuracy the effect that a particular piece of data will actually have upon an individual's privacy in anything approaching a context independent manner.

Despite, these difficulties, one thing that may be said for this concept is that it may provide a more effective means of protecting an individual's privacy than an ideal type preoccupied exclusively with identification.

In summary the Context Independent 'Affects' ideal type:

Has the advantages that it:

- Potentially allows you to draw up a list of data types that will 'always' or 'never' be personal data (at least within a specific social context)
- Offers more specific protection of an individual's privacy than 'ideal types' centred simply on the notion of identification.

Has the disadvantages that it:

- Fails to explicitly recognise that considerations of context are, to some extent, unavoidable when assessing an individual's privacy.
- Recognises that in theory *any* information that might affect an individual in a relevant way could constitute their personal data
- Relies upon judgements about the likelihood of a particular piece of information having an effect upon individuals which may prove incorrect in specific cases.

IDEAL TYPE 4: The Context Dependent 'Affects' model

The possibility/probability of relevant effect

This ideal type holds that 'personal data' is data capable of impacting upon an individual's 'privacy'. Whether, in actual fact, a particular piece of information is capable of affecting an individual's privacy is determined by the contingent circumstances of the instant case: it is necessary to take account of the specific context.

This ideal type expressly recognises that the 'meaning' and 'value' attached to any piece of data for an individual's privacy will be determined by context. It may be distinguished from the previous model (within which a 'background' social context had to be acknowledged) by recognising the relevant context to extend to the contingent circumstances of the individual (potentially) affected in the instant case. This concept of personal data seems to employ a concept of privacy that is more consistent with the sociological and psychological literature on the subject which recognises privacy to be an *interaction* between an individual and others and/or environment.

As with previous ideal types described, this concept has to be developed in a way that will take account of the latent ambiguity within it. Does it recognise as 'personal data' any data which might possibly 'affect' an individual's privacy or does it only recognise that which may 'actually' affect a specified individual's privacy in the contingent circumstances? Also, similar to the previous 'context dependent' ideal type, the consequence of allowing merely 'theoretical' affect may extend the class of 'personal data' to include *any* information at all.

Interestingly, however, the consequence of restricting the scope of 'personal data' to only that which may actually affect an individual's privacy in the circumstances may not be to substantially restrict the *range* of information that might constitute personal

data. Given the intensely subjective process of interaction between self and environment an individual's privacy *could* be affected by just about anything.

Restricting the scope to the 'actual' possibility of a relevant effect in the circumstances would have the consequence that only *ex post facto* determinations could be made. Given the idiosyncrasies of an individual, and the constructive process that constitutes the generation of an identity, the circumstances would have to be *experienced* before their effect upon privacy could be known. That is not to say however that some kind of predictive judgment could not be made. It is just to emphasise the fallibility of such a judgement.

Any prediction as to the effect that a particular type of data will have upon an individual's privacy will clearly require a determination as to the (relative) *likelihood* of privacy being affected. In assessing likelihood, again as with previous 'ideal types' described, it may be necessary to have regard to the 'likely contexts' in order to assess the 'likely effect' that a particular piece of data will have. Unlike the previous ideal type associated with privacy however, this concept does allow for information about the contingent circumstances of the individual to be taken into account when performing this calculation. It must still be emphasised however that, while one may anticipate a particular data type as more or less likely to affect an individual's privacy in a particular way, this prediction is inherently fallible.

For example, it may be expected that information about the football team that one supports will not usually have the potential to affect banking transactions. It may also not usually be expected to affect one's physical safety. However, it might affect one's physical safety if one were to find oneself in a pub full of supporters of opposing teams (see the **Country 35** response to Q2, A, q.3(c)). In such circumstances, information about which football team one supports might assume 'sensitivity' that one would not normally associate with such information. In some ways this may actually be considered a more 'predictable exception' to the general rule than encountering a football biased bank manager. In both cases, the data puts the individual at risk (although, probably, a different kind of risk) and the chances of the risk *actually* materialising are difficult to accurately predict. The perceived risk of vulnerability may contribute toward the perception of whether the information in question is capable of affecting one's privacy.

If one wishes to assess the relative impact that various data types are likely to have upon an individual's privacy then it may be necessary to assume a stable context. That context will have to take note of not only what data is usually available to others within that particular context, but also how those various pieces of data could ordinarily be *used* within that context. The informational context will have to take account of the relationship between the data in question and the participants (both data subject and data controller).

A classificatory model constructed around 'affect' will then have to take note of the various *uses* to which different data types may be put (by all involved) and the impact that such uses may have upon an individual's privacy. The vulnerabilities exposed by the acquisition and use of different data types will have to be assessed. Many of these vulnerabilities may be commonly shared by people, e.g., most people are vulnerable to credit card fraud if they are careless with their credit card details, often however individual vulnerabilities may be difficult to anticipate.

In summary the Context Dependent Affects model:

Has the advantages that it:

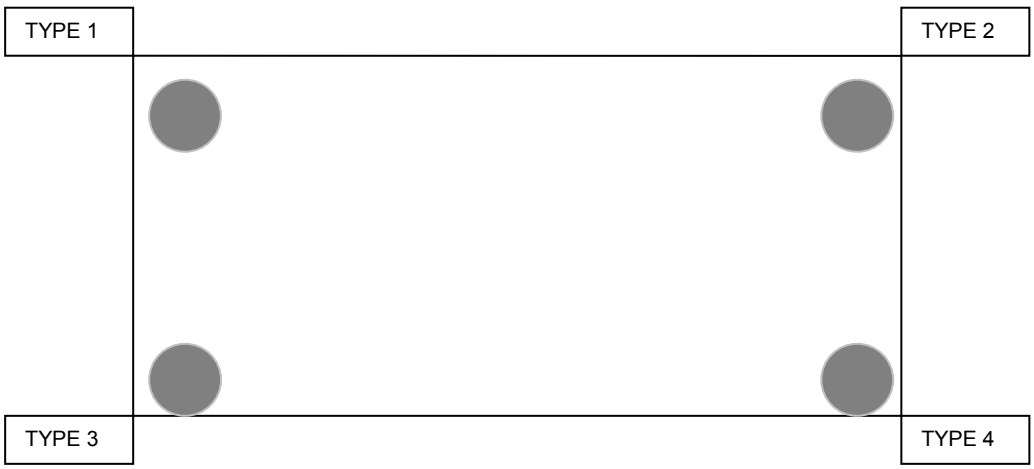
- Acknowledges the significance of context to the assessment of whether a particular piece of data will affect an individual's privacy
- Recognises that for an individual what information may affect their privacy may not be readily anticipated by another in advance
- *Any* information that might affect an individual in a relevant way would constitute their personal data.

