LexisNexis
RISK SOLUTIONS

Driving Smarter Outcomes:
Achieving a single customer view

Table of Contents

1. How do you create a single customer view across your business?

A single customer view is defined by Gartner as an 'accurate, timely and complete view of the customer across multiple channels, business lines, and potentially enterprises, where there are multiple sources of customer data in multiple application systems and databases'.[1] This unified view is consolidated online and offline and is a prerequisite of businesses today. Regulation demands it: GDPR requires that businesses collate and catalogue all the data they hold on their customers. Businesses such as AirBNB, Uber, Google, Alibaba and Amazon rely on it for competitive advantage: using comprehensive customer data to better understand their customers and make quicker and smarter decisions to improve customer service and satisfaction.

For a single customer view, an organisation needs to be able to quickly:

- Bring together different types of data from various data sources, often in different formats (for example integrating structured and unstructured data). There can be a great deal of difference in the type of information presented for a mortgage application versus a mobile phone contract. Combining data sources can be a particular challenge following mergers or acquisitions, where organisations are looking to integrate a range of different data-points on customers, from disparate systems;

- Compare the data records to determine whether they belong to different individual customers or the same customer. The difficulty here is that customer data decays at 30% per annum[2] - as people move house, change name, change contact details and don't necessarily notify the organisations of these updates;

- (Where data records are identified as relating to the same individual), determine which values are the most up to date and are aligned to the organisation's purpose;

- Create an accurate master record for this information that the right people in the organisation recognise as the right information and can access easily.

New and emerging technologies and applications can help with the speed and efficiency of data capture and processing. AI and other advanced analytics techniques can help with the analysis and interpretation of this data to make smarter decisions to tailor customer experience, but all of this relies on the customer data being complete and accurate.

Without high quality data, even the most frictionless customer journey is rendered meaningless.

2. Consolidating data records

Organisations typically try to consolidate their customer data records into one central master data management system – a MDM – which, as Gartner notes, is frequently misunderstood to be a panacea for data issues.[3] Moving the data into the MDM can be a mammoth task. One alternative to this is for businesses to instruct a third party vendor to connect all their existing systems to a specialised management platform. Either way, the organisation will need to be able to link customer data records and consolidate them into a single customer view.

The traditional method for linking customer data records is via 'rules-based linking'. This is where different data fields in the customer records are compared and evaluated according to a pre-defined rule or set of rules. Rules-based models struggle to cope with the data quality challenges that different data input methods and continuous data decay present. As the data-set grows and the information collected becomes more complex, supplementary rules need to be introduced to address the gaps and this increases the chance of spurious data links.

1. https://www.gartner.com/doc/455611/create-single-customer-view-customer
2. https://www.business.com/articles/prevent-data-decay-from-ruining-your-crm/
3. https://blogs.gartner.com/andrew_white/2017/10/05/whats-happening-in-master-data-management-mdm-land/

3. Statistical matching for better data record linking results

An alternative approach to linking data records is 'statistical matching'. In this case, instead of using rules to compare between two records, a statistical matching method looks at the record population as a whole and assesses the uniqueness of the information as well as looking for matches between individual data records. This allows for both better matching on an initial comparison as well as increasing the confidence with which data records are joined together, when consolidating historical information with new details obtained.

As an approach, statistical matching is nothing new. It's been around for decades. Historically, there were two limitations (as noted in the 'Automatic Linkage of Vital Records' paper by Newcombe, Kennedy, Axford and James - 1959):

1) Computer power (which is no longer an issue)

2) Data sources

The authors of the paper discovered discrepancies resulting from inconsistent recording of the data in one source; this made the process of linking the records difficult, especially where the frequency of a particular value was high, such as common surnames. Comparisons across two large datasets (with some of the same individuals), showed markedly better results than with the single source. In statistical models, the larger and more varied the underlying data sources, the better.

4. The importance of validating data against authoritative sources

The creation of master records, by definition, involves examining all an organisation's customer data and making programmatic choices as to which information can be trusted. These decisions can be informed by looking at all of the customer data within a business, but this insight alone will not be robust enough to make full use of a statistical model.

