

# The Business Applications of Identity Resolution

Robin Bloor, Partner  
Dr. Fern Halper, Partner





© Copyright 2006, Hurwitz & Associates

All rights reserved. No part of this publication may be reproduced or stored in a retrieval system or transmitted in any form or by any means, without the prior written permission of the copyright holder. Hurwitz & Associates is the sole copyright owner of this publication.

■ 470 Totten Pond Road, 2nd Floor ■ Waltham, MA 02451 ■ Tel: 781-890-7185 ■  
www.hurwitz.com



## The Business Applications of Identity Resolution

### Introduction

All organizations hold and use information about people. These people may be customers, employees, contractors, agents, suppliers, or contacts. The information that is held is gradually accumulated over time. Some of it may have been acquired through agencies or via indirect sources of data. Some of the information may have been provided on an as-needed basis by ad-hoc information sources.

Unfortunately, such information can be and often is inaccurate in many different ways and this lack of accuracy creates a need for Identity Resolution. Identity Resolution technology is software that applies a set of logical rules to information about people in order to increase accuracy and reduce the possibility of misidentification. Misidentification can, of course, have a large impact in specific areas such as money laundering and fraudulent claims as well as an across the board impact in corporate functions such as marketing.

This Hurwitz & Associates paper examines Identity Resolution technology, what it is, how it works, and why it is important. We also look at how Identity Resolution serves as the foundation for two other important technologies – Relationship Resolution and Anonymous Resolution and IBM's solutions in this area. Finally, we provide an illustrative case study that describes the experience of a financial institution using this technology to manage identity information, detect potential fraudsters and money laundering activities and comply with The Patriot Act.

### What is Identity Resolution Technology?

#### *Identity Resolution and Misidentification*

Personal information is used in a many different ways. Often, it will be connected to information from other sources in order to find out more about the individual. This might involve, say, a check for honesty (does the individual have a criminal record?) or a check for entitlement (is this a frequent flyer?). Whatever the circumstances, the misidentification of individuals is possible and, of course, undesirable. Nevertheless, it occurs quite frequently.

There are a variety of reasons for this:

- **Lack of a Unique Identifier for ID:** The primary personal identification attribute is the first and last name, but this is very rarely unique. Confusion is clearly possible between the records of any two people with the same or similar names.



## The Business Applications of Identity Resolution

- **Lack of Precise Data Standards:** There are no standards for the recording of many items of personal information. Even in the recording of a name, for example, David Jones, David R Jones and David Jones Jr. may all refer to the same person but be recorded differently simply because different data entry forms are used.
- **Changes to Identifying Information:** Some items of personal information, including names, addresses and telephone numbers, change. Change of name is a particular source of ID problems, especially with women where at least one name may change through marriage.
- **Data Entry Errors:** Data entry errors are a common cause for information being incorrect. The error rates of data entry clerks can be as high as 13% to 15% of the data entered and even when effective quality techniques are implemented, error rates rarely fall below 1%. Information gathered directly over the Internet is particularly susceptible to error as there is little possibility of applying quality control.
- **Deliberately Misleading Information:** People sometimes provide deliberately misleading information to conceal their true identity or to deliberately disassociate themselves with other information held about them. This may be done with criminal intent in mind, but can also have other motivations. Criminals in particular often maintain multiple identities.

*The error rates of data entry clerks can be as high as 13% to 15% of the data entered and, even when effective quality techniques are implemented, error rates rarely fall below 1%.*

### **Identity Resolution and Information Enrichment**

The goal of Identity Resolution Technology is to dramatically reduce or completely eliminate misidentification. It works by comparing all the information that might relate to an individual from all available sources and applying validation rules that are based on common sense, probability and experience to determine what is most likely to be the truth.

To provide a simple example: There may be many records with the name John Smith and there will be other records with names like Jon Smith or John F Smith that may or may not refer to the same individual. If you now try to resolve on, say, date of birth and spouse's name you can eliminate some of these records and confirm others as very likely to be accurate. The more associated details you can resolve on, the greater the certainty you can have that you have disambiguated the identity accurately.

