Compliance Check Project

STUDY OF COMPLIANCE WITH THE DATA PROTECTION ACT 1998 BY UK BASED WEBSITES

Final Report

May 2002

UMIST Contact:

Professor L. A. Macaulay, Department of Computation
UMIST, Sackville Street, Manchester, M60 1QD
0161-200-3354, lindam@co.umist.ac.uk

Office of the Information Commissioner Contact:

Iain Bourne
Office of the Information Commissioner
Wycliffe House, Water Lane, Wilmslow, Cheshire SK9 5AF
ibourne@dpexecutive.demon.co.uk

Acknowledgements

The Compliance-Check Team:

Linda A Macaulay, John Fowler, Kathy Keeling, Mauricette Scheurer, Alan Hartley-Smith, Joanne Doherty, Debbie Keeling, Rachel Davies, Daniel Macaulay, Ismail Adeniran, Abuzzer Tahir, Steven Moschidis, Tony Walker and the five members of the recruitment team.

Contributors:

Special thanks to all those who participated in the interviews, whether on the telephone or during organisational visits. This study would not have been possible without them. The contribution of time and information from small or large organisations despite the many competing demands on their time was often exceptional and always valuable.

The authors wish to thank **Jupiter MMXI** for kindly contributing the list of top UK domains from the **Digital Media Audience Ratings Report** and so supporting the conduct of this study. (Further information is available at www.jupitermmxi.com)

Office of the Information Commissioner Staff:

David Smith and Iain Bourne for providing expert advice on the Data Protection Act (1998).

CONTENTS

1. Introduction

1.0 Purpose of this Document

The purpose of this document is to present the key findings of a study of compliance with the Data Protection Act 1998 [The Act] by UK based websites. The Office of the Information Commissioner commissioned the University of Manchester Institute of Science and Technology (UMIST) to conduct the study. The study was undertaken in the eight-week period between mid January and mid March 2002 and is referred to as the Compliance Check project.

1.1 Aims of the Study

The aims of the study were:

- To assess the degree to which the operation of UK websites is in compliance with the Data Protection Act 1998
- To identify particular areas where there is a failure to comply with the Act in order that the Commissioner can target future efforts to secure compliance and therefore improve standards for on-line information processing;
- To generate awareness in both data controllers and data subjects through:
  - publicity arising from the results of the study;
  - contact with those data controllers in the study sample;
- To provide a basis for possible future enforcement action.

1.2 Conduct of the Study

The majority of previous Internet privacy policy studies are based entirely on features visible on websites, e.g., the privacy policy and its content.

- However, there are the limitations in such evaluations and in the invitation to tender. The Information Commissioner advised that the true range and nature of compliance cannot be determined unless a deeper investigation is undertaken.
- It was, therefore, important to conduct a study that examined not only the website but also the data handling practices beyond the information available on the website, e.g., how data security is maintained and retention of information.
- It was recognised that to achieve this within the timescale available for the study might restrict the number of sites that could be evaluated. This was agreed at 170 sites within the targeted website categories.

The range of information needed to assess compliance beyond that available on the website meant that both quantitative and qualitative research was necessary. Thus, for the collection of the main body of data, **two stages were applied:**

- **1. Independent analyst assessment of websites** including assessment of privacy information for intelligibility on readability indices;
- **2. In-depth interviews**
  - telephone interviews with site operators to ascertain compliance with data handling, security, etc.;
  - visits and face-to-face in-depth interviews.

Further information was also sought,

as a follow-up on the use of data given to website operators:

Information was posted to the websites selected for the in-depth survey using a unique identifier compiled for the purposes of this study. Website assessors were instructed that, where possible, they would request no direct marketing communication.

Post website visit communications to the unique identifier address, both terrestrial and e-mail, was collected;

inclusion of technology service providers;

For many companies, either an ISP (internet service provider) or ASP (applications service provider) is used to host their site(s). Of course, the company can and must rigorously check the content for compliance, but it can do little about the data integrity and security aspects directly - they are dependent on the provider's hardware and software configuration. So to the extent that these bodies comply to the provisions of the Data Protection Act 1998, this may provide at least some *de facto* compliance with at least some parts of the Act for those sites that are wholly managed by such bodies. Thus, the state of compliance of ASPs and ISPs is a matter of importance for in-depth examinations and interviews. Further, the advice offered by website designers is also an important factor in relation to the likely compliance of many websites.

Overall, the study included making url visits in excess of 3000; many of these were not active sites despite listings as such. The study team have spoken to at least 900 companies and organisations, 500 have been e-mailed with a letter from UMIST explaining the study. 180 interviews were conducted. Note that the main body of the text refers to 170 sites, the technical service providers are the subject of a separate report.

1.3 Samples

The objective of the study was to scope the potential problem of website compliance with data protection legislation, and therefore

- was concentrated on types of site where consumers would be likely to be asked to divulge personal information.