This is a significant challenge for master record projects. The data from these projects is consumed as a service by other business units and the impact on the organisation can be profound. As already highlighted, data that is keyed in wrongly or has changed or corrupted can severely hamper the usefulness of this information. Validation of the customer record against authoritative sources such as public records (for example, electoral roll), credit bureaus and large consumer datasets can enhance the data quality of a single customer record making any subsequent use of the data more effective and preventing costly errors. Statistical record matching is particularly well suited for this kind of data quality check, allowing the cross referencing against multiple sources, with a robust single customer view across all data in the system. The statistical models are actually (as noted in the 'Automatic Linkage of Vital Records' paper by Newcombe, Kennedy, Axford and James – 1959), strengthened by additional sources. This validation work is most effective for organisations who repeat the process on an ongoing basis to mitigate for the almost one in three records that decay each year.

Such a programme of ongoing data hygiene, statistical matching and distribution across an organisation is a big undertaking. Many do not have the budget or the appetite for tackling this challenge. Where such projects take place, they have a high failure rate due to the complexity and their frequent reliance on a single additional source for validation.

5. In summary

Organisations need a holistic view of their customers in order to deliver the best possible customer service. Statistical modelling and advanced analytics techniques, designed to improve customer experience, rely on companies being able to extract high quality data on their customers.

The rate of data decay, combined with human error in data entry and issues with integrating data from different sources and systems, means that it can be very hard to consolidate data records and arrive at a complete and up-to-date picture of customers.

Statistical matching enables better matching of customer data records across an organisation and throughout the customer journey. Statistical matching is most effective when there is high volume of data records. For it to be most successful, data should be validated against authoritative sources on an ongoing basis.

6. In practice

LexisNexis® Risk Solutions in the UK decided to make the switch from rules-based to statistical linking for our customer data records. Our experience has been very positive, both in terms of increasing record matches and improving the quality of the data we hold on our customers. We invested significantly in expanding the underlying third party data sources we use to validate our data, bringing in data from credit bureaus, multiple public records and other consumer data to support statistical matching and ultimately, to help provide a more complete picture. In terms of overall impact, statistical matching has enabled us to increase matching rates from 52% to 79% and with improved data accuracy.

7. Conclusion

There is no doubt that statistical matching is a better solution than rules-based matching for organisations with a significant volume of clients. However, the costs (investment in IT infrastructure and external data sources for data validation) and resource (data scientists who understand and can apply the technique), can be prohibitive.

A more cost-effective alternative for organisations who do not wish to scale up their information sources and build their own capability, is to match against a data service built on statistical modelling, using a large volume of authoritative sources to achieve high levels of accuracy.

LexisNexis Risk Solutions specialises in statistical matching. In the UK, we have one of the most comprehensive consumer data-sets and are investing continuously to build this out further. We can work with you to help you cleanse, enrich and supplement your data and to achieve a more holistic view of your customer.

Please contact us for a free consultation.

ukenquiry@lexisnexis.com
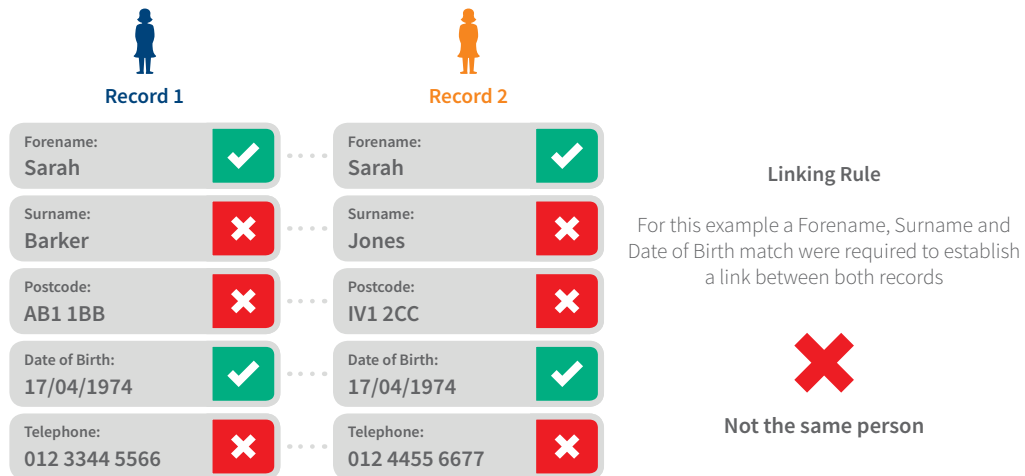
8. Statistical matching: How does it work?

Data capture is multi-channel and is collected at different times throughout the customer journey, in different formats (eg. online and offline interactions). Data input is subject to human error and data decays as customer situations change, so customer data can quickly be outdated.