Has the disadvantages that it:

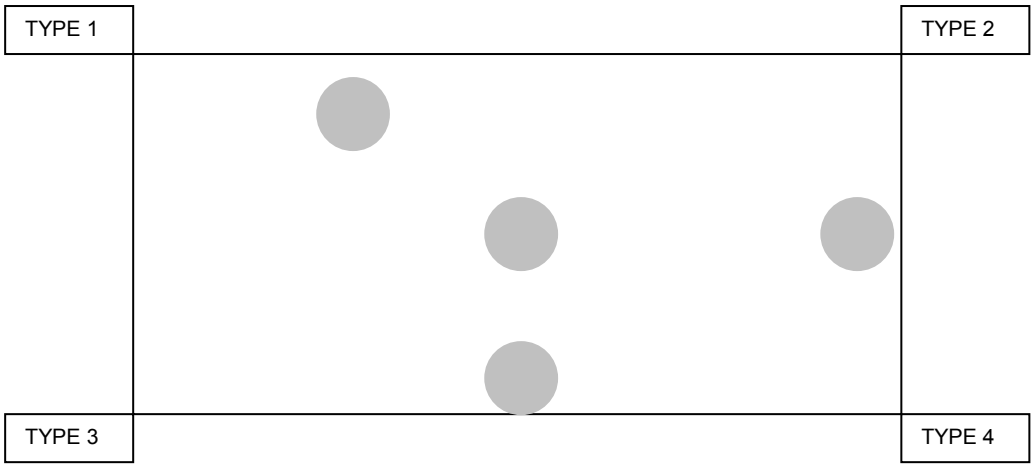
- Does not allow you to draw up a list of data types that will 'always' or 'never' be personal data.
- *Any* information might affect an individual in a relevant way given idiosyncratic vulnerabilities to the acquisition and use of information of different types.

The relation of the Ideal Types to one another

It is possible to illustrate the position of the ideal types described so far relative to one another within the following diagram. Each of the ideal types may be seen to occupy a position in one of the four corners described.



While the ideal types so far described have been located within one of the four corners of the above chart it is possible for a concept of personal data to fall *anywhere* within it. A concept of personal data could be developed that combined elements of the 'ideal types' portrayed. Such 'composite' concepts might be represented within the chart thus;



It is worth noting that countries may operate in practice with concepts that do combine in this way different elements of the ideal types specified e.g. a country may operate with a concept which considers personal data to be data that *both* identifies and affects an individual. Recognising that concepts operating in practice may constitute such composite concepts does not however undermine the relevance of the comments made about the ideal types. These comments simply have to be recognised as relevant only to one 'element' of the composite concept. Their continued relevance to the 'composite concept' overall will depend on how the elements interact. In understanding their interaction a more sophisticated 'understanding' of the concepts in operation may be developed.

It was noted in B4 that one group of countries appeared to operate with just such a concept underlying their decision making model. It may now be appropriate to return to that concept of personal data being operationalised by some countries.

The 'Identifies *and* Affects' Concept

This is the only 'concept' articulated by respondent countries that expressly requires the assessment of a particular piece of information in two different ways before it may be classified as 'personal data'. According to this concept information must be capable of not only identifying but also affecting an individual before it will constitute 'personal data'.

Countries occupying this position could operate with either a context dependent, or a context independent, concept of either 'identifies' or 'affects'. All of the countries whose responses seemed to most closely align with this particular concept indicated that they considered the classification of 'personal data' to be 'context dependent'. It is however unclear whether they consider 'identification' and/or 'affect' to be context dependent variables.

This in fact may point to one feature of this concept, indeed any concept that operates a composite of the elements identified in each of the 'ideal types', and that

is its complexity. Combining features found within the ideal-types necessarily adds complexity to the resultant concept and makes clarity of exposition especially important if the concept is to be properly understood.

From the questionnaire responses received it is not immediately apparent how these elements are combined in practice but clearly, in theory, any combination of 'context dependent identification', 'context dependent affect', 'context independent identification' and 'context independent affect' would be possible.

As well as the disadvantages that may be associated with increased complexity, there are a number of advantages to also be noted. Most obviously a more 'sophisticated' concept may help avoid some of the disadvantages associated with each of the ideal types identified above. For example, if the concept were to operate with the notion of 'context dependent affect' being a pre-requisite of personal data, then coupling this feature with another element, namely the additional requirement that personal data also be capable of identifying the individual (with identification being understood in a context independent manner), would avoid one potential disadvantage of operating exclusively with a 'context dependent' 'ideal type': the concept would no longer hold that *any* information could be capable of being personal data in appropriate circumstances.

If a 'context independent' notion of 'identification' were to be operated in tandem with a 'context dependent' notion of 'effect' then one consequence may be that certain information is *never* held capable of constituting personal data, due to the fact that it may *never* identify an individual. All other data would presumably be held to be 'personal data' 'sometimes' due to the fact that, while it might always be capable of identifying an individual, it might only sometimes also be capable of 'affecting' them in a relevant way.

Of course, simply combining elements doesn't necessarily disassociate them from the disadvantages that they have been previously associated with. In the above example (in which 'context independent identification' was associated with 'context dependent affect') the 'composite' concept would still suffer at least one deficiency associated with the notion of 'context independent identification': the idea that the possibility (or probability) of identification may not actually be ascertained in a context independent manner.

The disadvantage associated with the potential breadth of the information encompassed by a context dependent concept of personal data might however be ameliorated, if not avoided entirely, through *combining* elements of the two context *dependent* concepts described. If a concept of personal data were to be developed which held that data were only personal if it both identified and affected an individual then it may, even if both notions are understood in a context dependent way, offer an internal qualification of each component part.

Such a concept would isolate a particular way in which an individual's privacy may be affected, i.e. through their identification, and it would recognise only data capable of compromising an individual's privacy in this respect to be personal data. That is not to say that the data in question could *only* be capable of affecting an individual's privacy in this respect, it might *also* be capable of affecting their privacy in additional ways, but it would have to be capable of, at least, affecting this particular kind of intrusion into an individual's privacy. Likewise, this concept would not recognise *any*

identification of an individual to be sufficient to categorise the identifying data as 'personal data'. The identifier would itself also have to be capable of affecting an individual's privacy in a particular way: additional to simple identification.

It might be thought that *any* identification of an individual would, by definition, be capable of affecting an individual's privacy. This interpretation would not utilise a concept of 'privacy' consistent with the context dependent approach. An individual's identification does not *necessarily* constitute an invasion of their privacy: whether it does, or not, will depend upon the circumstances (this may be seen to be consistent with the notion of privacy being in some way defined by an *interaction* between an individual and others and/or their environment).

