



THE DIRECT MARKETING ASSOCIATION (UK) LTD

HAYMARKET HOUSE · 1 OXENDON STREET · LONDON SW1Y 4EE · T 020 7321 2525 · F 020 7321 0191*

29 WELLINGTON STREET · LEEDS LS1 1QQ · T 0113 244 7103 · F 0113 244 7224

41 COMLEY BANK · EDINBURGH EH4 1A · T 0131 315 4422 · F 0131 315 4433

E dma@dma.org.uk · W www.dma.org.uk

The Direct Marketing Association (UK) Limited is a company limited by Guarantee.
Registered in England No. 2667995. Registered office as above*.



DMA Best Practice Guidelines:
**FOR USE OF DATA IN
DIRECT MARKETING**

Acknowledgements

The DMA would like to thank the following practitioners who contributed to the development of these Best Practice Guidelines as members of a special task force:

David Coupe, Experian
Tony Masters, WWAV Rapp Collins
Ian Goodman, Printronic
Paul Turner, Acxiom

Thanks also go to David Reed, freelance journalist, for his input.

Contents

Why Best Practice?	Section A	1
Data Protection	Section B	3
Caring for Customer Data	Section C	5
Data Capture	Section D	6
Receipt and Transfer of Data	Section E	10
Name and Address Conversion and Standardisation	Section F	12
Name and Address Matching	Section G	15
Deduplication and Merge-Purge	Section H	18
Screening	Section I	20
Data Tagging and Enhancement	Section J	22
Sortation and Output	Section K	24
Appendix – Useful Addresses		25

Section A

Why Best Practice?

The use of data is key to most direct marketing activities. Whether the campaign is a simple list selection and mailing, or utilises complex databases and processes to arrive at a targeted audience, adoption of best practice in use of data has an equal importance.

The dynamics of today's marketplace means that data held about the individual begins to decay as soon as it is gathered.

For example:

- approximately 13% of the UK population move house each year (OPCS);
- approximately 11% of addresses are mailed incorrectly each year (DMIS);
- 45% of the UK population believe that a mis-spelt name or address is an indication of 'junk mail' (DMIS);
- 460,000 UK households are registered with the Mailing Preference Service (MPS).

In most direct mail campaigns the list (or file) of target prospects or customers may undergo a journey between a number of organisations involved in different parts of the production process e.g. list owner, broker, computer bureau, laser printer, mailing house, etc.

It is therefore vital that during all these processes, accuracy, integrity and security of the data is maintained to the highest standards (please refer to Section E: Receipt and Transfer of Data).

The contents of this document therefore, are designed to address the common issues that may occur and provide guidelines to avoid mistakes and advise on best practice for use of data in direct marketing.

Best practice means the standards which it is desirable for all those involved with data to achieve. Members of the DMA are already required to comply with the Association's Codes of Practice as a condition of membership. Best practice guidelines go beyond these levels by establishing benchmarks for members to aim at in the way they hold, handle and use data.

The benefits of best practice in use of data are quite clear:

- helps direct marketing become more cost effective, avoids waste for the advertiser and saves money;
- reduces potential annoyance to recipients through duplicated mailing, incorrect or mis-spelt names and addresses;
- helps advertisers target mail more effectively, enhancing the advertiser's image with his customers and prospects;
- well targeted and produced mail provides a more confident message to consumers about direct marketing and DMA members;
- best practice is an important part of industry self-regulation.

This guide also covers areas of data usage which are included within our Codes of Practice or by specific legislation. The Data Protection Act 1998 is the principal legal framework within which all personal data may be handled. These guidelines set out where responsibility lies between clients and bureaux for ensuring compliance with the Act at each stage of data usage.

The outcome of following the best practice guidelines should be better performing communications which build customer relationships and long term loyalty. These are objectives which the whole industry shares and which the Data Council and the DMA endorse.

Section B

Data Protection

Whether using customer or prospect files, in business to business or consumer markets, it is important to remember that commercial access to data is a privilege, not a right. Except for public domain information, every item of data used has been given freely and voluntarily by the data subject, with the expectation that it will be used fairly, appropriately and legally.

The Data Protection Act 1998 came into effect in March 2000 implementing the European Data Protection Directive. Companies which complied with the previous Act's eight principles should experience few problems in complying with the new legislation, although there are some important changes.

To maintain best practice in the use of data, marketers should also have in place facilities to ensure that data is as up-to-date as possible and that all suppressions are respected. A change of address file, either proprietary or commercial, should be used to validate addresses before they are mailed.

Notifications of changes to details, such as address, telephone number, job function, etc, should be recorded and added to the master file within a reasonable period of time. Requests not to be mailed, phoned or faxed (separate from Mailing, Telephone and Fax Preference Service registration) must be logged on 'Do not mail', 'Do not telephone' or 'Do not fax' lists and used as a screen before carrying out any communication.