- to achieve a rounded representation of the state of compliance with regard to specific areas, the Commissioner wished websites from the following target populations to be assessed:

1. Mail order or similar that collect credit or debit card details;
2. Companies offering financial services;
3. Government departments or agencies;
4. Companies or agencies offering health services or products;
5. Companies seeking to collect other sensitive personal data;
6. Employment or recruitment agencies;
7. Companies offering travel services;
8. Local authorities;
9. Companies whose website is aimed primarily at children.

A further target population of technology service providers was added.

1.4 Sample Sources

Groups 1-8

- As a foundation for the study to ensure as far as possible a comprehensive and random coverage of websites within each category, the Dun & Bradstreet Corporation's ("D&B") Electronic Commerce Registry database served as the starting point for creating the sampling frames for groups 1-8. These represented large, medium and small businesses based on annual sales and number of employees.
    - However, these lists may not contain enough data about websites run by unregistered traders, and so may present some bias.
- Therefore, a further sample of sites was compiled from websites offering lists of 'small' websites, e.g., portals such as home user internet access providers, shopping malls, on-line business directories. The target for interview of these sites was 30.

Most popular UK sites

- Additionally, since the majority of businesses in the UK are classed as Small to Medium Sized enterprises, a small sample taken from the Dun & Bradstreet list would reflect this and may not adequately cover the sites that large numbers of users visit.
- One way to overcome this, used in previous studies, is to also reflect the websites that large numbers of users visit.
    - Therefore, a random sample of 20 of the most visited UK websites was drawn. For this we are indebted to Jupiter MMXI for a list of the top URL's based on unduplicated traffic by UK consumers surfing the Web from home during November 2001. Again, substitution was by next on list.
    - The audiences of the sites in the sample ranged from just under 43% to 1% reach of the UK online population.
    - The list covered a wide range of heavily used sites, including portal sites, email and internet access providers, news sites, weather sites, travel services, magazines, television and radio stations, online directories, and games and leisure sites.

Group 9 Sites aimed at children

The sampling frame for children's sites was compiled from several sources:

- directories of sites for children, e.g., uk-click.co.uk; watsontheweb.co.uk.
- review of 12 magazines aimed at children and teenagers across the age range 5-16.
- search engine search.

The final list included sites built around fanzines, games sites, those giving health advice, product promotion, high street magazines, publishing houses, record companies, cosmetics, mobile telephone, charities and UN agencies. Table 1.1 presents the final sample broken down by website sector and size.

Table 1.1 Sample broken down by website sector and size

			SIZE			TOTAL
			Small	Medium	Large	
Website sector	Other services	Count	4		8	12
		%	33.3		66.7	
	Recruitment	Count	6	3	2	11
		%	54.5	27.3	18.2	
	Local authority	Count	3	7	4	14
		%	21.4	50.0	28.6	
	Financial & Insurance	Count	4	5	6	15
		%	26.7	33.3	40.0	
	Government agency	Count			10	10
		%			100.0	
	Travel related	Count	6	4	1	11
		%	54.5	36.4	9.1	
	Health related	Count	7	4	2	13
		%	53.8	30.8	15.4	
	Other sensitive data	Count	8	4	2	14
		%	57.1	28.6	14.3	
	Sites aimed at children	Count	13	1	3	17
		%	76.5	5.9	17.6	
	Retail related sites	Count	41	4	8	53
		%	77.4	7.5	15.1	
	Technical service					10
Total		Count	92	32	46	180
		%	54.1	18.8	27.1	100.0

1.5 Report Overview

This draft report is organised into four sections. Section two provides some general statements about the level of compliance of UK websites and includes examples of good practice and bad practice. Section three provides a summary of key issues covering explanation, privacy statements, direct marketing, the amount and type of information requested, data retention, data security and sites aimed at children. Section four presents a brief evaluation of the project against the aims and raises a number of issues concerning communication of the Act to smaller organisations.

1.6 Overview of Appendices

There are almost fifty pages of appendices associated with this report. These are organised according to key Data Protection principles (Security, Access, Retention, Notice, Choice and Adequacy) and contain more detailed reports and tables. In addition there are more detailed reports about sites aimed at children; about the understanding of personal-sensitive data and about sites that collect credit card data. There is also an analysis of the results by category of website and a comparison of the website assessment of those that agreed to be interviewed (i.e. 180) and those that did not (i.e.142).

Note that the appendices are provided in a separate document.

2. Level of Compliance

This section presents some findings related to the general level of compliance of UK websites, and highlights a number of examples of good and bad practice.

2.1 General

There is good general awareness of the Data Protection Act across both large and small companies. Most view customer confidentiality is important and 'good for business'. The level of compliance is variable depending on the size of the company and the extent to which the business is regulated by an outside body.

Larger and/or companies within regulated industry sectors exhibit a high level of compliance. Regulated companies are those such as insurance companies whose business is regulated by the Financial Services Authority. Banking and insurance company websites exhibit some of the best examples of compliance. Most of the larger companies have 'Compliance Units' who are responsible for setting up company internal processes to ensure compliance. The best examples of Compliance Units set controls for web developers and for web content.

