



# Solving the Address Search & Matching Problem for Insurance Fraud Application

An overview of the identity search & matching problem space, and an introduction to Identity Systems technology

@2004-2005 Identity Systems, an Intellisync Company. All rights reserved. All logos, brand and product names are or may be trademarks of their respective owners

## Introduction

A primary focus in the fight against Insurance fraud is in the areas of non-disclosure in policy applications and in claims processing to reduce the "one incident, multiple claims" or "one identity/one address, multiple claims" scenarios.

In this fight, Insurance Fraud Investigation and Policy Application Systems depend upon data about the names, addresses and other identification attributes of the people and organizations involved in policy applications and insurance claims, including applicants, the insured, payees, claimants, physicians, lawyers, witnesses, agents, employees and other parties.

Historical claims data is typically loaded into large-scale claims repositories. This data is then made available to Insurance Fraud Investigators for case-led inquiries and claims audits; to Claims Processors for claims screening; to Policy Application systems for pre-screening new applicants; and to Data Mining applications for claims analysis (e.g. grouping historical claims by address or identity). In addition, data will often be kept about historical policy applications.

It is a fact of life that all such identification data suffers from unavoidable error and variation. It is normal that spelling and typing errors occur, nicknames and abbreviations are used, and words are missing and out of order. Often the entity committing the fraud or perpetrating the crime is in fact trying to defeat existing matching algorithms by subjecting the identification data to deliberate, abnormal or extreme variation. Despite this error and variation, Insurance Fraud investigators must be able to search, match, link and examine pieces of information from multiple sources, both internal and external, to discover connections that would otherwise remain hidden. Policy application systems must be able to thoroughly search claims history and previous application data to determine if the organization has had prior dealings with this identity.

There may also be a need to screen policy applications for compliance reasons, against national watch and alert lists, such as OFAC (Office of Foreign Assets Control) and the DPL (Denied Persons List). Underwriters, brokers, agents, primary insurers, and others are prohibited from engaging in transactions that involve various OFAC blocked countries and entities. These types of lists require extremely thorough screening, especially for commercial transactions where the penalties can be high.

All of this identity data requires sophisticated indexing that performs well regardless of the quality, format or country of origin of the data. It must be possible to reliably search or mine the large-scale databases using the names of people and companies, as well as their addresses and other identity data.

The technology that supports such searching and matching must be able to ensure candidates are found despite the unavoidable or deliberate variation and error in name, address and other identity data. The technology must cope well with data of any quality and completeness, as the source of such data can vary greatly in reliability. Such search technology must not require that the data be cleaned or formatted, for reasons that include:

- The data may be not legally changed without approval of the customer or source organization.
- Statistical techniques for enhancing data are "good for statistics" but introduce error that can be destructive for matching.
- Many cleaning techniques are not reversible e.g. changing Bobby to Robert; changing St to Street when it is possible that it could be Saint.
- Users believe that the transformed data is true and base decisions on it.
- Rejecting invalid data, which simply means it can not be used for any purpose and all business value of that data is lost.
- It is also highly desirable, and often essential, that the search technology be able to search and match data from any country, and potentially in any character set.

Some solutions to the searching and matching requirements of such systems require skilled investigators who know when and how to vary a search or change the search data to cause the system to work more successfully. Boolean based and wild-card searches are examples of these. A far better solution uses automated search strategies that satisfy all permutations and variations of the search.....the real solution needs to be designed to find all the candidates regardless of the way the search data was entered, regardless of the quality of the data stored in the database, and regardless of the **experience of the user.** Such search strategies must of course provide real-time searching of all name and identity data. On-line usage must satisfy the investigators and source systems need for fast response without any loss of quality of search, despite the quality of the data.

Another aspect of this problem area is the ranking of the results returned to the searcher. While diligent investigators can use sophisticated search tools well, it is not possible for the average user to spend day after day simply browsing historical data and do a good job selecting candidate matches; even the diligent user can get ineffectual at the job if it is a continuous activity. **To better automate** the searching, matching and screening process it is necessary that computer systems are designed to "mimic" the very best users when choosing amongst the possible matches. In the same way as human operators' use names, addresses, dates, identity numbers and other data, the system must be able to use matching algorithms that effectively rank, score or eliminate the candidates.

The volume of data that is today available to these systems is growing explosively. Today it is time to invest in the core objective of these systems; that is to make sure that the highly valuable data that is stored in these systems can in fact be found, despite its error and variation. Similarly the value of high-end tools of investigation that provide "link analysis", "data clustering", or "visualization" can be significantly improved if they make use of the very best search and matching algorithms.