The client is responsible for ensuring that:

- where processing of personal data is carried out by a bureau on behalf of a client, the client must in order to comply with the seventh principle of the Data Protection Act 1998:
 - a) choose a processor providing sufficient guarantees in respect of the technical and organisational security measures governing the processing to be carried out, and;
 - b) take reasonable steps to ensure compliance with those measures.
- where processing of personal data is carried out by a bureau on behalf of a client, the client is not to be regarded as complying with the seventh principle of the Data Protection Act 1998 unless:
 - a) the processing is carried out under contract:
 - i) which is made or evidenced in writing, and;
 - ii) under which the bureau is to act only on instructions from the data controller, and;
 - b) the contract requires the bureau to comply with obligations equivalent to those imposed on the client by the seventh principle.

- any in-house database and use of personal information is correctly registered;
- all data has been acquired and processed in accordance with the Data Protection Act 1998;
- any marketing communication using personal data complies with the British Code of Advertising and Sales Promotion (BCASP);
- facilities are in place for updating data, registering requests not to be mailed or phoned, or to permit access to data by its subject;
- third-party data rented from a commercial data owner in membership of the DMA is supported by a list owner warranty.

The supplier is responsible for ensuring that:

- the data owner is fully and correctly registered with the Data Protection Commissioner (all current notification details are now accessible online at <http://www.open.gov.uk/dpr/>);
- data is held and used in accordance with the Data Protection Act 1998 including notification of its business and security of the data;
- lists are screened against the most recent version of MPS, TPS or FPS;
- any communication using data which they own/manage complies with BCASP;
- requests to be removed from a list, to have details corrected, or to have access to the information held are properly dealt with;
- list owners hold a list owner warranty, where appropriate.

Section C

Caring for Customer Data

Information held on individuals who are customers of the advertiser's company is of particular value. One of the principal drivers of direct marketing has been an emphasis on retaining customers, through loyalty programmes, and cross-selling. It has been estimated that it costs one-fifth the amount to sell to an existing customer as it does to make a new sale.

For this reason, the DMA's Data Council recommends that particular care is taken when handling, processing and using files of customer data. During the course of a customer's lifetime, it is also likely that a greater wealth of data will be generated. That means more sensitive items, such as date of birth or financial status, may be held. To retain customer trust, this data must be processed to the highest possible standard.

The DMA recommends that the following points are given special consideration when treating customer data:

- particular care should be given to the data capture process, especially where the information being captured may be sensitive and for which special rules may be in place;
- the fourth principle of the Data Protection Act 1998 should underpin the management of customer data. This states that, "personal data shall be accurate and, where necessary, kept up to date." The data owner should be conscious of the time sensitivity of data and its inherent attrition rate;
- an appropriate updating cycle should be implemented to ensure that the data held is accurate and up to date. Immediately prior to a communications programme, extra emphasis should be placed on refreshing customer data;
- 'address' management procedures should be in place to maintain the accuracy of customer addresses. These should include automatic updating of the postal address using Royal Mail's Postcode Address File (PAF), using customer notification of change of address, or employing proprietary change of address or suppression files;
- regular screening against a deceased file should be undertaken to avoid causing distress;
- responsibility for managing customer data extends to the prompt removal of deceased records;
- DMA members are required to screen customer records against Mailing Preference Service (MPS). It is a statutory requirement to screen records against Telephone Preference Service (TPS) and Fax Preference Service (FPS). Customers registered with the Preference Services file have indicated they do not wish to receive such marketing communications. Companies must operate their own 'do not mail/telephone/fax' files, as required by the Telecommunications Act and the unfair processing in data protection legislation;
- prompt action should be taken to respond to a customer request for access to his or her personal data. Under data protection legislation, all information held has to be provided promptly and in any event within a maximum of 40 days, subject, if thought appropriate, to a fee with a current maximum of £10. Such requests should be dealt with as a customer service in a spirit of transparency and good faith.

Section D

Data Capture

To make use of data – whether name, address, telephone number or any other element – the information has to be captured at some point. Achieving the highest level of accuracy and consistency at this stage will reduce problems with data processing and use later, as well as improving results from any analysis or contact made using the data.

Three key sources exist from which data may be captured – print (mailed/faxed response), telephone or electronic media.

Coupon response

When designing a marketing communication which will generate a physical response in the form of a coupon, consideration should be given as to how data capture will be facilitated. Whether cut from a press ad, mailshot, catalogue, leaflet or other printed material, the coupon should prompt the consumer to provide data that is both sufficient and accurate.