Headline:	Large companies and/or companies within regulated industry sectors have websites that exhibit a high level of compliance. Small companies and/or companies in unregulated sectors have websites that exhibit a low level of compliance.

Smaller and/or companies operating within unregulated industry sectors exhibit a low level of compliance. Those who were compliant tended to be so more by accident that by design. There is a general level of awareness of privacy issues but only in as much as they might impact the interests of the business. There were some exceptions but even the best examples were not 100% compliant, the key areas for concern being those of Data Retention and Data Security.

Headline:	Key areas for concern are Data Retention and Data Security

Small companies who have extended their existing business onto the web are of particular concern. Data is typically backed up regularly in case of data loss, however, the data is then not physically secured and often there is no policy for the subsequent destruction of the data. There was generally a low level of company internal security. In addition there was a low level of system security with encryption rarely mentioned or used, though most companies did have an overall password for system entry. A number of companies in this category have 'Terms and Conditions' on the website that directly contradict the Privacy Statement. Most small companies do have good intentions and are often not aware of the danger of unintentional misuse of data or of the potential for external attack.

Headline:	Small companies are often not aware of the danger of unintentional misuse of data

Many small companies assume they are protected through their Internet Service Provider (ISP) when they are not. Many small companies were unsure of who is responsible for compliance if their data is held by the ISP. Most ISPs do not see it as their responsibility to ensure that their client's data is compliant, to give advice on Data Protection or on the inclusion of a Privacy Statement on a site they are hosting. However many ISPs do tend to give help if asked. There are examples of good practice among ISPs.

Headline:	Many small companies wrongly assume they are protected through their ISP

It appears that many organisations do not fully understand what is meant by 'data collection'. They assume that if they haven't explicitly asked for the data then they haven't 'collected' it. Particularly worrying in this respect is the collection of free form data (i.e. where the user can enter whatever information they choose) as used in emails, chat rooms and discussion groups. Self-help health and advice sites provide examples of free form personal data being collected and stored.

Headline:	The scope of 'data collection' is not well understood, especially with respect to free form data

More specific points relating to compliance are reported in section three. The next two sections present some examples of good practice and bad practice that the interviewers encountered.

2.2 Examples of good Practice

Business as Usual

Many small companies with existing manual processes do not retain data. They collect a customer order and credit card details on-line, print these out and then process the order in the normal (manual) way. Many are not aware that the data could be used for other purposes and would not want to use it in any case.

Internet Service Provider

One ISP who provides a shopping mall framework for retail shops to set up their own site provides an example of good practice. The mall site has a Privacy Statement for the general public with their own contact point and address so that the public can contact them if there are any problems or concerns. In addition the ISP actively encourages each retail site to have its own Privacy Statement and to display contact details.

Large Regulated Business

A large company who sells a range of products including insurance has six staff in their Data Protection Compliance Unit and has specific controls covering web builders. There are procedures in place to control all changes to the website. One particularly good policy they have enforced throughout the site is that of positioning a Privacy Policy button alongside the Data Submission button at the end of each data collection page. This ensures access to the Privacy Policy wherever it is needed rather than just in one place on the site.

Communicating with Customers

The Privacy Statement of one bank achieved an excellent Reading Ease score of 62%.[1] The bank promotes the use of a reference guide for plain English to be consulted for all written communication with customers.

Children's Sites

Some sites take children's privacy and safety very seriously. For example:

- Site A enables children to create their own web page but requests parental authorisation in writing (not by email).
- Site B gives very specific instructions to children on how to protect their privacy (which could be used as a model).

[1] The Flesch Reading Ease score rates text on a 100-point scale; the higher the score, the easier it is to understand the document. For most standard documents, the aim is to score between 60 to 70. The Privacy Statement for sites surveyed scored an average of around 45.

- Site C requires a written enquiry with proof of identity before it releases the details of what information it holds on children, although it is not really a UK site, more a US site with a UK front-end.
- Site D aimed at teenagers has a "banner" saying "Hi, I'm Julie, I'm 14, I'm into Kylie" and a text saying – roll here to see what Julie looks like – showing the picture of a middle aged man with the caption "people on the web are not always who they say they are!" which makes an important point in a striking way.

2.3 Examples of poor Practice

Health Advice and Counselling

Of concern for examples of poor practice, although doubtless unintentional, are the types of site often run by two or three people from home, offering a simple Privacy Statement if at all. While the people who operate such sites no doubt keep the information confidential and trust each other, unfortunately they frequently display a basic ignorance of PCs, data handling and data management. For example, on one site, the database was created by a friend, they "were not sure if they had a contract with the ISP but he is a friend of the family and they trust him". Data security is often thought of only in terms of physical security with confidential data and information passed using attachments with no encryption and no encryption of stored data, e.g., one website operator, when asked about data security, said they 'lived in an apartment block with a security man in reception'. A practice to be discouraged is the display of words such as "Registered under the Data ProtectionAct" on the website as a symbol implying that customers can have confidence in the site.

