

<b>DATE</b>	March 2008
<b>TITLE</b>	Database Marketing
<b>CIRCULATION</b>	6,000



Prospect Data Management

Once you've decided to build a prospect pool, how should you select the data to put in it? James Lawson investigates.

## Fill the pool

**W**orking with a prospect database is increasingly acknowledged as the most cost-effective way for serious direct marketers to acquire new customers. Farming a semi-permanent file rather than hunting and gathering lists every time you want to send a campaign out of the door takes marketing up the evolutionary chain, and learning can be fed back to a central point for future application too.

So the argument is over and a prospect pool is on the horizon. But what are you going to put in it?

### What's in the pot?

The initial attraction of a prospect database for many is the opportunity to save money when buying data by negotiating terms in advance. There no doubt that buying all the data you need for the year up-front is a lot cheaper than *ad hoc* list purchasing. However, many companies tend to maximise the size of their discount by doing a deal with a single data provider and so sticking to a single source. Does this lead to a lower performing pool?

Opinions tend to be split on the route to take, with some advocating simply selecting a tranche of lifestyle data or geodemographically-coded Electoral Roll data

that matches your customer profile, while others would take a more traditional route in bringing together tens of lists and updating them quarterly thereafter while testing new sources. The middle way is to use some sort of base file and pull in valuable extra sources if they will really make a difference.

"You need to use a base file with other information added to it," says Richard Higginbotham, head of marketing at CDMS. "On top of that, you can layer your response data and other elements of prospect tracking like how many times a contact has been mailed."

Multisourcing a file may make the task of building a pool more complex but has a number of benefits. For a start, even if all you are doing is adopting worst practice and carpet bombing the Electoral Roll – the approach used by US credit card companies over the past decade to build their UK market share – maximum coverage will come from multisourcing different versions of the Electoral Roll. If the aim is to whittle down a UK base file to match your needs, having the biggest possible footprint to start with has to be a good idea.

In a study by The Customer Partnership in 2005, Julian Berry and Paul Rowson found only a 75 per cent overlap between two competing replacement



You need to be able to process quickly and accurately

Rob Salmon, managing director, meta-morphix.

Electoral Rolls. So the consequence of going with just one Electoral Roll supplier is either a 25 per cent loss of mailable volume, or having to mail parts of the file that are much lower down your propensity models' gains charts, with a consequent drop in response rates. As the opt-out level has risen by nearly ten per cent since that study, the figures would be even more compelling now.

Using multiple files should also give more variables to use for predictive modelling. Berry and Rowson found that the four separate sources they used were all involved in each of the five propensity models they built. "We could not have scrapped any of them without seriously weakening our models," says Julian Berry, managing director of The Customer Partnership.

Multisourcing will also help reliability by providing multiple instances of the same individual or household, and so multiple examples of critical variables like date of birth or insurance renewal date. Based on the reliability of each source, the data processor can choose to pick one piece of data from a record, merge other fields or simply ignore it.

"We currently take 38 different feeds to build our own pool and we have a hierarchy of data survivorship rules which govern which source takes precedence," explains Zoe Vine, head of Data Services at The Trading Floor. "Multiverification helps increase confidence in many areas, particularly residence."

Set against these benefits is the extra work and therefore cost involved in the initial build, then the ongoing maintenance cost plus the liaison work with each supplier: the more sources a file has, the more complex it will be to refresh. Each data supplier will supply updates at different intervals, and each one will have a different idea of what "update" means, so you could be processing the same data over and over again.

Building a model may also be slightly problematic if, despite a lengthy search, variables are not populated or consistent across the whole data set. However, there are ways around this.

"You may have to build separate models for individuals that appear on one source but not on others," says Berry. "But usually have 80 to 85 per of them across your main sources so it's not as bad as it sounds."

### Test your files

How to choose the size and composition of the base file, and which lists to layer on top is the heart of the challenge. Tim Beadle, managing director of Marketing Improvement, says that paring down the size of the "addressable target universe" should be the first step.

"Take a brutal and pragmatic view," he says. "If you are selling Jags, then you don't want high-rise council block tenants. Far too many people try to make their prospect database as large as possible which carries a massive cost."

Just as in classic profiling and targeting, understanding which are the key attributes of your own customer base will help inform which other files might help improve prospecting performance – though restricting a pool solely to the profile of your existing customers could limit growth.

The desirability of extra files will obviously vary for each sector: car insurance renewal dates will be top of the list if you are marketing car insurance, while indicators of a propensity to go on foreign holidays will attract travel operators. Co-op data showing multi-buyers is a likely purchase for a mail order company and mortgage providers may be attracted to some of the warm lead-to-purchase data aggregated by the likes of The Trading Floor.

Some marketers may have their heart set on using a certain set of lists that have worked for them in the past. Alan Thorpe, commercial and operations director at G2 Data Dynamics, gives the example of a financial services company that wanted to employ lists of readers of financial magazines to drive its contact programme during the ISA season.

"But they could only get a couple of hundred thousand names from those," he says. "By imputing variables from the lists across a larger EuroDirect base file, they got the volume for roll-out along with the accuracy they needed."



Far too many people try to make their prospect database as large as possible

Tim Beadle, managing director, Marketing Improvement.

However, it's possible to go too far in adding response-built lists, and Beadle advises caution. "Sure they responded, but do you know why?" he asks. "Or why they purchased? It's ok if the list is precisely right but the overwhelming majority of lists don't have the information on them to help you make that decision."

He states that it can often be best to stick to a simple base file along with a few highly predictive variables like age, gender, location and income. Targeting can then be based on those in combination with geodemographics.

"Using the most comprehensive sources to get the right footprint and selecting rigidly on tight geodemographic values works very well," Beadle argues. "Where people live is the most highly predictive variable out there."

So when it comes to deciding which selection of sources will work best in a pool, the only really accurate way to do it is, as ever, to test in advance. No discount is worthwhile if the data involved doesn't work.

"You need to get your short list of potential data suppliers to match up their data to your model development data sets, and then let the statisticians tell you which data items you need, and how much you need any particular piece of data," advises Berry. "No-one can tell you which are the right variables until you model and test, but very few people do that. If the data item is very expensive and the least important variable in your propensity model, you could decide to drop it and save the money."



**No-one can tell you which are the right variables until you model and test**

Julian Berry, managing director,  
The Customer Partnership.

But having to think two or three years ahead and set up a testing programme can be rather daunting when the priority is to get the pool built. No wonder so many companies listen to the sales pitch and sign up with a single data owner.

"It is very hard to test the pool components with different timings and creative," says Vine. "It's a different thing to classic DM list testing. But if you're going to licence up front for three years then you simply will

have to do it."

One option that holds out the promise of cost and time savings is to employ pre-merged pools. Why reinvent the wheel? This encompasses products like CACT's Ocean and Alchemy's AMS pool and also includes Ai Data Intelligence's pooling proposition.

"We have a large merged database as standard," says Ai's managing director Jon Cano-Lopez. "We can build client-specific views based on which supplier data they want to buy."

The company has recently added credit data to its merged file and also offers Pipeline, a web-based pay-as-you-go list sales service that pulls records from the same central database.

"This might not necessarily be the whole answer but it has to be better than a single source," comments Berry.

In all cases, it's well worth checking that the company building your pool will be able to source all the data you might need from the major owners. In some cases, it has been known for data owners to refuse to supply some service providers and not just because they threaten their own prospect pooling business. For example, one major data owner has previously refused to supply its file to a third-party provider because the terms of the proposed contract meant that the provider would have had unlimited use of the variables when transformed into scored values within a pool.