To support good quality data capture, a coupon should:

- provide sufficient space for all the data elements required. The average UK name and address record is 48 keystrokes long, but it can involve up to nine separate lines of information. While use of data for postal purposes may only require the delivery point and postcode, consumers have preferred address elements which might include locality and even dependent locality. Business addresses can be even longer and business titles more varied still;
- prompt the respondent for key data elements. Separate lines or boxes for title, initial or name, address and postcode will improve the quality and accuracy of data compared to free form text boxes. A specific prompt for postcode will also improve the speed and accuracy of data capture, since automated processes can return the full address from this element alone;
- include a separate prompt for country if data is being gathered from multiple countries. This precludes the need for subsequent assignation of each response to a country of origin;
- allow variations in handwriting, ink colour, etc. Using blocks or 'tiger teeth' to denominate spacing for characters can be a useful way to improve legibility. This may also allow coupons to be data captured using faster optical character recognition (OCR) processes. Coupons should not be printed on strong colours (ie, reversed out of black) or over images as this will make completion and reading harder;
- be tested before signing off on a campaign by asking a friend or colleague to complete it.

Telephone response

An increasing volume of response is being generated via the telephone. This gives an ideal opportunity for data capture of both name and address as well as other information, subject to time and cost constraints.

To support best practice in data capture, telephone response mechanisms should:

- allow the respondent to provide information at his or her own speed where automated call handling/interactive voice response (ACH/IVR) is being used. The system should prompt for name and address elements, including a double check for key elements, such as postcode. A prompt to spell difficult words should be included to facilitate transcription unless these are covered by reference to PAF;
- include an initial request for the caller's postcode in live operator scripts. Computer software can return the correct postal address from this, allowing operators to validate it with the caller and add the house number and any preferred address elements. This will also reduce the duration of calls;
- avoid open-ended questions, as these extend call times. Data capture requests should be restricted to yes/no categories, banded information (such as age in ten year steps), or multi-choice prompted responses.

Fax Marketing

Fax marketing is increasingly being used as part of the marketing mix and this gives rise to a number of issues from a data capture point of view.

To support best practice in data capture, fax marketing mechanisms should:

- ensure that adequate arrangements are in place to receive and record details of recipients who do not wish to receive further fax mailings, whether such requests are made by telephone, fax or mail. Any such telephone requests should be dealt with sympathetically and sensitively by the issuing organisation;
- ensure that the fax number of such 'do not telephone' recipients are removed from the respective fax lists in a timely manner after notification and not called in future campaigns;
- ensure that any consumers, or businesses operating from a consumer premise, are made aware of the FPS contact number if they indicate a more general 'do not telephone' requirement beyond the actual fax marketing piece in question;
- ensure that any delivery report information following a fax mailing is diligently used in a timely manner to maintain up to date fax lists for future use.

Electronic media

Growing use of the Internet and email will allow highly personalised communications, based on detailed information about targets. Care is necessary at this stage, however, due to the distinctive culture of these media.

Online data capture mechanisms should:

- respect the prevailing 'netiquette'. Many users of the Internet object to commercial communications which they have not requested;
- use online address validation software to check the address details entered and confirm with the user any questionable details. Remember that an email address does not give any indication of where a user lives, or their age so prompt for country of origin and date of birth;
- make access to higher levels of an online site conditional on the provision of personal data, but do give some value to users who do not wish to leave their details;
- DMA members must comply with the online code of practice.

Using data capture bureau

Data capture involves two elements. The initial stage is the conversion of coupon or telephone responses into an electronic file via optical scanning, rekeying or transcription. This produces a list which is identical to the responses received, with information exactly as provided by respondents. In the second stage, this data may be validated, enhanced or corrected.

Choosing a data capture bureau requires decisions to be made about the level of accuracy required and the extent of further work required on a file. Best practice guidelines include:

- selecting a bureau with relevant experience and expertise. Data capture may be carried out from market research surveys, postal lifestyle questionnaires, coupon responses, etc. Each of these employs a different skills set (and possibly different technologies). Ensure the supplier has the appropriate knowledge;
- due to the commodity nature of first stage data capture, many suppliers are based off-shore. This can offer significant savings, even when allowing for freight costs. A potential challenge to this route may come from the new data protection legislation. To comply with the law, data will be allowed to be exported only to countries outside the European Economic Area which have an adequate level of data protection. Further advice will be issued by the DMA together with a model form of contract which may bring normal direct marketing transfers within the exemptions to the law's prohibition;
- selecting a bureau which offers the necessary facilities if data needs to be validated or enhanced. Data quality will be improved if a supplier has the appropriate services, such as default character setting, range checks on numeric data, address validation procedures, data enhancement and telematching;

- agreeing the format and schedule for data output in advance. Responses may be sent for data capture and returned in batches, or as a single consolidated set. If frequent, regular data transfers are required, it is advisable to use a bureau with electronic transfer facilities;
- obtaining signed, written agreements in advance of any work being undertaken. In addition to standard contractual obligations, these should also cover liability for data and confidentiality;
- ensuring that plans have been made for retrieval of original documentation. This may be a legal requirement for some product categories, in case a dispute or query arises. Documents may be converted into digital or optical files for faster retrieval and ease of storage. Ensure that correct procedures have been agreed for the disposal of all original documents, including incineration or shredding as appropriate.