# **Identity Systems' Approach**

The ideal identity search solution:

- Overcomes the error and variation in the data
- Maintains performance even in extreme volume situations
- Finds all valid candidates without generating false matches.

#### This implies:

Intelligent and scalable **algorithms**, which, through the use of rich keys and search strategies, return all of the candidates an expert user would consider as being the same as the search criteria.

These algorithms must be able to cope with data from the real world. This includes data that is not formatted or cleaned or which contains missing, extra, truncated, out of order, non standard, or noise characters/words, initials, abbreviations, nicknames, numbers, codes, and concatenations. The algorithms need a customizable rule base to incorporate the knowledge of the expert user, and a default population rule base in the case where the user is not that experienced.

The algorithms require phonetic and orthographic correction functionality, to address spelling and typing errors.

Intelligent matching routines must be available and tuned to mimic the expert user making a choice as to which candidates are the correct matches. Such matching routines need to take into account all of the error and variation in the identities' attributes, as well as weighting the attributes as the user would.

The Algorithms must work well regardless of the country of origin and language of the data and must insulate the application developer from the differences between country and language.

Identity Systems software is designed and built on these premises and is being used by over 500 commercial and governmental organizations around the world.

# Partial List of Identity Systems Customers using the technology to fight Insurance Fraud

Insurance Services Office (USA) Insurance Bureau of Canada Insurance Council of New Zealand The Hartford Insurance CSC Financial Services (@First<sup>™</sup> and PatriotProtector<sup>™</sup> products & services) NRMA Insurance (Australia) VIC Transport Accident Commission (Australia) Workplace Safety & Insurance Board (Canada)

#### **Other USA Users include:**

Bank of New York Internal Revenue Service (IRS) Sprint General Motors CitiCorp VISA Equifax Experian General Reinsurance Insurance Services Office The Hartford Cendant (Resort Condominiums International) L. L. Bean State of California (several systems)

#### **Other International Users include:**

Canadian Dept of Justice Australian Bureau of Criminal Intelligence Brønnøysund (Norway; Government)) Bausparkasse Schwabisch (Germany; Bank) Office of the Narcotics Control Board -Thailand Australian Customs Service Benkar Tuketici Finansmani (Turkey; Credit Info) New Zealand Immigration Service New Zealand Police Hong Kong Customs UK Home Office - Offenders Index Institute of Medical & Veterinary Science (Australia) Czechoslovakia Security - Information Services

## **Identity Systems Products**

There are three products:

- Identity Search Server (ISS): Online / Batch
- **SSA-NAME3:** Core SDK/API (included in other products)
- Data Clustering Engine (DCE): Batch

All these products can search, match, and rank identity **data of any quality or format** and do not require any "cleaning" or "scrubbing" of the source data nor is the source modified in any way by Identity Systems products. **All countries/languages** are supported, both singularly and in combination. All products can be used to centrally index and simultaneously search, match, and group identity data from disparate sources such as from Current Sales, Legacy, and External databases. An overview of each product follows.

**The Identity Search Server** (ISS) is the product that provides **on-line and batch searching, matching and duplicate discovery** for all types of identification data stored in relational databases (**DB2/UDB, Oracle, and SQL/Server**) on OS/390, NT, AIX, HP/UX, Linux, & Solaris platforms. High-performance indexes are automatically maintained **without changes to existing application programs**. Custom search clients can be developed using an easy-to-use API. ISS uses an **n-tier** architecture, making it the ideal solution for organizations implementing the latest technologies, including web-based applications. With ISS, the capability to **centrally** index identity information from **disparate** source tables, databases, and computers is particularly advantageous since the central index can then be used to search, match, rank, group, and maintain all the data **simultaneously** and in **real time**.



Customizable Identity Search Server™

**SSA-NAME3** is Identity Systems **core technology** and included in all the other Identity Systems products. It is a system development kit that enables organizations to build business application programs to **search, match, rank, and analyze** records about people, companies, products, addresses and various other identity data. It can be executed on most any platform and operate on data stored in any database. The software contains algorithms for computing 'fuzzy' search keys and strategies, as well as algorithms for complex matching of identity data.