Identity Resolution is self-correcting. The more information context you provide, the better it identifies wrong information regardless of the data feed sequence and format. Consider the situation where you have Jon Smith Jr living



## The Business Applications of Identity Resolution

at 111 First St with the phone number 111-111-1111 and another Jon Smith at the same address, but a different phone number 111-111-1211. They might be the same person with two different phone numbers or the second phone number may be a data error. However, if you add a third record, with a Jon Smith Sr and phone number 111-111-1211, you can now conclude that Jr and Sr live together and each has his own phone line. Without this self-correcting feature, databases generally grow more corrupt over time and ultimately a concerted data cleansing effort is required.

Identity Resolution is always as valid as possible for the information it has, and the more information it has the better able it is to determine who really is (or isn't) who. Furthermore, if you join together all the information that has been identified as valid from multiple sources, you assemble a much richer set of information about an individual than if you had taken information from a single source.

Identity resolution makes it possible to assemble an enriched set of information about an individual because it enables the accurate merging of information from multiple and possibly diverse sources. Without identity resolution techniques it might not be possible to accurately resolve, say, applicant information against a database of professional qualifications or even a database of criminal misdemeanors. Once you are confident you can reconcile such information sources, it becomes possible to exploit them effectively. Your potential data universe expands and your ability to use the information increases.

### **Relationship Resolution**

The idea of data enrichment quickly leads us to the possibility of unearthing hidden relationships using identity resolution technology. Consider, for example, the fairly common fraud, where an employee places orders for goods or services with a sham company that he has directly set up himself. If you use identity resolution technology to resolve employee information against information about executives of supplier companies and the technology finds a match in a supplier company, perhaps identical names or maybe just identical addresses, there is clearly cause for concern.

Using identity resolution technology to unearth hidden relationships between apparently unconnected individuals is referred to as “relationship resolution”. A clear area of application for this is in identifying suspicious relationships that

***Without identity resolution techniques it might not be possible to accurately resolve, say, applicant information against a database of professional qualifications or even a database of criminal misdemeanors.***



## The Business Applications of Identity Resolution

may indicate criminal activity or, perhaps, unacceptable conflicts of interest. In the case where someone maintains two or more identities, it is likely that records relating to these identities will have the same address or possibly the same phone number. Conversely, a common address or phone number might indicate two people living in the same abode – but that in itself may indicate a suspicious relationship, if one is a known insurance fraudster and the other is someone who has just submitted an insurance claim. Data items that may identify hidden relationships include:

- Name (relatives, spouses, or even ex-spouses)
- Address (living at the same abode or a shared PO Box)
- Phone number (shared)
- Credit Cards or Bank Accounts (used jointly)
- Membership of Organizations

There are also “upside” areas of application. Using other data items, it is possible to identify customer target groups for cross-selling and up-selling. Sophisticated customer segmentation is possible using a whole host of different attributes.

An optional feature of the IBM’s Relationship Resolution capability, which extends the same techniques, is IBM Degrees of Separation. This can recognize affinity groups from common or related data items. It can identify customers, employees, suppliers, third party involved persons and subgroups within such groups. This is an extensive capability, which can evaluate relationships with up to thirty degrees of separation. The number of degrees determines the “sensitivity” of the results.

### **Anonymous Resolution and Privacy**

With ID Resolution technology, there is an opportunity for law enforcement agencies and all organizations that wish to reduce the risk of fraudulent activity by customers or staff, to assemble rich sets of data from multiple sources and more accurately identify criminals or potential criminals. However there are practical and ethical issues that can constrain data sharing.

Organizations that have built up large databases of person-related information want to prevent the theft or illegal use of their data. But if such data is to be analyzed along with data from other sources it is difficult to think of any effective way of keeping it secure. The security of the data will also be a concern when the data provider is under legal or ethical obligations to

***With ID Resolution technology, there is an opportunity for law enforcement agencies and all organizations that wish to reduce the risk of fraudulent activity by customers or staff, to assemble rich sets of data from multiple sources and more accurately identify criminals or potential criminals.***



## The Business Applications of Identity Resolution

guarantee the security and privacy of the data beyond its legitimate area of use. These constraints have, for many years, acted as a barrier to organizations making more effective business use of the data they own and manage. At first sight, this might seem to be an irresolvable Catch 22, but surprisingly, it is not. There are clever one-way hashing mechanisms, which can both protect the data from being read and yet allow it to be used in conjunction with other data. Names and other matching data can be hashed in such a way that similar names (say Smyth and Smith) have similar hashed codes. In other words, the data can be hidden in a way that does not interfere with either identity resolution or relationship resolution logic.