A concept of personal data specified in this manner may, through isolating a particular way that privacy would have to be affected, and through isolating a particular circumstance within which identification would have to take place, operate the notions of identification and effect in mutually restrictive fashion.

Such a 'composite' concept may result from an appreciation that there are certain disadvantages to be associated with any one of the ideal types. These disadvantages may be felt if relying upon any one of these ideal types exclusively when developing a strategy to make decisions on personal data in practice. In the next section (C4) we explicitly consider the challenges that would be faced by a country seeking to develop a decision making strategy reliant upon any one of these ideal types. We shall specifically consider the significance of each of the ideal types for a country attempting to construct such a strategy around a definition of personal data which utilises the key terms identified within Directive (95/46/EC).

Conclusions:

The Significance of "Ideal Types" and "Composite Types" for Decision-Making Strategies

From the empirical surveys and the literature review it is clear that countries do not use clearly defined concepts and models to establish which data are personal data. However, there are significant elements that are observable within the literature, in the ways that different countries make practical decisions about the classification of data types and then in the ways that they talk about their decision-making processes. From these significant elements, a series of ideal types can be established characterising ways of defining personal data according to the weight placed on single significant elements. A further step is then to make composite types by mixing the single significant elements together to reduce the disadvantages associated with the single element ideal types. From this, by way of conclusion, the implications and possible outcomes and experiences that a country relying on one of the ideal types in decision-making will experience can be considered and the difficulties of the disadvantages of each type can be assessed. The ideal types and composite types are tools, with robust justifications, that could inform the decision-making of countries, and in particular The UK. The final evaluation of the appropriate concept and definition of personal data to apply rests with the competent authority (and not researchers): the models give theoretical frameworks, inspired by empirical observation, within which to think and to make the decision.

It should be clear that, given the issues that may be raised with each of the ideal types, any country relying exclusively on just one of them when implementing a decision making strategy will face a number of challenges. The precise nature of these challenges will depend, in part, upon the formal definition of 'personal data' adopted within the authority's jurisdiction. The words within their formal definition will have to be interpreted within the classificatory strategy developed.

Any country aiming to design a decision making strategy consistent with Directive 95/46/EC would then have to meet some particular challenges. Crucial to an understanding of the Directive are those key terms identified in Section A of this report. How might these terms be understood in the light of a decision making strategy based on each of the ideal types described? How might each of the concepts described assist an interpretation of the key terms? What other, more general, issues would be faced by countries relying upon any one of the ideal types described to found a classificatory strategy?

Unique Identifier

A country whose decision making strategy drew almost exclusively on the Unique Identifier ideal type would face a serious difficulty arising from the fact that very little (if any) data could truly be classed as a unique identifier. Failing to develop a strategy capable of accounting for context would lead to only that data unique across *all* contexts being labelled personal data. The result would be that very little data could possibly be described as 'personal data' by such a classificatory strategy.

A potential solution to this difficulty is to classify data as personal or otherwise according to its 'degree of fit' with the ideal of the unique identifier. As suggested in C3, DNA is perhaps the paradigmatic case of a unique identifier. DNA would thus

probably sit at the top of a scale, with all other data types falling somewhere below DNA on that scale. On this basis, one could see how 'fingerprints' might be considered closer to the top of the scale than 'name'. 'Name' itself may be more 'unique' than 'shoe size'. 'Degree of fit' may then be ascertained through establishing the relative number and availability of contexts within which a specific piece of data may retain its status as a unique identifier.

A classificatory strategy built upon the unique identifier concept may thus recognise the significance of context but yet still use the unique identifier ideal type as its guiding principle. In effect, it is the practical, real-life manifestation of the unique identifier ideal type. The closer a particular identifier is to being 'unique' across all contexts, i.e. in *all* circumstances, the more likely it is to be classed as personal data.

This strategy does not actually draw a distinction between two discrete categories of data. Instead, it assigns all data types a position on a continuum. This continuum itself only has relevance within a specific context. To take a rudimentary example, the data type 'hair colour' may be largely useless as an identifier in a nationwide context. However, in the context of a single room it may be an extremely effective identifier indeed. To place 'hair colour' on any continuum it is therefore necessary to at least implicitly assume a context. As there are relatively few contexts within which 'hair colour' may constitute a 'unique identifier' (when assessed nationally) it would be unlikely to be classed as 'personal data' by this classificatory strategy (if national context were adopted as the relevant context).

The challenge faced by any country operating this strategy is in justifying where to 'draw the line' between data that is to be considered personal and data that is not. The solution to this question cannot be found within the concept itself for the ideal type seeks to relegate the importance of context. Any country working with a Unique Identifier concept therefore has to find some way of establishing at least three things: the relevant context within which to judge the 'uniqueness' of a piece of data, the appropriate height at which to set the 'bar' of uniqueness and, finally, whether a specific piece of data is sufficiently unique within that context to 'clear the bar'.

If the construction of a classificatory strategy upon this ideal type necessitates reference to context (if a paucity of personal data is to be avoided) then this has some significance to the interpretation that this concept would offer of at least one of the 'key terms' identified as relevant to an understanding of Directive 95/46/EC. Were reference to context *avoided*, and the ideal type unmodified, then it may support interpretation of 'direct' identification as identification enabled by a truly unique identifier i.e. identification enabled by a piece of data that may be 'unpacked' to provide all of the information (substantially) need for identification. It may then also support interpretation of 'indirect' identification as that identification made possible only via such cross reference with extrinsic information (e.g. database) as may be associated with an 'attributed' (and potentially non unique) identifier.

The recognition of context does however go some way to undermine this simple bifurcation. Even those identifiers which retain their unique status *across* contexts are only capable of identifying an individual when placed *within* those contexts e.g. even a head and shoulders photograph of an individual will only directly identify that individual if the identifying individual possesses sufficient information, or may posses it, to uniquely link the photograph to an individual (they may, for example, be in the

same room at the same time). [15] Invariably then, data operates as a unique identifier because, *within the context,* it both enjoys a unique status *and* functions as an identifier.

This places rather a different hue upon the difference between 'direct' and 'indirect' identification. The difference between 'direct' and 'indirect' identification no longer appears resident within the nature of the data itself, but rather within the nature of the context within which the data resides. The difference between the two may be that data may enable 'direct identification' when occurring within a context that already possesses sufficient information to enable identification. This reduces description of data as enabling only 'indirect' identification to a simple description of the need for a change in context. Whether a specific piece of data will enable identification either directly, or indirectly, thus depends upon whether sufficient information is already present to attribute the identifier with 'unique status' and to 'directly' enable identification.

