

Align Journal Online Exclusives

Resolving Identities Improves the Bottom Line by Robert G. Barker

November 14, 2007

The aftermath of the events of 9/11 has impacted many facets of life in the U.S. While numerous articles on the effect on the collective psyche have been written, 9/11's role in driving new technologies and affording more precise determination of identities in both commercial and government settings has begun to draw increasing attention.

R&D funding in a new technology space called identity resolution rapidly increased in the first few months following the destruction of the Trade Center and Pentagon. Elements of what eventually became the Department of Homeland Security (DHS) began a determined search to discover technologies that could prevent a recurrence. In April 2002, the Federal Aviation Administration (FAA), which later became a part of DHS, initiated a formal call for proposals from companies possessing technology that could improve the screening of airline passengers for potential terrorists. Eventually, more than 100 technology companies stepped forward to compete in the DHS Request for Proposal (RFP).

Although 9/11 was the catalyst for the emergence of identity resolution solutions for government, analogous challenges in knowing "who's who" and "who knows whom" have existed for years in commercial markets (e.g., retail and insurance). According to an annual survey conducted by the University of Florida, the average retailer loses an estimated 1.5 to 2.0 percent of their revenues each year due to customer, vendor, and employee fraud with 2006 total losses tagged at more than \$40 million (see "2006 National Retail Fraud Survey," Richard Hollinger and Lynn Langton, University of Florida, 2007). The Insurance Information Institute estimates that fraud accounts for 10 percent of the property and casualty insurance industry's incurred losses and loss adjustment expenses, with an estimated \$30 billion lost in both 2004 and 2005 (see Insurance Information Institute Website home page: www.iii.org/media/facts/statsbyissue/fraud/). Other industries (e.g., banking) face similar attacks upon their profitability.

Commercial losses from deceptive activities impose a non-governmental "fraud tax" each year on every household in the country. Calculating the fraud tax attributable only to retail and property/casualty insurance deception yields an average of more than \$600 per U.S. household that's lost each year through fraudulent activities and recovered through higher prices for goods and services. Adding similar losses in other industries would drive this fraud tax number even higher.

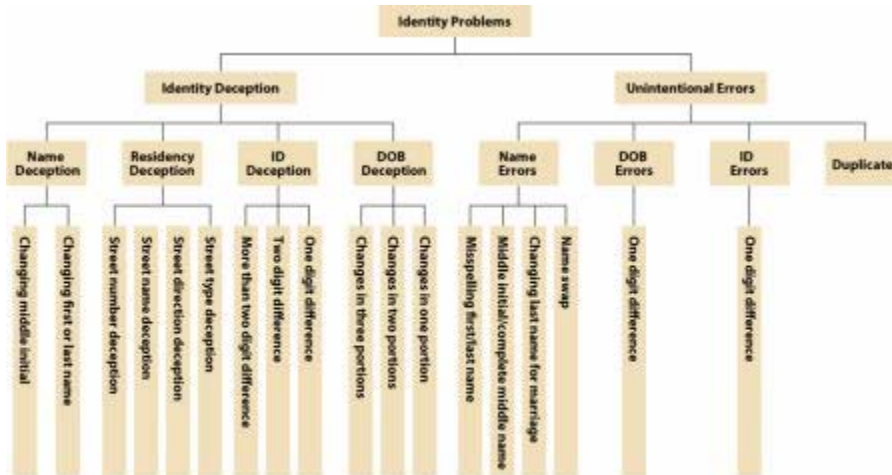
The price paid for fraud impacts all of us: higher prices for retail goods, higher insurance premiums, or greater risk for airplane travel. How can a commercial organization protect itself against such fraudulent activities? The answer isn't simple.

Guarding against fraud requires knowing the true identity of customers, vendors, employees, and other entities. It's ironic that most organizations already possess the internal data needed to detect potential and real fraud. What prevents an easy solution? Obstacles arise because the required data repositories were created for a single purpose and weren't designed to be combined with other data in other locations and other formats.