The sharing of personal data in this secure way is of particular use in situations where the data being shared is highly sensitive. Consider for example security forces from several different countries sharing data on terrorist suspects. Each security organization will doubtless want to keep its information completely secure and yet it will be important to allow these multiple databases to match information on known terrorist suspects or to carry out relationship resolution. With anonymous resolution this can be done and information shared without providing any access to information that is not intended to be shared.

Another example of the need for this functionality occurs with mergers and acquisitions, where due-diligence processes can require determining how many customers two companies have in common or whether there are compliance issues (for example, does the company being acquired do business with undesirables). This technology makes it possible to answer such questions without the need to reveal specific data.

### The IBM Entity Analytic Solutions

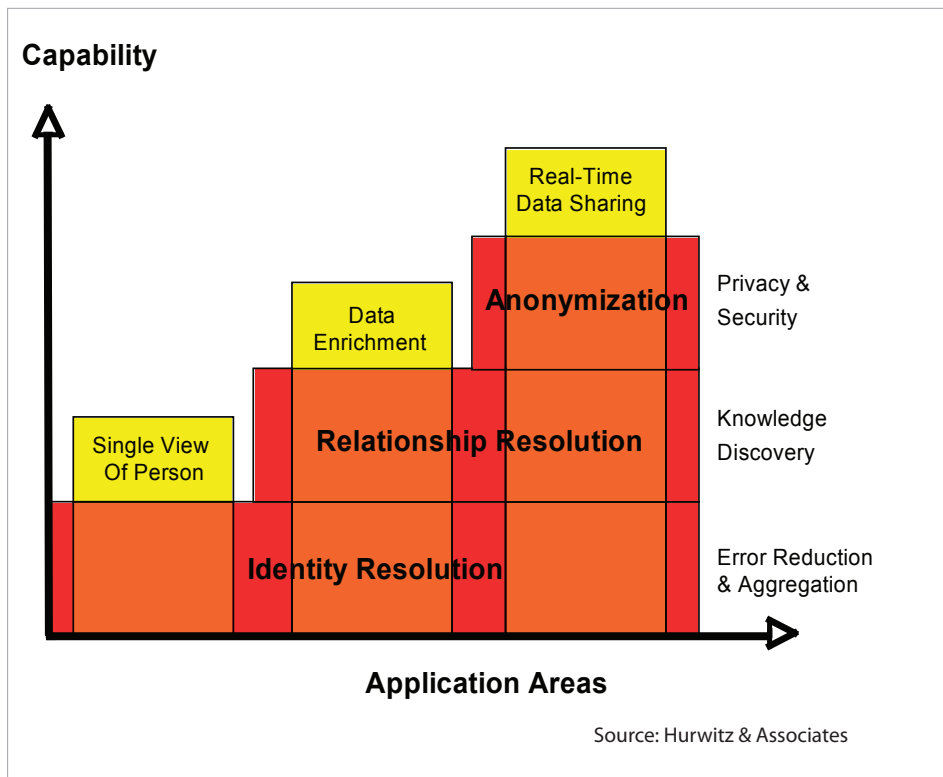
IBM refers to the software technologies and techniques we have described above as Entity Analytics and has unique technology which provides all of the capabilities we have discussed; Identity Resolution, Relationship Resolution and Anonymization.

As a software product, IBM's Entity Analytic Solutions (EAS) is unusual because it is built to operate as a service that can be used by any application. In order to implement it, it is only necessary to load the software and identify all sources of personal information that are intended to be used. From that point on, any application can pass requests to the EAS system and it will be process the requests against its data sources and provide answers.