If, in constructing a decision making strategy on this ideal type, a country were to adapt it in order to take account of context then it would still be necessary for the country to establish whether it is simply the existence, or the accessibility, of the additional relevant/necessary information that is significant. If it is accessibility that is crucial, then there is also a question of *who* is able to access it (e.g. the data controller (as in Country 17 and The UK) or absolutely anybody).

Perversely, the unique identifier concept may not provide any guidance on the nature of identification itself. While it classifies a data type according to its relative status as unique, the concept itself doesn't necessarily presuppose any particular concept of identification. This is an issue which may be seen to run through all of the ideal types described (see, 'Towards a composite approach?' below).

Context Independent Affects

A decision making strategy based on the Context Independent Affects ideal type faces similar general problems to the Unique Identifier variant. To create a prospective list of data that will 'affect' an individual is an arduous task. Narrowing the scope of 'affect' to 'an effect on privacy' does not make matters any easier. This is because a strategy based upon the Context Independent Effects ideal type relies upon the possibility of reliable judgements on the effect particular information would have upon individuals. It will be extremely difficult to provide such reliable judgements in advance.

However, as outlined in C3, a decision making strategy based on this ideal type may be able to find room to take account of the context dependent nature of privacy. Such a decision making strategy may accept that what constitutes privacy is indeed determined to a degree by *social context*. This may then be taken into account in a classificatory strategy. However, any individual contingencies or particular circumstances will *not* be taken into account. As argued earlier, such an approach does not *wholly* ignore the significance of context, but it assigns to context a more limited role on a day to day basis.[16]

[15] Even then, the 'directness' of the identification may be arguable. See discussion of the difference between data and information in the Literature Review (A2).
[16] See C3 for a fuller exposition of this point.

However, an approach of this nature comes with its own special set of difficulties. Perhaps the most fundamental of these would be the need to specify which society an individual relates to, and having done that, to ascertain which data types could be said to unavoidably affect an individual's privacy in that society. The most obvious 'society' that an individual can belong to is perhaps the national society. However, one could make the case that an English region or a Country 6 Lander (or perhaps a particular socio-economic or demographic group) better represents the 'society' to which any given individual 'relates'. Of course, the further down this road one goes, the further away from the ideal of a context independent test one moves.

This reinforces the seriousness of the challenges associated with operating any context *independent* approach. By definition context independent concepts do not allow for contextual factors, such as the availability of databases or 'frameworks of understanding', to be taken into account. In reality however, rather than any inherent quality of the data itself, it is precisely the influence of such contextual factors that enables data to function as an identifier of, or to produce some effect upon, an individual.

As with the 'unique identifier' ideal type, a decision making strategy based on the Context Independent Affects ideal type would struggle to provide any guidance on the notion of 'direct and indirect identification' contained in the Directive. This is primarily because identification, be it direct or indirect, is not central to this decision making strategy. However, a country seeking to operate with this type of model would have to address the issue of how identification could fit in with their approach, if it is an approach that is to remain consistent with the Directive. [17]

Context Dependent Strategies

Recognising the significance of context and operating with a context dependent decision making strategy is however not itself free from difficulty. The same fundamental criticisms can be levelled at both the 'identifier' and 'affects' variants of the context dependent family.

Any country that utilises a context dependent approach can potentially choose between a decision making strategy holding personal data to be that data which *actually* identifies or affects an individual, or a strategy holding that personal data is that data which could *possibly* identify or affect an individual.

A decision making strategy based on the Context Dependent 'Identifier' or 'Affect' ideal types, which took the position that data was personal if it could *possibly* identify or affect an individual, would be forced into an awkward situation. This is because *any* data can identify or affect an individual *in the right circumstances*. Clearly it is not practical to afford protection to *all* data.

Therefore, any country operating with such a context dependent strategy may be compelled towards acting in a reactive fashion. The country will be driven to ask, not if the date could *possibly* identify or affect the individual, but rather whether it *actually* does so in the set of circumstances at hand. This would reduce the amount of data

[17] This is because the Directive states that 'personal data' "shall mean any information relating to an *identified or identifiable* natural person" (Article 2a).

classed as 'personal' but it would also have another consequence that may not be considered so desirable.

The strategy would accept that the *particular context of each individual case* was the critical factor in enabling the data to either identify or affect the individual. Assessing the *actual* context of *each individual case* would however have the consequence that the classification of data as personal could only occur 'after the fact': when the particular context of the case can *actually* be known. A purely *reactive* decision making strategy of this kind may be undesirable due to the lack of predictability it causes. No list of data to be classed as personal could be drawn up under this approach, and this would contribute to the difficulty in predicting with any confidence if data would be considered personal in advance of any particular case.

A potential alternative to this decision making strategy would aim to provide clearer guidance on whether data would be classed as personal in a future case. It would seem to steer a path between the '*possibly'* and the '*actually'* approaches. This approach suggests that if the context enabling the data to identify or affect an individual would '*probably'* arise then the data may be classed as personal data. This decision making strategy allows the country to make a proactive, prospective judgement on which data is to be classed as personal and afforded protection without drawing *all* data into that category.

This approach would allow for the drawing up of a 'list' of data that would be considered 'personal' because of the 'probability' of relevant context arising. If this list were considered definitive then the country would have to accept that, in any given case, data on this list may not *actually* be personal data and certain personal data may not be on the list. An alternative would be to hold the list purely indicative: listing data which will *probably* be held 'personal data'. A determination on any particular piece of data would however be made in the circumstances of the instant case.

Such an approach may represent a more workable (but by no means problem free) decision making model than either the 'possibly' or the 'actually' models. A judgement on which data to include upon such a list would involve predicting which contexts or circumstances are most likely to occur, or are dominant in day to day situations. If data could identify or affect an individual in such circumstances, then it could be classed as personal. However, this approach relies on judgements on which contexts or circumstances are likely to arise or are dominant. These judgements would be extraordinarily difficult to make and they would most certainly be fallible.

Some of the difficulties associated with holding a context dependent line *and* creating lists may be illustrated by looking at the responses given to Questionnaire 2, Part 2, on the 'significance of context'. This question provided the respondents with a set of relatively stable contexts, and asked them to class various data types as personal or otherwise within the particular contexts given.

Some countries declined to fill in the table provided. However, the reasons for not doing so varied considerably. **Country 37**, for instance, stated that "[we] do not see how, for the most part, context matters". Conversely, **Country 35** felt unable to complete the table because "context is very significant". They felt that the 'stable

contexts' outline in Part 2 contained "insufficient information", particularly as there was no indication of how the information was to be used.