Benchmark

The level of accuracy for data capture should be agreed in advance between client and bureaux. This should be defined for each job undertaken and is usually expressed in percentage terms, for example, % rejects and % invalid addresses. Request and check a test file to ensure these levels are being met. A 1 in N sample of the final file may also be checked.

Where data is to be validated and enhanced, agree separate rates of accuracy for matching to address verification files. The type of verification and definition of a match should also be specified.

As a minimum standard, data capture should achieve 95 per cent accuracy of addresses captured.

The client is responsible for ensuring that:

- methods of data collection are designed to maximise accuracy, legibility/audibility and completeness of information supplied by respondents;
- suppliers comply with Data Protection laws, as appropriate;
- specifications for data capture are supplied, stating levels of accuracy and matching required;
- procedures are in place for retrieval or destruction of original documents after data capture.

The supplier is responsible for ensuring that:

- work is only accepted for which it has the appropriate skills and technology;
- up-to-date verification tables are maintained;
- data and documents are processed and stored securely, in line with the Data Protection Act, and are returned or disposed of according to the client's brief once work is complete.

Section E

Receipt and Transfer of Data

The transfer of data files is a basic, but critical process. A single project may involve the sourcing and transfer of hundreds of separate files, with a delivery schedule covering several weeks. Managing and controlling this process is critical to the orderly handling of data.

All transfers, handling and storage of data must comply with the Data Protection Act 1998. Data owners are responsible for checking the security arrangements operated by any third party to which they transfer files. Data owners must ensure data will be held securely and files processed lawfully.

Every data owner will have a preferred file format. Where data is sourced from multiple owners, both from the client's in-house database and from third parties, this means the bureau will have to convert each file into a common format for processing. Documentation of each file layout needs to be supplied to allow the bureau to prepare conversion procedures.

Each file then needs to be notified to the bureau by the client, and logged on receipt. Accurate management of files in this way will allow cross-checking of planned data use with actual use. On receipt, the bureau should confirm the file layout, size, readability, and status against the client's data schedule. All files must be stored in a secure environment.

The bureau should be capable of handling data in most common media formats, including DAT, diskette, magnetic tape, cartridge and electronic data transfer. Notification should be given by the client of the format in which each file will be delivered, together with any unusual media to be used. Resupply of data will be necessary if a file does not tally with the documentation, or if it is corrupted.

Electronic data transfer requires particular care because of the increased possibility of data being corrupted during transmission. Documentation should be supplied either as part of the file transferred, or separately on paper.

Once processing has been completed, the data must be returned by the bureau to its owner. Industry practice is for the original data files to be returned exactly as supplied. Records on all data transfers must be maintained to allow for an audit trail once the process is complete, if required.

The client is responsible for ensuring that:

- a schedule of files to be used is supplied in advance;
- notification of media to be used, including uncommon formats, is provided;
- use of data complies with the Data Protection Act 1998 and is held, disclosed and processed lawfully;

- full documentation is provided for each file, covering project reference, file layout, supplier contact details, sample print, number of records and return instructions;
- test files are supplied when requested by the bureau;
- delivery schedules are met, unless previously notified;
- reasonable steps are taken to ensure that files supplied are virus free.

The supplier is responsible for ensuring that:

- all files received are checked against the client's schedule and examined for readability and size;
- any discrepancies or problems are notified to the client promptly;
- all data is processed and stored according to the Data Protection Act 1998;
- incoming data is checked for viruses and data owners are informed immediately of any problems;
- the schedule for processing is followed as agreed, subject to prior notification of any changes or delays;
- each file is given a unique identification to allow it to be reconstituted at the end of the project and tracked through processing;
- data is returned to the client/data owner at the end of the project in its original format, subject to prior agreement;
- the supplier should also consider whether the bureau should have professional indemnity insurance to cover the bureau's liability for loss, damage or theft of data whilst held and processed by the bureau.

Section F

Name and Address Conversion and Standardisation

Screening, deduplication and tagging or enhancement of data will all be carried out more effectively if name and address data is converted into a standard format. Each data owner – whether a commercial list provider, client, or computer bureau service provider – will store files in a unique format. This is designed to meet their internal processing requirements, but must be converted in order to allow multiple data sources to be simultaneously dealt with.

To ensure the highest levels of data accuracy in matching and outputting data, computer bureaux need to be able to identify and control all elements of an individual name record. There are four key elements:

- title;
- forename or initials;
- surname;
- suffix.

The management of these data elements is usually achieved through the maintenance of data tables for each. These will include standard abbreviations for titles (e.g. Mr, Mrs, Prof., Rev.), variants of forenames, and correct decoration suffixes (e.g. OBE, FRCS). Each data element is matched during processing to the relevant table for validation. Bureaux also maintain dynamic tables to allow for mis-spellings and omissions, as well as abbreviations, word joins or splits and transpositions.

Benchmark

The recognition and correction rules used by the bureau should be available for inspection by the client on request.