***The sharing of personal data in this secure way is of particular use in situations where the data being shared is highly sensitive.***

## The Business Applications of Identity Resolution

The IBM EAS is a turnkey capability that can be used by any application once it has been loaded and configured. It has been built to operate at on-line speeds so that the processing it carries out can be included in on-line programs running within the organization or over the web.



**Figure 1. IBM EAS Application Spectrum**

Figure 1 illustrates the spectrum of application areas for IBM EAS. The technology embodies three distinct capabilities; Identity Resolution which aggregates identity information, reducing errors and information conflicts as it does so, Relationship Resolution which enables the discovery of relationship information that is hidden within the data and Anonymous Resolution, which provides privacy and security without reducing other capabilities. These capabilities can be used in a variety of different ways, Identity Resolution providing a fundamental capability that both Relationship Resolution and Anonymization enhance and extend. We can classify the areas of application as being providing a single view of person, data enrichment and the exploitation of new data resources (data sharing).



## The Business Applications of Identity Resolution

### ***The Single View of Person***

IBM EAS dynamically creates a single data view of a person from multiple data sources within the organization when requested. EAS provides a full attribution capability where it never discards information. Every identity and attribute is linked, preserved, and consolidated. So, this single data view includes a complete history of data sources and real-time updates to them. It can be used by applications of any kind to reduce ambiguity or errors that may result from confused or inaccurate name information. The accuracy of the single view that is built increases as more data sources are used, so there is no reason not to use every data source available.

It is important to note that the identity resolution capability is not intended to be a data cleansing capability since it does not presume or assume that any specific form of a name is “correct” and the software makes no attempt to update any of the personal information it accesses. It validates personal data or removes data inaccuracies in real-time and in doing so arrives at a reliable and up-to-date single view of a person’s data. It also naturally carries out a data audit, revealing the level of data corruption or inaccuracy, which exists within the various personal data stores that the organization owns and maintains. It can thus feed information to data cleansing systems – or pull information from such systems.

### ***Data Enrichment***

Identity Resolution enriches data by enabling sources of personal data to be aggregated. Relationship Resolution provides further enrichment by making it possible to discover previously unidentified relationships in the collection of data sources being examined.

Again, it is important to understand that this is a real-time capability. The discovery of relationship information, whether to identify possible criminal connections or to link together customers in a CRM system means that it can be acted on immediately.

### ***Anonymization***

Anonymization dramatically opens up the potential pools of data for an organization to exploit. Indeed, as the use of IBM EAS increases, it is quite likely that it will stimulate the commercial market for data sharing considerably, simply because it guarantees the security of the data. A company that shares its data using this technology would be confident that the data was safe and also able to prove that it wasn’t compromised by any data sharing activity it was involved in.

***...[the identity resolution capability] validates personal data or removes data inaccuracies in real-time and in doing so arrives at a reliable and up-to-date single view of a person’s data.***

***Anonymization dramatically opens up the potential pools of data for an organization to exploit.***



## The Business Applications of Identity Resolution

If we think of Identity Resolution as providing a single view of person, then Anonymization makes it possible to build a far more comprehensive view of person or alternatively to build very focused views of person, by leveraging external data sources that deal with specific aspects of an individual's activity (medical information, educational information, employment history, etc.). Similarly, Anonymization extends the amount of information that can be gathered from applying Relationship Resolution.

The fact that IBM EAS provides a real-time capability also means that organizations are able to try out different data sources to determine what value they may add.

### IBM EAS and Scalability

The full identity resolution process consists of 4 phases.

- Name and Address Standardization
- Normalization
- Data Enhancement
- Identity Resolution

IBM EAS first converts name and address into a standardized form and normalizes the information, merging identical identities. It then enhances the data, accumulating details from different sources. Finally, it applies all the rules that it can to resolve identities. If Anonymization is turned on, it carries the same process out on encrypted values.

The software has been designed so that this process can be carried out on data streams sent from multiple sources and run in parallel on multiple computers, each accessing different large data sources, all with a response time of no more than two seconds. It is built for scalability and to operate in a real-time environment.