The countries that did fill in the table demonstrated a considerable degree of variation. In the context of 'Police Force', for example, only 'name' commanded any significant level of agreement. All respondents felt that it was 'more likely' to be personal data. None of the other data types could engender anything like this level of consensus. National Insurance Number, for example, was ranked at both 10 (most likely to be personal data) by **Country 27** and 2 (with 1 as least likely) by **Country 40** (and others).

This lack of consensus demonstrates the difficulties in creating a prospective list of personal data, even where future context is relatively stable. This may be partly due to the different weights each country attaches to the other variables of 'identify' and 'affect'. Even then, it is clear that while accepting the importance of context for any DPA, simply acknowledging that importance does not automatically provide a problem-free solution to the issue of classification.

Towards a composite approach?

It is clear that operating with a decision making strategy based overwhelmingly on one ideal type presents some serious difficulties. Problems of conceptual coherence face countries that ignore the part context plays in making data personal. However, to operate a model totally dependent on context would lead to a degree of unpredictability that is probably unacceptable to most countries. As suggested above, one potential response to this problem is to construct a 'composite' approach that draws on ideas and themes from more than one ideal type in order to construct a more robust decision making strategy.

Recognition of the advantages and disadvantages associable with the different paradigms might also help in understanding those questionnaire responses where it was more difficult to identify any clear concept or guiding principles. **Country 22** and **Country 33**, for instance, provided responses where it is difficult to clearly identify a favoured ideal type. **Country 22** appear to implicitly attach a significant amount of importance to context. However, this makes their claim that some pieces of data are 'never' personal data difficult to sustain. If an individual *was already identified and the data types in question linked to them*, then how could certain data types 'never' be personal data. **Country 22** may consider that the data types in question do not 'affect' the individual in a relevant way, thus justifying their classificatory stance. However, such a line remains difficult to hold, as any potential 'affect' is itself surely dependent upon context. Likewise, **Country 33**, who also appear to be close to the 'context dependent' conceptualisation insist that some pieces of data are 'never' personal data. If context is determinative then how can such blanket statements be justified? These apparent anomalies may perhaps be indicative of an awareness (implicit or explicit) of the relative strengths and weaknesses of a decision making model based on one ideal type. Both **Country 22** and **Country 33** appear to favour a decision making strategy informed by the context dependent ideal types. The responses that appear inconsistent with this line may be motivated by a wish to avoid the disadvantages that are associated with a strictly context dependent paradigm.

Countries attempting to avoid the disadvantages associated with developing a decision making strategy upon an ideal type of personal data may choose to either incorporate a number of 'ideal types' (or derivative concepts) within their strategy, or, incorporate *a* bespoke concept specifically designed to hold the various elements identified within the ideal types in optimal relation to one another. The difficulty associated with the first response is that the operation of the classificatory strategy developed may be difficult to predict. There is also the danger of internal inconsistencies if inconsistent concepts are relied upon in materially similar cases.

The difficulty associated with the second response is in the development of such a concept in the first place. Furthermore, even if such a 'composite' strategy could be cultivated, it may not provide a trouble free solution. However, a self-conscious, deliberate, and explicit positioning by a country of their operative concept of personal data amongst the 'ideal types' described may assist with the transparency and predictability of their strategy.

Decision Making Models and the key terms 'Identification' and 'Natural Person'

The decision making strategies outlined in this section can provide only the framework for the classification of data as personal or otherwise. They cannot, in isolation, explain all the key terms that were discussed in the literature review in Part A. The focus of the decision making strategies is on the meaning of 'relating to', for as has been demonstrated throughout the course of the report, the interpretation of this seemingly innocuous phrase has the most serious repercussions for classification.

However, the way in which those terms that are not directly addressed by the decision making models or ideal types are interpreted also has consequences for the classification of data as personal. A good example of this is provided by 'identification'. None of the decision making models engages with the idea of 'identification'. Yet clearly, what one understands by 'identification' will have an enormous impact on what is classed as personal data.

'Identification' can potentially refer to at least two very different concepts. The first could be termed '*handshake'* identification. This concept of identification requires that the individual concerned can actually be physically located, in order to enable a 'handshake' to take place. The second could be termed '*isolate and affect'* identification. This holds that no such physical location is required; instead, 'identification' is achieved if an individual can be effectively isolated from others and deliberately targeted in some way. Such identification may be regularly realised within electronic environments. An example may be provided by those individuals who 'date' online in chat rooms while concealing details of their 'real' identities or physical locations. While incapable of locating each other for the purposes of a 'handshake' they may nevertheless be able to consistently and reliably 'identify', and 'affect', each other in their virtual environment.

The decision on which concept of identification to utilise is based is an important one, as it has serious repercussions. One such repercussion is that each notion of 'identification' supports a particular view of how far the term 'person' may be

extended. Under the 'handshake' concept of identification, it would be difficult to see how the individual to be identified could be anything other than a living, natural person. However, if the second 'isolate and affect' concept of identification is employed, then this could *potentially* apply to legal persons. In fact, taken to its extreme, such a concept of identification could legitimise the protection of personal data belonging to imaginary persons.[18]

A country may not find any explicit guidance on which concept of identification to favour from the ideal types outlined in C3. The ideal types, in isolation, appear to offer no direction on whether the concept of identification ought to be understood in the 'handshake' or 'isolate and affect' sense. However, a decision making strategy that employed a composite concept of personal data: one requiring identification and affect *to privacy* may provide some guidance. Identification would only be significant if 'isolation and affect' were sufficient to affect an individual's privacy. This may itself often be associated with the possibility of 'handshake' identification, but it would not necessarily be so limited. As has been apparent, time and time again, context would be all.

Clearly, then, while the ideal types may not themselves provide a comprehensive understanding of the key terms that make up any definition of 'personal data', they may assist toward an interpretation that is conceptually coherent while remaining rooted in practice. A decision making strategy constructed through self conscious adoption of a composite concept may recognise the limitations associated with each of the ideal types. The development of such a decision making strategy is bound to be a challenging task. Nevertheless, if it is held explicitly and openly it may aid the transparency, accountability and predictability of any data protection authority's decision making strategy.

[18] Of course, the Directive uses the term 'natural person', perhaps to avoid such an extension of its scope. Nevertheless, this has not prevented Korff from arguing that the principles of data protection should be extended to cover legal persons (see Literature Review (A2)).

Appendix One: Summary of International Legislation (definitions of "personal data")

Country 36:
personal information means information or an opinion (including information or an opinion forming part of a database), whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion.