Constant changes to your customers' circumstances causes your data to become inaccurate
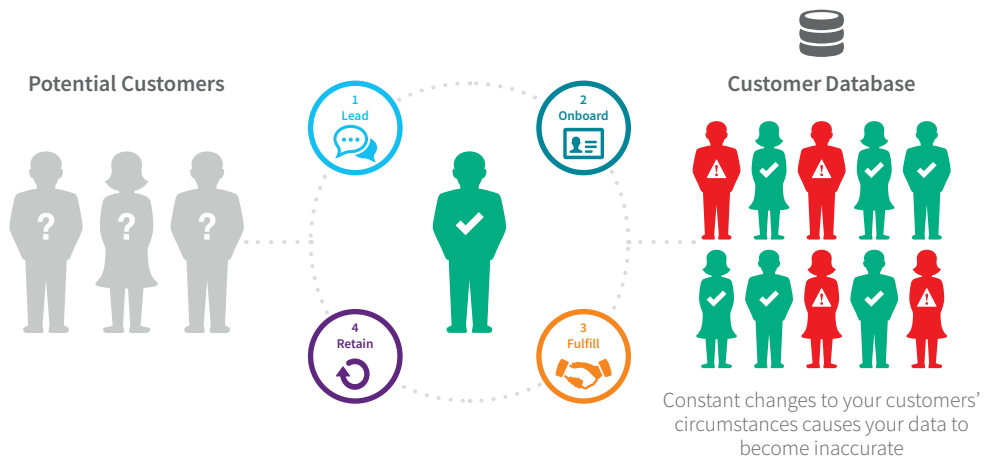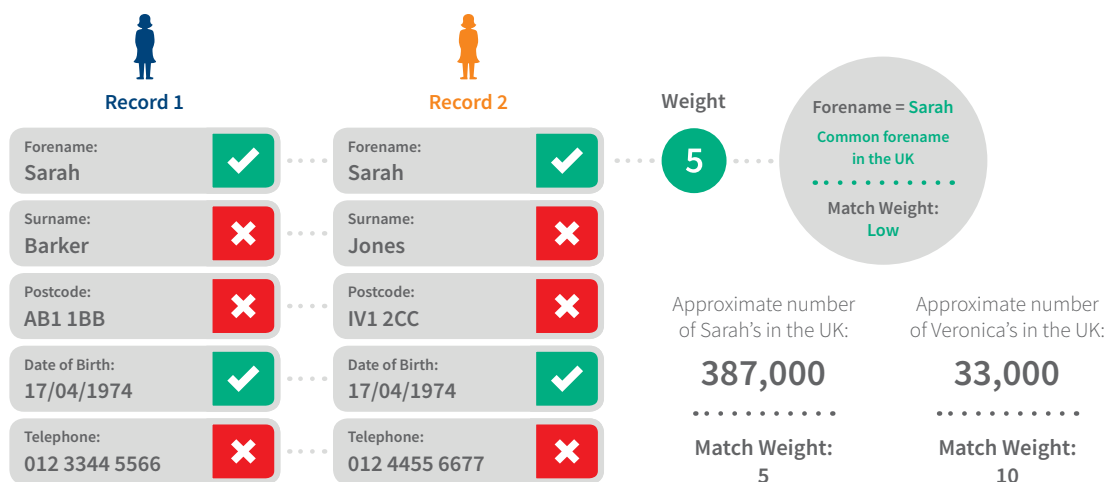
As organisations grow, whether organically or inorganically, customer data increases in size and complexity, perhaps due to new information channels or else resulting from the need to integrate different databases. Bringing this data together, for example in data lakes or other consolidated views, can give rise to multiple errors, which impacts on the organisations ability to deliver an optimal customer experience.