Self Help Groups (not Registered under the Act)

Self help groups thought they were not subject to legislation because they are not businesses and not registered charities but none the less process personal data. One such group runs a site for people with a particular type of illness to share their experiences. A bulletin board is used for individuals to put up whatever they like about their pain, medical condition etc. People (including children) are also encouraged to upload photographs of themselves. Two people vet the site (but this is limited). Clearly, there could be cause for concern. Several of these groups use discussion group software which is operated from abroad, usually in the US. The data is kept on a server in some unknown location.

Large distributed organisations

Large distributed organisations, such as universities and conglomerates, can show a lack of control of compliance. For example, university sites frequently contain a large number of registration forms, some ask about race, religion, health etc. It is all too often unclear who is responsible for the content of the website, e.g., one web master said he mounts the pages but doesn't vet their content. The Data Protection Officer of the university said he circulates Head of Departments to raise their awareness of the act and indicated that all the data being collected is collected on written paper forms anyway. A large conglomerate and a group of government agencies each had a single Data Protection Officer responsible for the entire member groups, a difficult task. Once data is collected in the constituent parts of distributed organisations it is difficult to tell how secure it is or who is in overall control.

Children's Sites

One site has online discussion groups where the material (some of it containing sexual content unsuitable for children) is visible to all visitors, although registration is required to take part. A couple of sites which appear to be offshoots of a phone company collect information on children (and adults) in return for free ring tones etc. for their mobile phone. At least one of them also requires the use of a premium rate telephone call. However the fact that they are an offshoot of the phone company is not made clear.

The next section presents further findings from the study.

3. Key Issues

This section highlights a number of key issues arising from the study and is organised as follows:

- Explanation given to individuals and the privacy statement
- Explanation related to Direct Marketing
- The amount and type of information collected
- Retention of information
- Security of information
- Information about children

3.1 Explanation to Individuals/The Site's Privacy Statement

This section highlights significant findings related to the experience of individuals using a site and to the site's privacy statement. Most of the findings from this section are the result of visits to sites by actual users and some are the result of both user visits and interviews with site developers.

The identity of the person or organisation operating the website is made clear in most cases. A small proportion provided only an email address and no postal address, while two sites provided only a postal address.

Headline:	75% of sites surveyed do provide contact information

Half of the sites surveyed do carry either a privacy policy or a fair collection notice. Information practice statements were posted on only a quarter of sites. Taking into account both types of statement 42% of sites did not post any form of privacy information. Small business sites are much less likely to carry privacy information than those of larger businesses. Of those sites with privacy statements most are accessible from the home page.

Headline:	42% of sites do not post any form of privacy statement

Privacy notices and explanations may be written in such a way as to be misleading or unclear to the ordinary visitor who may not be familiar with the exact terms of the law. Alternatively, the position on

the website may be such that it is effectively unavailable, whether intentionally or unintentionally. Only about 5% of privacy statements reached the recommended level for intelligibility to the average reader when assessed using the Flesch Reading Ease[2] score. Financial and Insurance sites faired worse while children's sites, travel and retail sites scored better.

Headline:	Privacy Statements are unlikely to be intelligible to the average internet user

All sites surveyed collect at least one kind of personal, demographic or sensitive information, yet less than half give an explanation as to why this is being gathered. Only 24 out of the 170 surveyed collect sensitive data, of these a little over half provide clear and unambiguous reasons for so doing.

Almost half of the organisations surveyed place a cookie[3] on the user's computer. In almost a quarter of these cases a third party place a cookie on the users computer.

Headline:	Almost half of all sites placed cookies on the user's computer

The interviews with site developers found a noticeable lack of awareness about web bugs[4], only 8 respondents stated that a web bug was placed anywhere on their site.

Headline:	Few sites use web bugs, those that do, use them for advertising or measuring site traffic but in one extreme instance investigators found 22 third party web bugs active on one page!

Although 43% of those with Privacy Statements tell users how to gain access to data held about them, only around a third of notices mention it is a user right to see data. Of the total 170, this is around 18%.

Headline:	Only 18% of sites surveyed tell users how to gain access to data held about them

Overwhelmingly, when this information is given it is in the Privacy Statement (90%) – in only 2 cases was this information given at the point of data collection. There is no statistical indication that giving information on data access rights is related to size, but smaller operators show slightly lower numbers doing so than larger concerns. Only just over a third of those mentioning user rights also mention a cost and the amount.

[2] See footnote 1 above
[3] See Definitions
[4] See Definitions

36% of those with a Privacy Statement indicate how inaccuracies in data will be handled. Of the total 170, only 27% of sites give any information on how to question or complain about privacy or misuse of data.

Headline:	Only 27% of sites give any information on how to question or complain about privacy or misuse of data

Only 39% have procedures for recording what personal data is collected and kept so the actual ability to perform a full data access request may be in question. Of course, for many small companies this does not pose as much of a problem since they have all their information in one place.

Headline:	Less than 40% of sites have procedures for recording what personal data is collected

The table below shows how personal data is located if there is a request for access.