Country 1:
"Data" ("Personal Data") Information relating to data subjects (sub-para. 3) who are identified or identifiable; Data are "only indirectly personal" for a controller, a processor or recipient of a transmission when the Data relate to the subject in such a manner that the controller, processor or recipient of a transmission cannot establish the identity of the data subject by legal means;

Country 37:
''personal information'' means information about an identifiable individual, but does not include the name, title or business address or telephone number of an employee of an organization.

Country 22:
"personal data" or "data" means any information relating to a living data subject; consolidated data of a statistical nature, from which the data subject cannot be identified, are not deemed to be personal data.
(person means any natural person or any public or private corporate body whether or not it has legal personality and includes the Government of the Republic)

Country 3:
'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject');

Country 6:
"Personal data" means any information concerning the personal or material circumstances of an identified or identifiable individual (the data subject).

Country 7:
"Personal data" means any information concerning the personal or material circumstances of an identified or identifiable individual (the data subject).

Country 8:
"Personal data" means any information concerning the personal or material circumstances of an identified or identifiable individual (the data subject).

Country 9:
"Personal data" means any information concerning the personal or material circumstances of an identified or identifiable individual (the data subject).

Country 33:
"personal data" means data which relate to a living individual who can be identified-
(a) from those data; or
(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual;

Country 35:
"personal data" means data which relate to a living individual who can be identified -
(a) from those data, or
(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,
and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual;

Country 12:
"personal data", as any information relating to a natural or legal person, body or association, identified or identifiable, directly or indirectly by reference to any other information, including a personal identification number;

Country 34:
"Personal data" means data consisting of information which relates to a living individual who can be identified from that information (or from that and other information in the possession of the data user), including any expression of opinion about the individual but not any indication of the intentions of the data user in respect of that individual.

Country 27:
Personal data - any information relating to a natural person - the data subject who is identified or who can be identified directly or indirectly by reference to such data as a personal identification number or one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

Country 40:
Personal information'' means information about an identifiable individual; and includes information contained in any register of deaths kept under the Births and Deaths Registration Act 1951:

Country 20:
Personal data: any information and assessments that may be linked to a natural person,

Country 29:
Within the meaning of the Act personal data shall mean any information relating to an identified or identifiable natural person.

Country 17:
Personal data - All kinds of information that directly or indirectly may be referable to a natural person who is alive.

UK:
"personal data" means data which relate to a living individual who can be identified-
(a) from those data, or
(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,
and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual;

Appendix 3: Questionnaire 1

The University of Sheffield

Department of Law

Crookesmoor Building
Conduit Road
Sheffield S10 1FL
Tel: 0114 222 2000

Dear Sir/Madam,

We are a group of researchers at the University of Sheffield (UK). We are working on a research project for The UK Information Commissioner, focussing on the question "What are Personal Data?". As part of the research specification, the Commissioner has asked for a survey of the views taken by other countries as to the meaning of personal data and the practical issues that other countries have experienced in defining the coverage of the law in this respect.

The methodology that we have adopted involves a short email questionnaire which we are distributing to a sample of countries. We are sending the questionnaire to you since you responded positively to a previous email sent by our research team. This questionnaire covers general questions and basic principles and should not take too long for you to complete.

We would be grateful if you could complete the questionnaire and return it to us by Wednesday 24th December 2003. Completed questionnaires should be returned to us via email to : D.Moxon@sheffield.ac.uk

The questionnaires will be analysed by the research team and the major findings presented to The UK Information Commissioner in the form of a written report. The findings will be made available to participants in the project and published in selected academic journals. Due to the comparative nature of the project, it may be necessary to refer to the specific practices adopted by individual countries. We will not refer to named individuals unless explicit consent for this is sought and obtained in advance.

We will use a follow-up questionnaire in order to explore some of the issues in more detail. The second questionnaire will use scenarios to highlight problem areas and different approaches shown in the first questionnaire. Again, this questionnaire will be kept as brief as possible.

If you have any questions, the research team can be contacted via the email address above.

Many thanks for your time.

Yours sincerely,

David Townend
Mark Taylor

Natasha Semmens
David Moxon

Questionnaire

SECTION 1: Examples of data that might be described as 'personal data'.

1. Please indicate whether the following information about an individual would ALWAYS, NEVER or SOMETIMES be considered 'personal data' in your jurisdiction? If SOMETIMES, please indicate the circumstances in which the information would NOT be considered personal data.

	Always	Never	Sometimes This information would NOT be considered 'personal data' if…
Name	☐	☐	
Home Telephone number	☐	☐	
Shoe size	☐	☐	
National registration number	☐	☐	
Blood group	☐	☐	
Countries visited in last 5 years	☐	☐	
Salary details	☐	☐	
Head and shoulders Photograph	☐	☐	
Political party voted in last election	☐	☐	
Car registration/licence plate number	☐	☐	
Email username and password	☐	☐	
TV viewing habits	☐	☐	
Dental Record	☐	☐	
Sexual orientation	☐	☐	

Religion	☐	☐	
Parents' Names	☐	☐	
Credit card number	☐	☐	
History of Addiction	☐	☐	
DNA profile	☐	☐	
Details of time, place and cause of death of data subject	☐	☐	
Education/qualifications	☐	☐	
E-commerce transactions	☐	☐	
Mothers' Maiden Name	☐	☐	
Fingerprint	☐	☐	
Medical History of Family Members	☐	☐	
Bank Account Details	☐	☐	
Computer IP address	☐	☐	
State Benefit Received	☐	☐	
Alias/ pseudonym used in internet Chat Room	☐	☐	
Football Team Supported	☐	☐	
Family Portrait (painting)	☐	☐	
Vehicle Ownership (make, model and colour)	☐	☐	
Date of Birth	☐	☐	
Still image taken from Closed Circuit Television (CCTV)	☐	☐	
Natural Hair Colour	☐	☐	

SECTION 2: The term 'personal data' is defined in Directive 95/46/EC:

'personal data shall mean any information relating to an identified or identifiable natural person ('data subject)…' (Article 2, para. (a))

The questions in this section relate to the interpretation of this definition in your jurisdiction.

2. **How have you formally defined the following terms (a-e) in your jurisdiction? If so, please type in the formal definition in the space below each term.**

   a) **'information'**

   Answer:

   b) **'data'**

   Answer:

   c) **'relating to'**

   Answer:

   d) **'identified or identifiable'**

   Answer:

   e) **'natural person'**

   Answer:

3. Have you had any problems with the interpretation or definition of these terms? Please type in your answer in the space below each term. Please give examples or attach cases/documents to illustrate your answer.

a) 'information'

Answer:

b) 'data'

Answer:

c) 'relating to'

Answer:

d) 'identified or identifiable'

Answer:

e) 'natural person'

Answer:

4. Article 2(a) distinguishes between 'direct' and 'indirect' identification.

a) Has 'direct' identification been formally defined in your jurisdiction?