Address data is a particularly dynamic element. The postal address used for mailing purposes is not a geographic or physical description of a location – it is an instruction to Royal Mail on how to route post to a delivery point. To respond to changes in physical geography, such as new buildings and demolitions, or to improve its delivery efficiency, Royal Mail is continually changing postal addresses.

These changes are issued quarterly on the Postal Address File (PAF). PAF has been commercially available since 1974 as the standard point of reference for data matching. Bureaux hold historical tables to allow them to match old records as well as adopting each new version of PAF within a reasonable period of time.

Benchmark

Best practice is for bureaux to adopt changes to PAF within six months of their issue. Bureaux should be able to match address records using historic postal addresses as well as current formats.

Example

In 1993, Royal Mail made a significant change to addresses in Gretna.

OLD ADDRESS:	NEW ADDRESS:
149 Central Avenue	149 Central Avenue
Gretna	GRETNA
CARLISLE	Dumfriesshire
Cumbria	DG16 5AA
CA6 5AA	

Many individuals living in Gretna will continue to give their address to companies as Cumbria, since this is the geographic location. Bureaux will need to be able to match new records using Dumfriesshire to those using the old format.

By standardising addresses to match PAF, direct mail can be more efficiently delivered by Royal Mail. This allows clients to achieve discounts on their postal charges. By adding the Delivery Point Suffix, derived from PAF, further discounts are also possible.

Address matching does require decisions to be made by the client, however. While PAF represents the address format used by Royal Mail for its services, it is not the only address format in use. Local authorities, health authorities and Government all use differing addresses which reflect their own administrative needs.

Individuals may also add their own elements, such as house names or locality, while businesses may add building names and internal addressing details (e.g., fourth floor, North Block). These elements are referred to as customer preferred items.

The client must specify the level of match required before the bureau undertakes any processing. There are several options:

- matching to PAF – this will alter the address supplied to exactly match that used for postal delivery purposes;
- retaining customer preferred elements – as Royal Mail only requires the house or flat number and postcode in order to deliver mail, a matched record can still include elements supplied by a customer;
- retaining the original record – there are reasons why an ambiguous or inaccurate address may have to be retained, for example, when sending voting papers to building society members.

Additionally, each bureau will maintain its own matching files which may result in different outcomes for the same data. The client must define what source is to be used to validate a record and what level of matching accuracy is appropriate, together with the amount of customer preferred data which will be retained.

Benchmark

Best practice is for bureaux to be able to identify ambiguous addresses and report on the options for correction. The client may choose which match to apply or to leave records unchanged.

Example

An address may be supplied for validation which is ambiguous. It could refer to either of two real postal addresses.

SUPPLIED ADDRESS:	OPTION ONE:	OPTION TWO:
School House	School House	School House
Dorrington	Rodington	Church Road
Shewsbury	Shewsbury	Dorrington
SY4 4QL	SY4 4QL	Shewsbury
		SY5 7JL

Where overseas data is to be converted, the bureau will need to hold up-to-date reference tables for international address formats, including local language spellings and abbreviations. Early notification will be essential for such data sets owing to the non-standard nature of the processing.

Business data conversion involves the recognition and standardisation of an additional data element – job title. Job functions are given widely varying descriptions across industry. The bureau must be able to recognise these variants, including new buzzwords, in order to standardise them into a single format. Agreement should be reached in advance on the range of source titles to be standardised into a single new title.

The client is responsible for ensuring that:

- the supplier is advised of the nature of the data to be processed, particularly whether it is consumer or business, UK or international;
- clear instructions are given on the standard to be used for conversion and the tolerance applicable when converting data;
- any data queries raised by the supplier are dealt with promptly.

The supplier is responsible for ensuring that:

- external verification tables, such as PAF and Royal Mail postcode changes, are accurate and up-to-date;
- internally-generated tables, such as commonly used abbreviations, are kept up to date and are capable of recognising ‘wild card’ spelling errors;
- agreement has been reached with the client on the standardisation processes to be used and the levels of accuracy which will be tolerated;
- any data issues are raised with the client in a timely manner;
- results are audited and analysed and a report on any failures or exceptions is made available to the client.

Section G

Name and Address Matching

Matching name and address records is carried out for three key purposes:

1. to identify and/or remove duplicate records from a file. Lists and databases often contain internal duplicates, usually where the same individual has been recorded twice, or where the same individual has provided details in different formats. Matching these and suppressing them reduces wastage and improves the performance of a file;
2. to screen a file against other data sources for validation or suppression. The file may be matched against external data sources, such as the edited version of the Electoral Register, county court judgements, deceased or gone-away files. This process ensures that a communication is made only to individuals who have been verified to be at an address, or to be in the appropriate target group. Matching against the TPS or FPS file is required by legislation, where a communication is unsolicited matching against MPS, is required under the DMA's Code of Practice;
3. to enhance, or 'tag', data to a file. Additional data, such as date of birth, telephone number, or lifestyle characteristics, may also be added to a file from a matched external data source. This will help to improve targeting and may also be used when planning communications.