***Anonymization extends the amount of information that can be gathered from applying Relationship Resolution.***



## The Business Applications of Identity Resolution

***The USA PATRIOT Act.*** In response to the terrorist attacks on 9/11, the USA PATRIOT Act was passed. It gives the federal government more authority to track and intercept communications and it also gives the Secretary of the Treasury power to combat money laundering through US financial institutions. Section 314(a) of the PATRIOT Act allows law enforcement to communicate with financial institutions directly about suspected money laundering via 314(a) requests. These requests come via the Financial Crimes Enforcement Network (FinCEN) to the financial institution to determine whether the organization has done any business with anyone on the list. Typically, a financial institution might receive fifty names on a bi-weekly basis and they are asked to identify relationships they have with these potential criminals and their close associates. The government typically gives the organization a few weeks to respond. Any known terrorists or drug smugglers are put on the office of foreign assets list (OFAC).

### ***An EAS Case Study: Global Anti-Money Laundering***

Hurwitz & Associates interviewed a vice president of a large financial institution who was responsible for generating policy and standards for the organization, in order to understand the institution's use of and experience with IBM EAS. One of the vice president's areas of activity was in global anti-money laundering. The company had begun to address the issue of managing and investigating identities for account opening and suspicious financial activity even before the PATRIOT Act gave them an immediate impetus to accurately identify people and organizations.

This global enterprise knew that they needed a powerful resolution engine to deal with the identity issue. The vice president had previous experience in this area and was well versed in issues regarding name matching. These included normalization as well as a number of issues pertaining to matching international names and characters. For example, Arabic and Asian names have certain characters that are complex and often rendered differently in the western alphabet. So, a person with an Asian passport may have one rendition of a name, but when they travel to another country, a character is recognized as something completely different. Simple pattern matching, such as recognizing similar name spellings, was not going to solve their problem.



## The Business Applications of Identity Resolution

Historically, a type of inquiry such as a 314(a) would set off a flurry of email activity to people around the company – in effect, around the world - asking to see if this name appeared in any of their systems. There was no fool-proof way to ensure that these people were actually checking their data sources for these names, or for that matter, checking them effectively. According to the vice president, “[Before EAS] At the end of the day, we wouldn’t know if a clerk had actually looked through the list and we wouldn’t know if we were just getting exact string matches or something more sophisticated. With EAS, we now have the power and expanse of the identity analytics piece.”

The financial institution has recently completed phase 1 of the project. In phase 1, the goal was to be able to respond to requests, such as the 314(a) with the highest quality entity resolution. The institution uses its own data source as well as sources like World Check to find close associates with the named entity. In this phase, the organization did not make use of the relationship resolution capabilities of EAS – although it is likely to implement this in the future.

To date, the company has seen that they can replicate what they did under their old, manual process with little difficulty. Additionally, they have revealed information that they did not find using their previous methods. In essence, they are carrying out the same process, but doing it much faster with much more consistent, reliable, and insightful output.

In the next phase the company plans to add additional customer data to the system to accomplish the following objectives:

- Identify more suspicious activity and report it to the database.
- Incorporate their case management system into the mix. A fraud investigator, for example, could then use the case management system to analyze suspicious activity. The company plans to use EAS to determine if there is overlap between what different groups in the financial institution may be investigating.
- Utilize EAS more thoroughly for their account opening system. For example, if they are processing 1,000 new accounts at night, they can run these accounts past EAS to determine whether anyone is on a “bad list” or whether there is some suspicious activity associated with any person.

The VP we spoke with has little doubt that these activities will help decrease risk as well as lower costs associated with case management and processing new accounts.



## The Business Applications of Identity Resolution

### **About Hurwitz & Associates**

Hurwitz & Associates is a consulting, research and analyst firm that focuses on the customer benefits derived when advanced and emerging software technologies are implemented to solve pragmatic business problems. The firm's research concentrates on understanding the business value of software technologies, such as Service Oriented Architecture and Web services, and how they are successfully implemented within highly distributed computing environments. Additional information on Hurwitz & Associates can be found at [www.hurwitz.com](http://www.hurwitz.com).

Robin Bloor is a Partner of Hurwitz & Associates and the founder of Bloor Research.  
Dr. Fern Halper is a Partner at Hurwitz & Associates.