**The Data Clustering Engine** (DCE) is a **stand-alone**, **batch** data **grouping and investigation engine** for all forms of identification data and can execute on AIX, HP/UX, Linux, Windows NT, Solaris, & Compaq Tru-64 Unix platforms. The DCE performs a thorough analysis of the relationships between people, companies, addresses, products and other entities. This powerful software is used by numerous leading corporations and institutions for **de-duplication**, **file matching**, **householding**, **fraud investigation**, and various other analyses. It **does not require programming** and is highly scalable for the processing of large and extreme data volumes.



## **Identity Systems**

Identity Systems, an Intellisync Company, is the pioneer in enabling organizations to build and maintain high-quality identity data search and matching software solutions. Identity Systems was formerly known as Search Software America (SSA) and personnel have been in this area of specialization for more than 19 years. Identity Systems now has over 500 clients worldwide who rely on its robust enterprisewide software. The Company has offices in the United States, the United Kingdom, Australia and Brazil.

## **Company History**

Identity Systems was originally established in Australia in 1987 with the purpose of formally developing and marketing its founders experience in building identity search and matching systems. This experience had shown that, while the significance of identity matching in different systems varies considerably, there has always been the same basic set of concerns:

- Performance problems arise because the most frequently occurring names are also usually those upon which searches are most often performed.
- Traditional phonetic algorithms create poor response time and prevent the user from locating a match among many candidates. Traditional match-codes fail to find matches when data is invalid or incorrectly parsed.
- True phonetics is only a subset of the errors in names, addresses and other identity data.

Some of our projects emphasized the need to quickly achieve a match, if there was one. Others placed their emphasis on proving that there was no match at all. One project presented the unusual opportunity to empirically develop and modify an algorithm designed to solve phonetic, orthographic, and Anglicization problems in more than 2,000,000 hand-written credit records. During the project development activity, some 300,000 computerized matches were compared to manual matches done by expert searchers. Whenever the searcher found data not found by the system, the algorithm was refined.

Another project involved the re-processing of 25 million records of international data from virtually every country in the world, where it was known that at least 99% of the records were in fact pairs about the same person. The project also demanded a performance breakthrough since the design objective was to develop an identity search system to handle over 30,000 inserts per day and to support a 50,000,000 record on-line database. Based on the fact that the project's purpose was to identify the records that were not pairs and the population was smaller than the error rate in the data, the successful solution required considerable research.

The experience gained from such projects and Identity Systems work with customers around the world for the last 19 years, has been incorporated into ongoing development and maintenance of Identity Systems product suite. Although the identity search and matching problem may never be completely solved, Identity Systems enables organizations to reach unparalleled levels of quality and performance.

## **Contact Us**

Visit us on the web at <u>www.identitysystems.com</u> or at the contacts below:

#### **USA, North and South America**

USA Sales <u>USASales@identitysystems.com</u> tel. +(203) 698 2399 Identity Systems 1445 East Putnam Avenue Old Greenwich CT 06870, USA

#### **UK and Europe**

UK Sales UKSales@identitysystems.com tel. +44 (118) 944 9688 Identity Systems 9 Headley Road Woodley READING RG5-4JB, UK

#### Australia and Asia

Australia Sales <u>AUSSales@identitysystems.com</u> tel. +61(02) 8904 9936 Identity Systems Suite 1, Level 3 194 Miller Street North Sydney NSW 2060, Australia

## Argentina

Argentina Sales <u>eguiterman@leviminond.com.ar</u> tel. +54(11) 343 2005 Levi Minond Estudio LmyA, Piedras 575 piso 1 "D" (1070) Buenos Aires, Argentina

## Brazil

Brazil Sales <u>USASales@identitysystems.com</u> tel. +0800 891 7052

#### France

France Sales FRSales@identitysystems.com tel. +33(08)10 85 93 62

#### India

India Sales <u>sanjayb@identlogic.com</u> tel. +91(22) 2496 0033 Identlogic Systems Hansraj Pragji Building 83C Moses Road Worli, Mumbai 400018, India

#### Japan

Japan Sales <u>marai@intellisync.com</u> tel. +81(3) 3511 0600 Hakuyo Building, 5F 3-10 Nibancho Chiyoda-ku, Tokyo 102-0084, Japan

## South Africa

South Africa Sales <u>sam@bateleur.co.za</u> tel. +27(11) 691 1600 Bateleur Software Solutions Homestead Park, Block A 65 Homestead Ave Bryanston, Sandton Gauteng 2120, South Africa

#### Turkey

Sales info@komtas.com tel. +90(212) 288 47 30 Komtas Yildiz Posta Caddesi Dedeman Ishani Kat 10 No.48/23 Esentepe 34349 Istabul