Multiple disparate systems can lead to inconsistent data and duplicate customer records
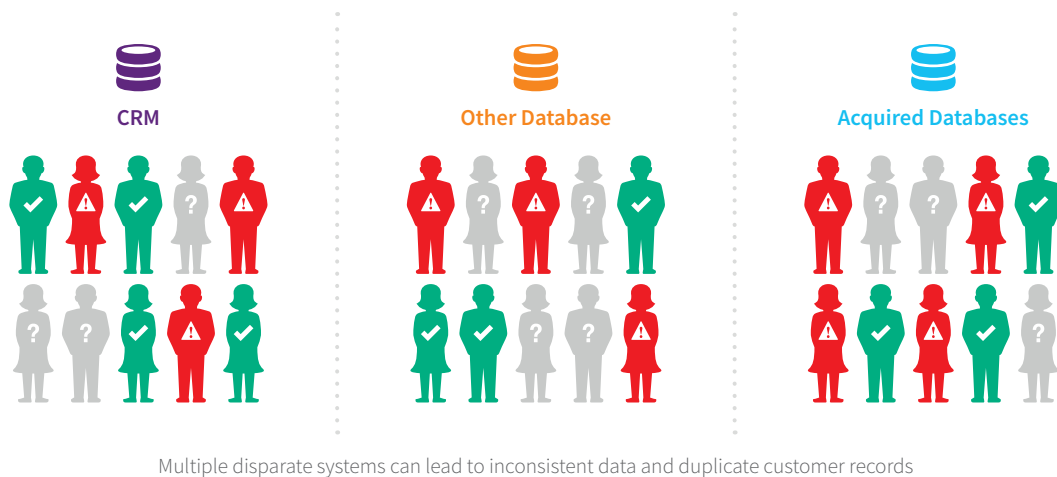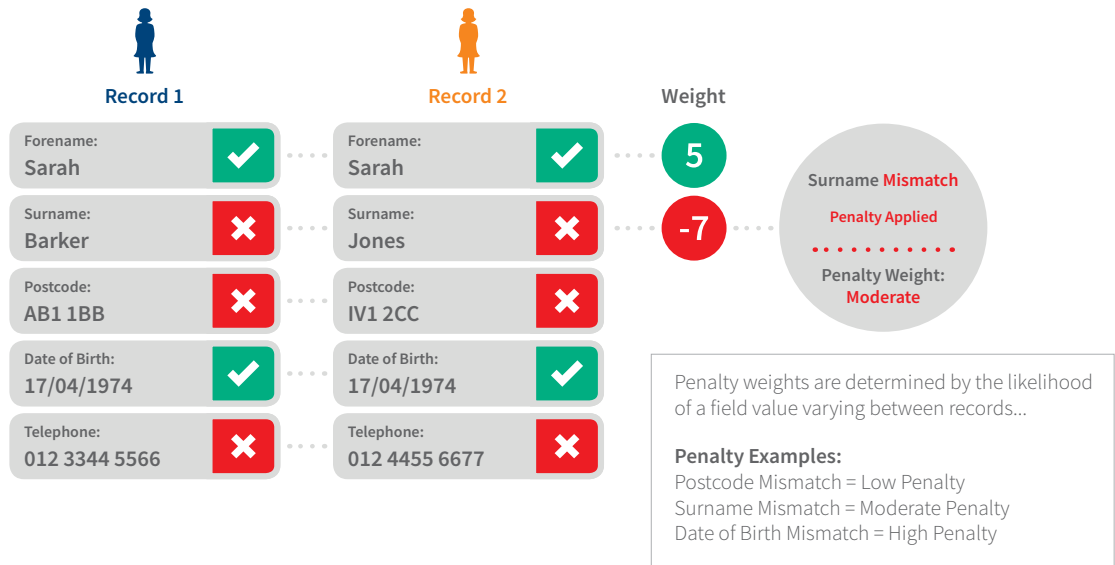
The most prevalent method for consolidating data records is rules-based linking. In the case below, the rules-based model compares various data fields and concludes there aren't enough matches to consolidate the view.

Record 1

Forename: Sarah	✔
Surname: Barker	✘
Postcode: AB1 1BB	✘
Date of Birth: 17/04/1974	✔
Telephone: 012 3344 5566	✘

Record 2

Forename: Sarah	✔
Surname: Jones	✘
Postcode: IV1 2CC	✘
Date of Birth: 17/04/1974	✔
Telephone: 012 4455 6677	✘