In each case, a decision will need to be made about the level of accuracy to be tolerated in matching. Each bureau will use a different technique. Consequently, different suppliers will achieve different levels of matching. It is important to understand how the technique used affects this accuracy and whether matches are in fact real. All software will also produce a certain error rate – the tolerance of this should be agreed in advance between the client and the bureau.

Example

In order to achieve correct matching, bureaux need to be able to identify ambiguous addresses and offer alternatives. Typical difficulties arise out of mis-spelt addresses that could be resolved into either of two places. Boston, Lincs and Bolton, Lancs are commonly confused.

INCOMING ADDRESS:	MATCH ONE:	MATCH TWO:
2 Church Road	2 Church Road	2 Church Road
Bolton	Farnworth	BOSTON
Lincs	BOLTON	Lincolnshire
	BL4 8AL	PE21 0LW

Where data is to be suppressed or enhanced, there are risks associated with matching. Removing a record which appears to be a duplicate, but which in fact is just very similar, such as identical surnames in the same households, could lead to a customer failing to receive information to which he or she is entitled. For this reason, financial services clients will often accept a lower level of match rate, for example. Equally, appending data to the wrong record might lead to inappropriate targeting of communications. A higher match rate could be called for in these circumstances.

Matching Levels

Where no data is to be overlaid, matching and deduplicating a file is usually carried out with a level of overkill – suppressing even doubtful duplicates in order to reduce wastage. Underkill is more appropriate where a data overlay is to be applied. This not only avoids the risk of incorrect targeting as noted above, it will also minimise the cost to the client of licensing this data.

As a rule, matching software should not be dependent on a single data element. This will avoid the suppression of a file as a result of a spelling error in the source file. A hierarchy should be agreed which weights each data element to be used in the match. For example, the postcode is a strong matching point, but should not be used in isolation since a single character difference could result in a failure to match. The second initial in a name is a weak matching point and may be overlooked where it differs, if all other elements are the same.

Consumer record matching

Consumer file matching can be undertaken at a number of different levels:

1. matched on title, initials/forename, surname and address;
2. matched on surname only and address;
3. matched on address only.

A single data processing project may require matching at more than one level. For example, when screening against a deceased file, the match is commonly undertaken at the surname and address level. Matching rented lists against each other might be at the finer, full name and address level.

The impact of each level must be clearly understood. Where address only matching is used, if multiple occupiers with different surnames are present at one address, only one of those records will be retained, for example.

Business record matching

Business file matching can be undertaken at a number of different levels:

1. matched on title, initials/forename, surname, company and address;
2. matched on company and address;
3. matched on address only.

Job titles and departments add a further degree of complexity to business data matching. For example, two records could share exactly the same individual and company name, but have a different job title – these may or may not be the same person.

Another difficulty is the ability to identify accurately all the supplied data elements. Both company names and job titles are often abbreviated and presented differently across files. The bureau must hold tables which are able to identify these as matches, for example, recognising International Business Machines and IBM as the same company.

Matching at a coarse level will have an impact on the final file size. For example, using address only will result in only one record being retained in a match where multiple companies share the same address.

Non-name and address matching

Another option available in a deduplication process is the use of non-name and address data. During a data tagging project where precise matching is important, the use of personal data such as date of birth or bank account number is a useful means of ensuring records to be merged are definite duplicates. The data element chosen will also have to be one which has a high level of population on the files being matched.

Benchmark

Best practice is for the text of the postal address to be identifiable and selectable at different levels:

- *fully correct;*
- *correct, but with a removable element after the thoroughfare;*
- *correct locality, town, county;*
- *address not recognised.*

Best practice for postcodes is to be able to identify and select as follows:

- *accurate;*
- *confirmed to thoroughfare level;*
- *confirmed to dependent locality level;*
- *confirmed to post town level;*
- *missing or unverifiable.*

The client is responsible for ensuring that:

- a clear and written brief is provided for the type and level of matching to be used;
- the acceptable degree of tolerance within each of the matching levels to be used is stated;
- further information requested by the supplier is supplied in a timely fashion;
- where a data audit (i.e., sample of file with verification of matching) has been requested, this is signed off promptly.

Section H

Deduplication and Merge-Purge

Where a variety of list sources are being merged to produce a mailing or phoning file, deduplication is an important process. It has several effects. Communications are made more efficient by reducing wastage. Customers and prospects are not alienated by receiving several identical messages.

To carry out the deduplication process, a hierarchy of the list sources being used must be constructed. This means that where a duplicate is found, the record on the list with the higher preference is used, while the list of lower preference is considered a duplicate.

This has an important impact on data costs, since under net name deals, only those records used will be paid for. How the hierarchy is constructed and used will depend on the client's marketing requirements, subject to the sensitivity of the message and the available budget.