If yes, please give the formal definition:

If no, what do you understand the term to mean?:

b) Has 'indirect' identification been formally defined in your jurisdiction?

If yes, please give the formal definition:

If no, what do you understand the term to mean?:

SECTION 3: The directive also refers to the anonymisation of data and the storage of
Data (Article 2 (c), Recitals 15, 26, 27). In this section we would like to ask you questions about these issues.

5. For the purposes of the directive what do you understand by the term 'personal filing system'?

6. Would you consider the following to constitute a 'personal data filing system'? (Please tick one box for each example)

	Always	Never	Sometimes
Newspaper	☐	☐	☐
Card Index	☐	☐	☐
Organisational Filing Systems	☐	☐	☐

Photograph Album	☐	☐	☐
Diaries	☐	☐	☐
Electronic Databases	☐	☐	☐
Electoral Registers (if applicable)	☐	☐	☐
Registers of Birth, Death and Marriage	☐	☐	☐
Membership Lists of Voluntary Organisations	☐	☐	☐
Archived minutes of Meetings	☐	☐	☐
CCTV Footage	☐	☐	☐
Organisational Websites	☐	☐	☐
Telephone Directories	☐	☐	☐

7. Is the process of anonymisation ever capable of transforming 'personal data' into 'non personal data'?

Yes ☐
No ☐

If yes, please answer questions a) – c)

a) Do you distinguish between different methods of anonymisation?

Yes ☐
No ☐

b) Why?

c) What are your procedures for monitoring the anonymising process?

8. Is there anything else you would like to tell us about your interpretation of the term 'personal data'?

Appendix 4: Questionnaire 2

The University of Sheffield

Department of Law

Crookesmoor Building
Conduit Road
Sheffield S10 1FL
Tel: 0114 222 2000

Dear Colleagues,

UK Information Commissioner "What are Personal Data?" Research

Thank you for your responses to our first Questionnaire. We had a very good response from a variety of Authorities and the full responses that you gave produced fascinating results. As we indicated in our preliminary methodology, we have a number of issues that we need to clarify and some of our theoretical understandings that we need to test. We wish to do this in the attached second and final Questionnaire.

We are very keen to understand more fully how, why and when you classify information as 'personal data'. Therefore, part of the Questionnaire relates directly to your responses to Questionnaire 1, with questions to help us to make sure that we have correctly understood your approach. Part B then offers a series of short scenarios, designed to focus on different aspects of personal data that particularly trouble us. Part C gives a series of statements that are designed to provoke rather more open general indications of how you understand the concept of 'personal data'.

Under the terms of our contract with the Information Commission, all the data relating to this study is owned by the Information Commission. We will ensure that in the final report and any subsequent publications resulting from this work, the data is anonymised by not referring to any country or individual by name.

We know how valuable your time is and we are enormously grateful for your participation in this study. We hope that you will agree that the question posed by the Information Commission, 'What are Personal Data?', is of great interest and that finding a robust definition with as wide an input as possible, will be of great value to the Data Protection Community in informing and advancing its debate on the point.

The timescale for the whole work is very short and so we respectfully ask if the responses to Questionnaire 2 could be with us by email to _____ @sheffield.ac.uk by Friday _____ February 2004.

We very much look forward to your response and once again thank you for your participation.

Yours sincerely,

David Townend
Mark Taylor
Natasha Semmens
David Moxon
Sharon Booth

Questionnaire 2

NOTE: The specific content of Section A was different for each country, as it was a follow-up from the answers given to the first questionnaire (See A3). Rather than include the content for each individual country in this Appendix, we have only included the Section A designed for Country 36. Sections B and C were consistent across all countries.

The University of Sheffield

Department of Law

Crookesmoor Building
Conduit Road
Sheffield S10 1FL
Tel: 0114 222 2000

Dear Colleagues,

UK Information Commissioner "What are Personal Data?" Research

Thank you for your responses to our first Questionnaire. We had a very good response from a variety of Authorities and the full responses that you gave produced fascinating results. As we indicated in our preliminary methodology, we have a number of issues that we need to clarify and some of our theoretical understandings that we need to test. We wish to do this in the attached second and final Questionnaire.

We are very keen to understand more fully how, why and when you classify information as 'personal data'. Therefore, part of the Questionnaire relates directly to your responses to Questionnaire 1, with questions to help us to make sure that we have correctly understood your approach. Part B then offers a series of short scenarios, designed to focus on different aspects of personal data that particularly trouble us. Part C gives a series of statements that are designed to provoke rather more open general indications of how you understand the concept of 'personal data'.

Under the terms of our contract with the Information Commission, all the data relating to this study is owned by the Information Commission. We will ensure that in the final report and any subsequent publications resulting from this work, the data is anonymised by not referring to any country or individual by name.

We know how valuable your time is and we are enormously grateful for your participation in this study. We hope that you will agree that the question posed by the Information Commission, 'What are Personal Data?', is of great interest and that finding a robust definition with as wide an input as possible, will be of great value to the Data Protection Community in informing and advancing its debate on the point.

The timescale for the whole work is very short and so we respectfully ask if the responses to Questionnaire 2 could be with us by email to D.Moxon@sheffield.ac.uk by Friday 14th February 2004.

We very much look forward to your response and once again thank you for your participation.

Yours sincerely,

David Townend
Mark Taylor
Natasha Semmens
David Moxon
Sharon Booth

SECTION A

Part 1. Analysis of Question 1 responses.

The table on this page shows the responses you gave to Question 1 in Questionnaire 1, as to which types of information are always, never or sometimes personal data.

Always	Never	Sometimes
National registration number		Name
		Home telephone number
		Shoe size
		Blood group
		Countries visited in the last 5 years
		Salary details
		Head and shoulders Photograph
		Political party voted in the last election
		Car registration/licence plate number
		Email username and password
		TV viewing habits
		Dental record
		Sexual orientation
		Religion
		Parents names
		Credit card number
		History of addiction
		DNA profile
		Details of time, place and cause of death of data subject
		Education/qualifications
		E-commerce transactions
		Mother's maiden name
		Fingerprint
		Medical history of family members
		Bank account details
		Computer IP address
		State benefit received
		Alias/pseudonym used in internet chat room
		Football team supported
		Family portrait (painting)
		Vehicle ownership (make, model and colour)
		Date of birth
		Still image taken from CCTV
		Natural hair colour

We have a few further questions that we would like to ask you, based on this table, to assist our understanding of your interpretation of certain terms.