	COUNT	% OF 170
Definite procedure/explanation		
Central database	66	39
Held individually on server	11	6.5
Access via website	3	2
Info not kept beyond purpose	10	6
Paper file	8	5
Electronic and paper files	2	1
		59.5
No definite procedure/explanation		
Responsibility of other department/s	17	10
Response indicating misunderstanding of concept	27	16
No formal procedures	17	10
		36[5]

[5] Missing data accounts for 4.5%

Nearly 60% of respondents could give an explanation of how they would locate all personal data. Nonetheless, 36% said they had no formal procedure, misunderstood the concept or passed the responsibility to another department.

Headline:	Only 60% of respondents could give an explanation of how they would locate all personal data

E-mail is by far the most common method of allowing people to ask for inaccuracies in data to be rectified. Up to 30% of operators allow users to have direct access to their own data on the website and change it as necessary. Operator 'size' does not appear to be a factor and multiple options are often given.

Only 27% of respondents say that they go back to check data is still valid. For this purpose, contact by letter is as likely to be used as e-mail. Multiple methods are used.

3.2 Direct Marketing

Approximately 35% of all the sites surveyed state that information collected from users may be used for direct marketing or other purposes. Of those sites with privacy policies 61% stated this. This suggests that sites with privacy policies are more likely to inform users that information they divulge may be used for direct marketing, although it would appear that a small number of sites without privacy policies are informing users of their intention to utilise information collected from them for direct marketing.

Headline:	Sites with privacy policies are more likely to inform users that their information may be used for direct marketing

Only 35% of all sites surveyed give users a choice about whether they want to be contacted by the organisation for marketing or other purposes. Of these, 50% adopt an opt-out policy and 28% adopt an opt-in policy.

While visiting websites for assessment, the Compliance Check site assessors completed personal details using the name PRIS UMIST (PRIvacy Study at UMIST) and the UMIST terrestrial address. A hotmail account was set up for PRIS UMIST where email addresses were required. Where possible, assessors indicated that they did NOT wish to be contacted for direct marketing activities.

Headline:	Some sites sent marketing information when requested not to do so

Within two weeks of completing the study several companies/organisations had sent marketing emails when requested not to do so; one had sent multiple emails.

Headline:	Most unwanted emails offered loans or home working schemes, but ironically, one offered software to help protect online privacy!

3.3 The amount and type of Information Requested

The following tables show the percentage of sites that gather each type of data and whether that data is required (i.e. which must be entered in order to proceed) or optional:

Personal Data	**Optional %**	**Required %**
Name	14.7	82.4
e-mail address	18.8	79.4
Postal address	11.2	74.7
Telephone number	14.7	58.8
Driving licence details	1.2	2.4
Fax number	15.9	5.3
Credit card number	10.6	31.8
National insurance number	1.8	1.8
Mobile number	14.1	.6
Other	9.4	3.5

About 80% of sites gather no more than 5 pieces of personal data, the average over 170 sites is 4.5.

Demographic data	Optional %	Required %
Age/date of birth	5.3	19.4
Family members	4.1	5.9
Occupation	7.6	12.9
Education	1.8	4.7
Income	2.4	3.5
Expenditure	1.2	2.4
Gender	4.7	12.9
Country of residence	9.4	28.2
Other	10	6.5

Generally, very little demographic data is collected, 42% collect none at all and only 20% collect more than 3 pieces of information, the average is less than 2.

Sensitive personal data	Optional %	Required %
Physical or mental health	5.3	5.3
Racial or ethnic origin	1.8	2.4
Political opinions	1.2	
Sexual life	2.9	
Religious beliefs		
Criminal convictions	1.2	1.2
Sensitive organisational membership		

Exceedingly few sites collect sensitive information, however from the interviews it was clear that when, for example, health information was collected there could be some depth in that.

There is substantial misunderstanding and ignorance of what constitutes personal data and what constitutes sensitive personal data. Only 60% of interviewees were familiar with the distinction between personal data and sensitive personal data. Of that 60%, 60% gave an accurate description of personal data (36% of 170) and only 37% gave an accurate description of sensitive data (22% of 170).

Headline:	There is substantial misunderstanding and ignorance of what constitutes personal data and sensitive personal data

An additional consideration here is that the user's presence on some sites could be considered as indicating something about them. The fact of a visit to a particular health information site, or particular pages within general health information sites could lead to inferences about the user. This is even more so for those visitors who register with special interest sites, e.g., if a person registers to get news about low-salt diets for regulating blood pressure, there is an inference that there is a need for that information.

Headline:	Exceedingly few sites collect sensitive information, however, the user's presence on some sites could be considered as inferring something about them

The amount of information collected does not appear to be a function of size but of industry sector. Financial and Insurance and Recruitment sites tend to collect the most information, a not unexpected result, which is probably not out of keeping with the conduct of their business and delivery of services. Most retail sites do not collect excessive information, but there are a small number - a feature found in many of the sectors - that are collecting up to 7 (or more) pieces of information.

Headline:	Excessive information gathering does not seem to be widespread practice

Excessive information gathering does not seem to be widespread practice but there are indications that there are some sites, spread across size and sector, that gather data that may be excessive. Coupled with web bug activity these sites would bear careful scrutiny, especially as most often third party web bug activity is non-UK based.