The most common options are:-

Random	All lists are viewed as equally valuable. The record with the highest level of addressing is retained. This is commonly used when there is no experience of the lists, or in testing.
Cheapest Lists First	This provides a file with the lowest list cost (where net names rebates are all similar). This is useful where the client has no experience of the list.
Lowest Nets First	Those lists for which the client has agreed the lowest net name rebates are placed first.
Cheapest Cost Per Response First	This produces the most cost effective list. This is only possible where the client has previous experience of the list.

Some industry sectors require special care. For instance, financial services companies often have joint customers. This produces single records which may contain more than one individual name sharing an address. Deduplication against one of these files will require additional care to ensure that both the joint names can be used to match and suppress any duplicate.

Many clients recognise that they require different standards of deduplication depending on the information that they have about individuals. For instance, clients may wish to take greater care not to send a duplicate mailing to an existing customer than they require for simple prospect mailings. In that case, they may define a duplicate as anyone sharing a postal address with a customer or, where there is no customer presence, individuals sharing a surname and address.

Net name rebates are affected by where in a hierarchy the list is introduced for deduplication. The later in the process, the higher the number of duplicates that will be produced. To maintain trust in this process and in negotiation with list suppliers, the bureau must maintain and provide to clients complete audit trails. Reports on the validity of duplicates, in the form of samples of duplicates and of the deduplicated file, must also be supplied.

The client is responsible for ensuring that:

- a clear and written brief is supplied of the required definition of a duplicate, the hierarchy of list preferences, and any non-standard processing that is required;
- external data sources to be used for deduplication are supplied on schedule;
- agreements with data owners on net names are complied with in the hierarchy constructed.

The supplier is responsible for ensuring that:

- an accurate and complete audit report is provided showing the numbers of duplicates identified and their distribution across list sources;
- all counts provided will be auditable by printing the corresponding addresses, if required;
- processing is carried out in the order agreed with the client and in a timely manner.

Section I

Screening

Additional data about individuals allows marketers to make their communications more efficient by screening lists before use. This might result in the client choosing not to mail those individuals they believe will not respond to an offer, for example, or who may prove to be of poor commercial quality or low value.

Data used for screening generally falls into three categories:

Client specific

Data that the client has collected on individuals with whom it has traded previously is a highly discriminating screen. It includes files on:

- existing customers and prospects;
- individuals with whom the client has a previous bad trading experience;
- individuals who have requested not to be mailed, faxed or phoned.

Sector specific

Data that has been built within an industry sector may be pooled by companies operating in that sector for sharing. The two best known are files on:

- individuals with a poor credit record;
- individuals who have made insurance claims.

Generic data

Address based files are often used for screening. The most widely used include files on:

- individuals who have registered with the Mailing, Telephone and Fax Preference Services;
- individuals who are known to have died ('deceased');
- individuals confirmed to have moved ('movers');
- individuals who have not registered on the electoral roll ('no confirmed residency').

Use of the TPS and FPS files is a statutory requirement. MPS is also mandatory for members under the DMA Code of Practice. Failure to screen against the TPS, FPS or an in-house 'do not contact' list will have legal consequences. Screening against client-specific data is highly recommended. Clients may choose whether or not to use other data always subject to the need to ensure that processing is fair and lawful.

This choice will usually be made on the grounds of improved cost-effectiveness of communications. The client will need to calculate the cost of using industry and generic data for screening, together with processing costs, and compare this with the savings made by not mailing or phoning those records.

Generic data requires extra consideration and needs to be treated with care. Points to note include:

- DMA members are required to screen against MPS except in relation to existing recent customers and enquirers. Screening against TPS and FPS is a statutory requirement;
- deceased data is derived from official records on which use of formal forenames is required. Many people do not give their full forename when completing forms for commercial transactions, so match rates against this file are usually low;
- Royal Mail's National Change of Address File allows existing customers to be tracked from one address to the next, but only covers those individuals who subscribe to the mail redirection service;
- movers' files are usually derived from returned mail and tend to be inaccurate. Some commercial suppression files use a second source to verify a move – check which method has been employed;
- no confirmed residency files are derived from the electoral roll and need to be treated with caution. Not everyone registers to vote and even when newly available to direct marketers, the roll is about eight months out of date. Individuals who are not confirmed as residents at that point may have moved to an address subsequent to the compilation of the voting register;

Screening often takes place within the deduplication process. However, with industry specific data and electoral roll residency files, screening is usually carried out separately after deduplication.

The client is responsible for ensuring that:

- a clear and written brief is provided of the screening process to be used and the level of screen to be used;
- external data sources to be used for screening are supplied to the bureau in a timely manner.

The supplier is responsible for ensuring that:

- screening is carried out in accordance with the clients' brief and schedule;
- an accurate and complete audit report of the entire process is provided showing the numbers of individuals screened.

Section J

Data Tagging and Enhancement

Data tagging is the addition of extra data to the client's database from external sources. The purpose of data tagging is two fold:

- to improve the data used in targeting, especially if statistical modelling is used for selection;
- to generate a more individual message within the mailing (for example, a sales message that relates to the individual's date of birth).