Linking Rule

For this example a Forename, Surname and Date of Birth match were required to establish a link between both records
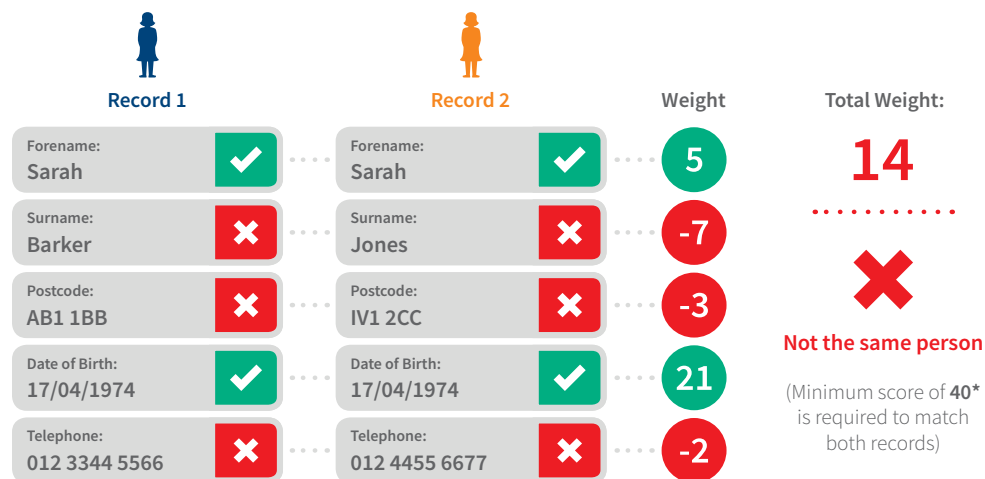
✘

Not the same person

However, using a statistical matching approach means that positive scores are applied to data fields that are matched and that weight is adjusted according to the frequency of that particular value occurring in the total record set. The more common the value, the smaller the weighting applied, as in the example below:

Record 1

Forename: Sarah	✔
Surname: Barker	✘
Postcode: AB1 1BB	✘
Date of Birth: 17/04/1974	✔
Telephone: 012 3344 5566	✘

Record 2

Forename: Sarah	✔
Surname: Jones	✘
Postcode: IV1 2CC	✘
Date of Birth: 17/04/1974	✔
Telephone: 012 4455 6677	✘

Weight

5

Forename = **Sarah**
Common forename in the UK

Match Weight: Low

Approximate number of Sarah's in the UK:

387,000

Match Weight: 5

Approximate number of Veronica's in the UK:

33,000

Match Weight: 10

In this same example, negative scores are applied for differences in the comparison. The weight of these penalties provides the model with context as to the likelihood that the difference means the records are different. For example, differences in date of birth would carry a more severe negative score than differences in surname.
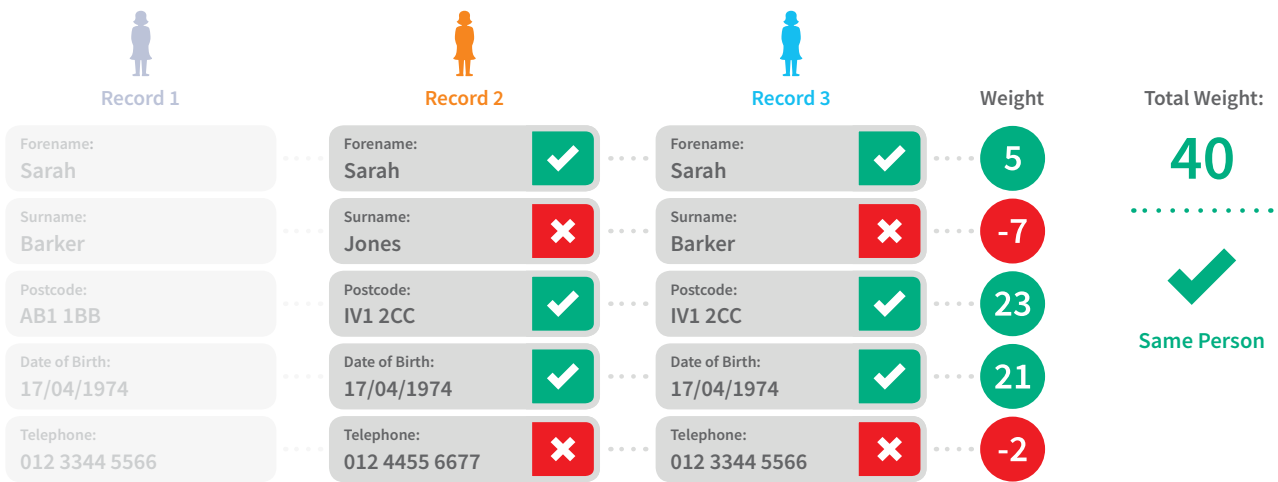

Record 1		Record 2		Weight
Forename: Sarah	✔	Forename: Sarah	✔	5
Surname: Barker	✘	Surname: Jones	✘	-7
Postcode: AB1 1BB	✘	Postcode: IV1 2CC	✘	
Date of Birth: 17/04/1974	✔	Date of Birth: 17/04/1974	✔	
Telephone: 012 3344 5566	✘	Telephone: 012 4455 6677	✘	