Headline:	Some sites are in need of careful scrutiny as they gather excess data and use web bugs

3.4 Retention of Information

Policies on retention of data are a reflection of the lack of procedures for assessing the amount and type of personal data collected for a particular purpose (see above). Where there is a policy in place, it tends to be based on pre-existing statutory or sector requirements. 36% of those that answered this question indicated that they have statutory or sector requirements that set out requirements for the retention of data.

However, 24% of respondents have no formal retention policy in place, 9% of these keep data indefinitely. A further 16% gave answers that were inappropriate to the question, indicating a mis-understanding of the concept.

Headline:	The intent of the Data Protection Act in respect of retention is not fully understood

Furthermore, other answers also indicate that the intent of the Act in respect of retention is not fully understood by:

- the 13.5% whose retention period was based on customer action only, e.g., data is only removed when customers tell organisations they no longer wish to deal with them;
- the 22% where data removal is based on technical or operational considerations rather than on the purpose of the data.

Therefore, the actual figure without a formal retention policy based on the intent of the Act must surely be higher than the 24% who admit to that and, on the whole, is an area for concern.

Headline:	At least a quarter of sites do not have a retention policy in place

When asked about their technical ability to set retention periods less than 30% of interviewees thought that their system included the facility to set specific retention periods for certain categories of data. Less that 40% thought that their system included the facility to flag records for review/deletion. Less than 25% thought their system capable of setting different retention periods based on purposes and data items.

Headline:	The technical ability to set retention periods is in question

3.5 Security of Information

Only 45% of sites interviewed have a data security policy. More large companies have a data security policy than small companies (60% compared with 37%). Less than half of all companies use some form of secure electronic link.

Headline:	Only 37% of small companies have any kind of data security policy

The most popular database technology used to store data were SQL Server (21%), Microsoft Access (25%) and Oracle (10%).

83% of companies claimed to make back-ups of the database, but only 55% of these have a back-up held off-site.

Headline:	Less than half of companies hold a back-up copy of data off-site

Even though many have no formal policies regarding security, most respondents felt able to explain how unauthorised people are prevented from accessing the system. Passwords are by far the most

common method of controlling access, they are 4 times more popular than the nearest rivals (swipe cards).

Protection against hacking and use of firewalls is more likely in larger organisations and less likely in smaller organisations. There is no statistical evidence to suggest that security is any more likely or unlikely in any particular sector, except for a suggestion that financial institutions and government agencies may be more likely to encrypt stored data and actively protect against hacking.

For 22% of companies there is the added complication that copies of data are held separately at different locations by different departments. Not surprisingly, this is more common in larger companies where the practice rises to 32%. However, central control of access is most usually retained, only 25 companies allowed departmental control.

Headline:	Working at home or off-site could be a cause for concern

Another potential lapse in security is if staff are authorised to take equipment or software for external use, e.g., to work from home or perhaps on a home visit as an insurance agent, mortgage or investment advisor. 36% of those interviewed said this was allowed by their company/organisation. 70% of these claim to give special training on safeguarding personal data.

3.6 Information about Children

30 children's sites were visited but only 17 of these agreed to be interviewed. This part of the report summarises the findings from the 30 site visits. The detailed analysis from the 17 in-depth interviews is not yet complete and is not included in this draft of the report.

The thirty were selected because they were either mentioned in magazines aimed at children and teenagers, or they came up at the top of a search engine search. The 30 Children's sites visited were mainly sites designed by professionals. 12 sites were part of a group of companies or had an associated site not for children. Many were aimed at children, but by no means all: 5 of the sites were "about" children, but not aimed at them.

Some of the sites visited are built around pop music and pop stars, others give health advice, others are built around a character, others promote their product. Several sites are not for children, but about them – aimed at parents, teachers, etc. Some sites have pages for children on a site not primarily aimed at them. Some sites seem to exist mainly to get information about people who own their products.

The vast majority (86%) of UK sites do not give any particular cause for concern:

- Most sites are sponsored by, or an offshoot of, "reputable" companies or charities (e.g. high street magazine, publishing house, record company, cosmetics, mobile telephone, UN agency).
- Most of the sites do not appear to mislead their visitors either as to who they are or as to what the site is about.
- About 60% of the sites have some contact details, although more often than not these were to the "parent" company or organisation. Some sites had no contact details at all and others were really difficult to find.

Headline:	Only 60% of children's sites provide contact details

- About 60% of the sites have a Privacy Policy. However, only about half of those display it prominently, whilst some of the others have it buried so deep into the site that it might as well not be there at all. (In one instance the Privacy Statement could only be found by using the site's search facility.)

Headline:	Only 60% of children's sites have a Privacy Statement

- Several sites have an excellent Privacy Policy and make a real effort to make it understandable to children. Several sites officially discourage children under 14 from using the site or subscribing to online services, and encourage children under 18 to seek parental approval.
- Some sites have a "catch all" clause "If you use our site you are deemed to agree with our terms and conditions" (which may include a statement to the effect that they share Personal Information with others, or reserve the right to do so).