1. We were very interested to see that you indicated that only National Registration number would 'always' be considered personal data. What is it about this data that distinguishes it from any other type of data?

2. For all the types of information in the 'sometimes' category, will different data types need different conditions before they become personal data? If so, how do you communicate that to potential data controllers?

3. We were very interested to see that you indicated that there is no information that would 'never' be considered personal data. Can you think of any types of information (not listed here) that would <u>not</u> be classed as personal?

Part 2. Analysis of Question 6 responses.

The table on this page shows the responses you gave to Question 6 in Questionnaire 1, as to what constitutes a 'personal filing system'.

Always	Never	Sometimes
Electoral register Register of Births Deaths and Marriages Membership voluntary organisation Telephone directory		Newspaper Card index Organisational filing Photo album Diary Electronic database Archived minutes of meetings CCTV Organisational websites

Your responses are very interesting and we would like to ask you a few additional questions in order to help us to understand the status of different types of information storage:

1. What is the difference between a 'telephone directory' and a 'card index'?

2. Can you suggest some circumstances in which the following examples *will* and *will not* be regarded as a 'personal filing system'?

a) Newspaper
b) Diary
c) CCTV footage

3. Are there any types of data storage (not listed here) which will never be regarded as a 'personal filing system?

SECTION B
PART 1 - Scenarios

Scenario 1

Having bought a local grocery store, Gordon Rocer finds an order book and five video tapes recorded from a security closed circuit television camera (CCTV) in the shop.

The order book lists the orders that customers have made by telephone for home delivery. The lists include food and goods ordered, and are headed by just the surname of each customer and no other identification.

The video tapes are from a single fixed camera in the store. The images are good enough to see facial features of customers as they approach the counter to purchase goods.

1. Would you consider the **order book** to contain personal data if...

a) The grocery store was located in a densely populated area and the surnames could apply to any one of a number of individuals?

Why?

b) The grocery store was located in a small village and some of the names could only possibly apply to one individual or family?

Why?

2. Would you consider the **video tapes** to contain personal data if...

a) The grocery store was located in a densely populated area, attracted a lot of passing trade and had few regular customers?

Why?

b) The grocery store was located in a small village, attracted very little passing trade and had very many regular customers?

Why?

Scenario 2

'MusicMaker' is an online service for young Musicians. Musicians put recordings of their music on the website and visitors to the site are encouraged to comment on the recordings. The comments are collected into an online profile on each recording, which all visitors to the website are able to view.

A positive profile can be of economic value as it increases the musician's chances of selling their music online to visitors to the website.

Some of the musicians put the music on the site under their own name, some under aliases, and some under the names of bands or groups. There is no other form of identification upon the website.

Is the information contained in the profiles the personal information of the musician if the musician

a) Is a musician putting music on the site under their own name,
b) Is a musician putting music on the site under an alias,
c) Is part of a group putting music on the site under a band or group name?

In each case, why?

Scenario 3

Seeing the success of 'MusicMaker' a record company sets up another website called 'CDSuccess'. The site is designed to assess the music market. It asks visitors to the site to answer questions such as, how often you are likely to purchase music, what your favourite styles of music are, and what you think of the groups or bands currently recording for 'CDSuccess'.

'CDSuccess' use these answers to inform decisions about which new groups or bands they should offer recording contracts to and which of those groups or bands they should continue to record with.

Can the information gathered by 'CDSuccess' be classed as the personal information of,

a) Groups or bands already recording with 'CDSuccess'
b) Unsigned bands who are hoping to record with 'CDSuccess'
c) The respondents to the online questions?

Scenario 4

Albert and Brenda own a small business together restoring antiques. Albert left the business taking all the records of the clients' names and addresses, leaving Brenda with no records of previous clients.

The legal agreement creating the business did not cover these records.

Brenda claims that the information constitutes her personal information because it affects her ability to conduct a business.

Do you consider the information to be Brenda's personal data?

Why?

Would your answer be different if Brenda claimed that, some of the clients had become friends and the information were her personal data because it affected her social life?

Why?

Scenario 5

George is young social science researcher who has questioned 200 students in his university about their reading habits.

He completed his survey on a paper questionnaire and included the name and address of each student on the questionnaire.

Would the information contained in the questionnaire be personal data if (please explain your answer in each of the different situations):

a) He keeps the paper questionnaires alphabetically according to the students' names

b) He keeps the paper questionnaires according to their favourite book

c) He keeps the paper questionnaires without any order in a pile in his office

d) He anonymised the data by removing the students' name?

PART 2 - The Significance of Context

How significant is context to the identification of personal data?

This table sets out a number of relatively stable contexts, and asks you to identify whether various pieces of information are more or less likely to be personal information *within that context.*

The first row illustrates how you might indicate the relative likelihood of a piece of information being considered personal data in a particular context. In our example G,H, and I are the most likely to be regarded as Personal Data. C and F are the least likely. The likelihood of the other pieces of information being considered personal data is somewhere between these two extremes.

	←Less Likely										**More Likely→**
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
CONTEXTS											
Example	*C,F*	*J*	*D*	*K*	*E*			*A,B*			*G,H,I*
Doctor's surgery											
School											
Small Business											
Local Authority											
Police Force											
Charity											
Medical Research											
Supermarket											
Fan Club											
Sports Club											
Book Club											
Internet Chatroom											
Car Insurance Company											

INFORMATION

A Name
B National Insurance Number
C CCTV Image
D Fingerprint
E DNA/mouth swab

F Internet chatroom alias
G Vehicle registration number
H Criminal Record
I Age
J Postcode
K Parent's medical records

SECTION C

The following statements are designed to test the importance of relative elements of the definition of 'personal data'. Having considered them, please respond to them below with a view to sharing how you understand the concept of 'personal data'.

1. It is impossible to create a list of what is or is not 'personal data' as the concept is entirely dependent upon the context in which the information is placed.

2.
a) Information can only be personal data if it can identify an individual.

b) Information can only be personal data if it does not identify an individual but can affect an individual in a different way.

c) Information can only be personal data if it both identifies and affects an individual.

3.
a) The effect of processing information about an individual can relate to his or her fundamental rights to private and family life in many ways. The definition of personal data should only reflect protection against significant harm to an individual.
or
b) Information that produces any effect upon an individual must be defined as 'personal data' and it is then the rest of the law that determines its protection.

4. Individual identity goes far beyond identification but is protected within the definition of personal data.

If there is anything you would like to add about any of the questions raised in this Questionnaire, please do so.