The sources most often used for the external data are:

- the edited version of the Electoral Roll can be used;
- lifestyle databases;
- pooled databases of clients' trading data.

The advantages of data tagging can be substantial. However, the costs of tagging are usually considerable. To be cost effective, it is likely that only selected segments of a customer file will undergo data tagging.

There are also considerable risks involved. For example, if the tagged data is not used correctly, customer annoyance may outweigh any marketing benefits. This may occur if:

- data has been incorrectly captured or is inaccurate;
- data has been incorrectly matched;
- data is incomplete.

The first of these two risks can be mitigated by careful matching. Matching procedures need to ensure that there is minimal possibility that two individuals within a household or company may be confused. Matching at surname and initial may not be sufficient if there is any likelihood that there will be two individuals with the same surname and initial within the household or company. Ideally, a match will be made using additional data, such as date of birth or job title.

The problem of incompleteness – not being able to tag data for all individuals on a file – applies mainly where the data is being used for the creative message. This could compromise the creative treatment if it relies on the missing data element being inserted into the message.

Statistical and modelling tools exist which can predict what the missing data should be. These can be used with varying levels of confidence, depending on the level of records for which that data element is present. Segmentation and selection tools will have procedures for handling data sets with missing data elements.

The client is responsible for ensuring that:

- a clear and written brief is provided on the type and level of data to be tagged;
- any external data source from which data is to be tagged is supplied in a timely manner;
- tagging levels are agreed and a sample of matched records is signed off.

The supplier is responsible for ensuring that:

- the client is advised on the tagging options and their implications;
- tagging is carried out in a timely fashion and to the agreed specification;
- tagged files and external data sources are returned promptly and in good order.

Section K

Sortation and Output

Data can be sorted into any sequence required by the client. Typically, output is sorted to Royal Mail's Mailsort sequence or DPS code to maximise postage discounts. If data is to be used for telephone marketing, other sequences may be appropriate, such as regional or random.

Unless the data processing is purely for analysis, the final step in many projects is the production of a tape file or label print-out. Most bureaux will be able to provide data on numerous different media, including various forms of both electronic and paper output.

It is important that early consideration is given to the final output required. This allows the bureau and printer to confirm output media and formats in advance of the project's completion. The supply of test data to the printer, including sufficient information for the client to be able to sign-off a test printing, should be included as a key task in the project plan.

If this procedure is followed, final output checking can then become the supplier's responsibility. The computer bureau should have internally checked and signed-off any file, report or label print run being despatched to the client or printer.

The client is responsible for ensuring that:

- a clear and written brief is provided detailing the sortation and output media required;
- sample files and print-outs are promptly checked and signed-off.

The supplier is responsible for ensuring that:

- sign-off has been obtained on file or label layouts, test files, print runs and output quantities;
- external tables, such as Royal Mail's Mailsort table, are maintained and up-to-date;
- sortation and output are carried out according to the client's brief and in a timely fashion;
- delivery of outputted data to printers is carried out on schedule.

Appendix – Useful addresses

DMA (UK) Ltd

Haymarket House
1 Oxendon Street
LONDON, SW1Y 4EE

T 020 7321 2525
F 020 7321 0191
E dma@dma.org.uk
W www.dma.org.uk

DMA North

29 Wellington Street
Leeds, LS1 1QQ

T (0113) 244 7103
F (0113) 244 7224
W www.dma-north.org.uk

DMA Edinburgh

41 Comely Bank
EDINBURGH, EH4 1AF

T (0131) 315 4422
F (0131) 315 4433

Data Protection Registrar

Wycliffe House
Water Lane, Wilmslow
Cheshire, SK9 5AF

T (01625) 545745 enquiries
(01625) 545740 registration
W www.open.gov.uk/dpr/

Mailing Preference Service

5th Floor
Haymarket House
1 Oxendon Street
LONDON, SW1Y 4EE

T 020 7766 4410
F 020 7976 1886

Telephone Preference Service

5th Floor
Haymarket House
1 Oxendon Street
LONDON, SW1Y 4EE

T 020 7766 4420
F 020 7976 1886

Fax Preference Service

5th Floor
Haymarket House
1 Oxendon Street
LONDON, SW1Y 4EE

T 020 7766 4422
F 020 7976 1886

Royal Mail Streamline

Streamline House
Sandy Lane West
Oxford, OX4 5ZZ

T (01865) 748768
F (01865) 780312

Advertising Standards Authority and Committee of Advertising Practice

Brook House
2 Torrington Place
LONDON, WC1E 7HW

T 020 7580 5555
F 020 7631 3051

List Warranty Register (LWR)

5th Floor
Haymarket House
1 Oxendon Street
LONDON, SW1Y 4EE

T 020 7766 4450
F 020 7976 1886

List and Data Suppliers (LADS)

5th Floor
Haymarket House
1 Oxendon Street
LONDON, SW1Y 4EE

T 020 7766 4430
F 020 7976 1886