A more detailed report is provided in the Appendix.

4. Summary

This section presents an evaluation of the project against the aims as described in section 1.1.

4.1 Aim 1: Degree of Compliance

To assess the degree to which the operation of UK websites is in compliance with the Data Protection Act 1998.

The results of this study should be interpreted with reference to the following:

- Any study of websites faces difficulties in defining the exact population since the true size of the website population is hard to estimate and extremely fluid, websites are added and withdrawn every day. The recruitment team often came across sites that were no longer active but still appeared on lists.
- Given the time-frame and the need for in-depth investigations, the sample size in this study was restricted.
- The focus of this study was on categories of UK based websites where users were likely to give personal information. As a consequence, the results of this study cannot be generalised beyond the population from which the sample was drawn, meaning no conclusions can be drawn using this study about online privacy practices for categories of UK websites not included or the rest of the World Wide Web.

Therefore, this study is best represented as a scoping exercise to demonstrate the likely size of the problem of non-compliance in UK websites most likely to gather personal information and to indicate particular areas of concern, perhaps for further investigation.

Notwithstanding, we believe that the results are a fair indication of the state of compliance of UK websites within the categories studied:

- the inclusion of samples from a number of different sources and the random selection within those sources provided depth and breadth to represent the variability of the websites in size, available resources and number of users
- the interview methodology provided a deep and rich picture of the issues
    - often with company personnel whose wealth of experience and knowledge provided much detail on issues beyond a discussion about their individual website
- although the interview methodology restricted the sample size, it is unlikely that certain information would have been available any other way - particularly from the larger organisations
    - information beyond the website would be difficult to gather, if not unworkable, by other means
    - website operators approached for interview could (and did) refuse, but it is commonly held that a personal approach, in this case by direct telephone call, results in higher response rates than approaches relying on replies to e-mails, letter or blanket questionnaires,

especially amongst those who are regularly solicited for information. The recruitment team was frequently told that the person recruited did not normally give interviews and, indeed, requested much further information. The permission for interview was often the result of a series of telephone calls.

4.2 Aim 2: Particular Areas

To identify particular areas where there is a failure to comply with the Act in order that the Commissioner can target future efforts to secure compliance and therefore improve standards for on-line information processing.

Particular areas of concern are highlighted in sections two and three of this report.

4.3 Aim 3: Awareness

To generate awareness in both data controllers and data subjects through: publicity arising from the results of the study; contact with those data controllers in the study sample.

The project has generated awareness in a number of ways, including:

- The study team have spoken to at least 900 companies and organisations.
- 500 have been e-mailed with letter from UMIST explaining the study.
- 180 interviews were conducted.
- sending a copy of 'FARSTARS' Guidelines for System Designers to respondents.
- sending a copy of Compliance Check Technical Questionnaire for use as checklist for respondents' own site.
- informing respondents of the web address ("url") of the Office of the Information Commissioner.

4.4 Aim 4: Future Action

To provide a basis for possible future enforcement action

A number of issues were raised by respondents during interviews with the Compliance Check team. These may assist the Commissioner when considering enforcement action. They mainly relate to smaller organisations.

Advice from the Office of the Information Commissioner

Larger companies can afford to communicate directly with the Office of the Information Commissioner, smaller companies cannot. Larger companies expressed satisfaction with the OIC although some commented that 'it was taking longer these days'. There is a major legal industry building up around the Act. (Interviews with larger companies took place in Solicitors Offices.)

Smaller companies have a broad understanding of the act but do not know whether they comply. They find:

- The Act difficult to understand
- They cannot afford legal advice

- The Office of the Information Commissioner website does not give an interpretation of the Act that is appropriate to their needs

The need for education

Smaller companies have expressed a need for:

- Reasonably priced and accessible education;
- 'Frameworks for Compliance', based for example on typical scenarios such as:
  - if you are a small shop who collects credit card data solely for your own purposes, then you should comply in the following way……
  - if you are travel agent (or similar) who passes information to others e.g. to an airline or hotel then you should comply in the following way…..

The need for better communication channels

- Many companies expressed concern that the Act might change in some way and that they would not be aware of the change. There is no obvious mechanism for communicating change to them.
- Many smaller companies felt that they had no voice with the Office of the Information Commissioner and that they need an organisation that can represent small business interests with the Office of the Information Commissioner.

The need for certification

- A number of smaller companies are using registration under the Act as a symbol of 'certification'.
- A number of children's sites claim to abide by the Data Protection Act, the BBB Code of Practice or COPPA legislation[6]. There is clearly a need for some form of 'certification' that a site is 'Data Protection Compliant'.

In conclusion, this report has highlighted some of the key findings of the Study of Compliance with the Data Protection Act by UK websites. It has presented details about the conduct and status of the study, reported on the general level of compliance and highlighted a number of key issues. Further details can be found in the appendices. These are provided in a separate document (Appendix).

[6] US Children's Online Privacy Protection Act – which applies to children under 13.