Surname Mismatch

Penalty Applied

Penalty Weight: Moderate

Penalty weights are determined by the likelihood of a field value varying between records…

Penalty Examples:
Postcode Mismatch = Low Penalty
Surname Mismatch = Moderate Penalty
Date of Birth Mismatch = High Penalty

In the completed example below, the total score does not pass the required threshold…

Record 1		Record 2		Weight
Forename: Sarah	✔	Forename: Sarah	✔	5
Surname: Barker	✘	Surname: Jones	✘	-7
Postcode: AB1 1BB	✘	Postcode: IV1 2CC	✘	-3
Date of Birth: 17/04/1974	✔	Date of Birth: 17/04/1974	✔	21
Telephone: 012 3344 5566	✘	Telephone: 012 4455 6677	✘	-2

Total Weight:

14

✘

Not the same person

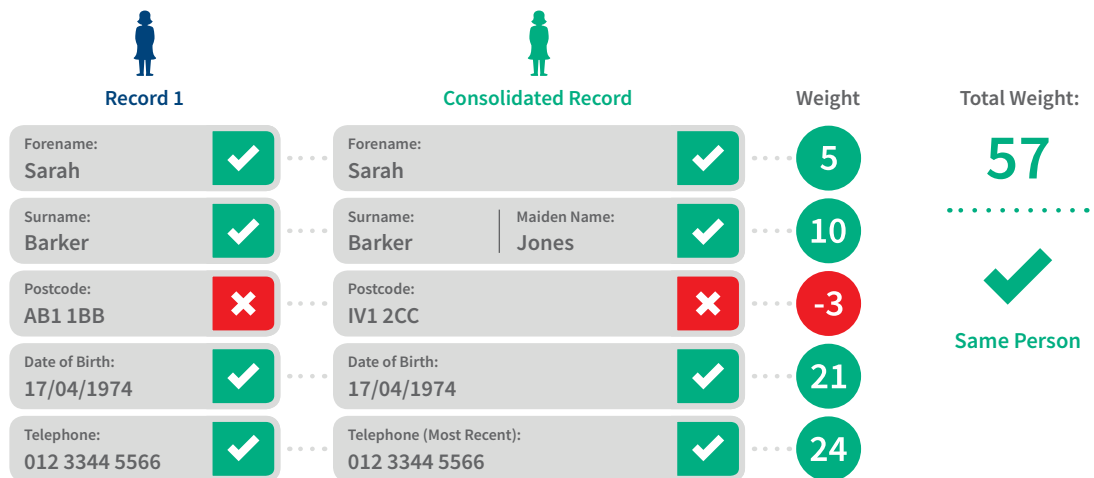(Minimum score of **40*** is required to match both records)

Record 2 is then compared to another record…

---

* Our Scalable Automated Linking Technology (SALT) determined that 40 is the threshold at which two records can be linked with a confidence level of over 99.9% - this threshold is continuously tested and calibrated for optimum linking accuracy.

This time the scoring passes the model's threshold and so the model considers the records to belong to the same person.

	Record 1	Record 2		Record 3		Weight	Total Weight:
Forename:	Sarah	Sarah	✔	Sarah	✔	5	40
Surname:	Barker	Jones	✖	Barker	✖	-7	
Postcode:	AB1 1BB	IV1 2CC	✔	IV1 2CC	✔	23	✔
Date of Birth:	17/04/1974	17/04/1974	✔	17/04/1974	✔	21	Same Person
Telephone:	012 3344 5566	012 4455 6677	✖	012 3344 5566	✖	-2	

A master record is created, which can then be re-compared to the first record, producing a more accurate view again of the individual, as more data contributes to a more comprehensive view.

	Record 1		Consolidated Record		Weight	Total Weight:
Forename:	Sarah ✔	Forename:	Sarah ✔		5	57
Surname:	Barker ✔	Surname: Barker	Maiden Name: Jones ✔		10	
Postcode:	AB1 1BB ✖	Postcode:	IV1 2CC ✖		-3	✔
Date of Birth:	17/04/1974 ✔	Date of Birth:	17/04/1974 ✔		21	Same Person
Telephone:	012 3344 5566 ✔	Telephone (Most Recent):	012 3344 5566 ✔		24	

For more information, call 029 2067 8555
or email ukenquiry@lexisnexis.com

risk.lexisnexis.co.uk

LexisNexis®
RISK SOLUTIONS