5. DEFINITIONS

Browser

Used to locate and display Web pages via a software application, e.g., Netscape Navigator and Microsoft Internet Explorer.

Cookie

A small text file, that a website can place on a computer's hard drive. This identifies a computer during the current and subsequent visits to a web site. Cookies have a variety of uses, many of which are important in the delivery of services over the web. They may make it possible to use an online 'shopping basket' (aka 'cart') to keep track of items you wish to purchase, personalise your visit, e.g., welcoming you by name; or maintain continuity from one page to another as you browse through an online catalogue or use on-line payment systems. The concern is that together with either web bugs and/or merging offline information, they could be used to build up user profiles without user knowledge or consent. See also session cookies, persistent cookies and web bugs.

Demographic information

Demographic information cannot be used to identify a person on its own: examples are age, gender, and income. In this form it is often use for market research, but combined with personal identifying information can be used to create personal profiles of web users.

Information Practice Statement

A single declaration that specifies one data privacy practice, e.g., 'We do not share your personal information with third parties'. These are often, but not exclusively, found at the point of data collection.

Persistent cookie

Once installed, it remains on the hard drive of your personal computer. Typically used to store information so that a website can identify you between visits.

Personal identifying information or data

This is information that can be used to identify a person, such as, name, e-mail address, driving license number, National Insurance number.

Privacy Policy Statement or Notice

A comprehensive account of the privacy practices relating to that web site that is located in one place on the web site. Usually reached by clicking on an icon or a hyperlink. May not always be termed Privacy Policy by a web site, e.g., other terms include: terms and conditions; customer relationship; important information.

Sensitive personal information or data

http://www.dataprotection.gov.uk/dpr/dpdoc.nsf

The Act defines categories of sensitive personal data as consisting of information as to the racial or ethnic origin of the data subject; political opinions; religious beliefs (or other beliefs of a similar nature); membership of a trade union; physical or mental health or condition; sexual life; any information about criminal offences or civil offences.

Session or Transient Cookie

Non-persistent cookies that are either completely removed from your computer when you close your web browser or within a fixed period of time. Used to enable shopping baskets and control advertising during one and subsequent visits within that period of time. Also may be used to automatically 'log-out' if web page has not been 'refreshed' within a set period of time.

Site classification by size

Included in the 170 web site interviewed was a broad sweep across large organisations to small businesses, charities and other not for profit organisations. For reporting contrasts between web sites, web site operators were grouped by 'size', the groups reflecting meaningful numbers to be able to report differences. Number of employees, on its own, is insufficient to classify a site by 'size' as it does not reflect the resources of finance and expertise available to a web site or the particular properties of web site use, such as site traffic. The web sites of many larger concerns are registered as separate companies. These separate companies may have few staff but large resources in terms of both finance and expertise. For the same reason and because some companies/organisations did not disclose, turnover was not used to categorise web sites. The study was given access to the Jupiter MMXI *Digital Media Audience Ratings Report* for November 2001, a list of top UK domains to 1% reach. 'Small sites' are those where the operator had up to 50 employees, no affiliation to a larger organisation and the site did not appear on the Jupiter MMXI list of top UK sites. Medium sized sites were those where the operator had approximately 51- 300 staff and the site did not appear on the Jupiter MMXI list of top UK sites. Large organisations are where the operator had more than 300 staff and/or affiliation to a large organisation and/or the site appeared on the list of top UK sites. The complexity of this assessment, meant, therefore, that sometimes the classification decision was a judgement appraisal, e.g., for the purposes of this study, all government web sites were classified as 'large'.

Web Bugs, also known as 'web beacons', '1by1 GIFs', 'invisible GIFs'

Web bugs are objects (images, iframes, etc.) imbedded on a web site. They allow a 3rd party web site to know the IP address and the page that was visited. Further, they can be linked to cookie activity. They are most commonly found in banner and other 'click-through' advertisements. Web bugs are often invisible because they are typically only 1-by-1 pixels in size. Not all very small GIF files are web bugs; they are also used for alignment within web page presentation. However, a web bug will often be loaded from a different web server than the rest of the page. Web bugs are used for a variety of purposes, e.g., an independent count of web site traffic; gathering web browser statistics; monitoring 'click through' traffic in response to particular advertisements; and tracking 'click throughs' between sites for determining commission payments. As they can be used to track the sites visited by users, the concern is that a web bug can be used along with cookie information to profile users. This happens without users' knowing; therefore, they cannot give consent.

For more information see:

http://www.cybersync.com/help/web-bugs/wbfaq.htm

http://www.eff.org/Privacy/Marketing/web_bug.html

http://www.privacyfoundation.org/resources/index.asp

http://www.intelytics.com/beacon.asp

Examples of websites connected to web bug activity

http://jc.visitorprofile.com

http://v0.extreme-dm.com

http://stky.brinkster.net/

http://www.qksrv.net - commission junction

https://www.netshopperuk.com - secure payment service

http://m.doubleclick.net

http://ad.uk.doubleclick.net

http://www.wundercounter.com

http://www.sitestatslive.com