Attacking the identity problem requires cooperation between groups that may not have worked closely in the past. In particular, Line of Business (LOB) managers and IT experts must join forces to break down barriers that prevent the effective use of existing internal data sources, and senior management must support and track these efforts to ensure success.

What Is Identity Resolution?

Many variations of identity problems exist. The chart below, from a University of Arizona study, depicts the authors' taxonomy of the most common problems that exist regarding identities of individuals. Few existing systems can effectively handle these types of ambiguities, which usually result in creating multiple records that are treated as different individuals. So business rules based on an individual aren't properly handled, resulting in waste and opening the door for those who wish to deceive.



Source: G. Alan Wang, Homa Atabakhsh, Tim Petersen, Hsinchun Chen, "Discovering Identity Problems: A Case Study," Dept. of Management Information Systems, University of Arizona's Artificial Intelligence Lab (www.ai.arizona.edu/), 2005.

Identity resolution is the process through which the true identity of an individual is determined based on available data. It improves decision-making by:

- Analyzing ambiguous identity information from disparate data sources
- Resolving multiple identities into one
- Discovering hidden relationships.

Other innovations can require an extensive replacement of existing solutions, but identity resolution greatly strengthens already deployed applications by augmenting them with powerful technology for searching identities. The deployment of identity resolution often requires little or minor changes to existing applications, including changes to user interfaces that could otherwise necessitate user retraining. Detecting similarities among multiple attributes (e.g., names, addresses, phone numbers), resolving multiple identities into one, and uncovering hidden relationships are all useful functions that identity resolution performs.

How Is Identity Resolution Different?

Identity resolution isn't the only technology that applies the power of data to increase operational effectiveness:

- Customer Relationship Management (CRM) systems help organizations track interactions with customers and maximize revenue and profit by optimizing sales and marketing activities.
- Customer Data Integration (CDI) and Master Data Management (MDM) merge different and often conflicting data about customers and transactions into a single authenticated master file to support a common method for all applications to retrieve information.
- Data warehousing solutions extract, cleanse, and combine information to provide an integrated repository that supports data analysis, reporting, and other operational functions.

While each of these approaches positively addresses the specific problems for which it was designed, resolving identities is a unique discipline addressing some specific needs:

Challenging and conflicting requirements:

Identity resolution is recognized as a complex, challenging discipline, separate and apart from traditional technologies. A common set of cross-industry, commercial identity resolution requirements resembles those that drive airline passenger screening:

- Identify risky individuals in less than a second
- Work unobtrusively to avoid annoying good customers
- Analyze information from multiple, disparate, and remote sources of data
- Minimize need to modify or move data
- Maintain confidentiality of all data sources while enabling their use
- Handle risky and non-risky individuals with the same system.

“Siloed” systems:

Resolving identities requires rapidly accessing multiple data sources since most organizations have data distributed across various data silos. Mergers, acquisitions, and rapid growth result in organizations having key business data stored in different silos, which makes it challenging to integrate and use that data. Identity resolution solutions should be capable of adapting to the attributes of individual data sources without requiring redesign; they should be able to analyze the information in each source and combine individual analyses to derive an aggregate answer.

High volumes of disparate data:

It often isn't feasible to move all the data needed to achieve identity resolution into a single repository, nor is it possible to transform this data into a single format. Identity resolution systems must work in real-time with multiple large data sets, each with a unique format and access protocol. Given the size of many data sources and the fact they must be combined for identity resolution, scalability of the underlying technology is paramount.

Sophisticated bad actors:

Criminals have become increasingly clever and traditional solutions haven't kept up. Most systems weren't designed to identify and handle the sophisticated ways that bad actors fool these systems. For example, if two names or two addresses are spelled slightly different, traditional systems create two different identities. Bad actors become adept at avoiding detection by finding ways to assume multiple identities and deceive these systems. Identity resolution solutions must discover this deception by searching for similarities and patterns in data and other techniques.

Maintaining privacy and confidentiality:

Resolving identities can have a strong positive impact on effective operations, but identity resolution implementation should be built around technology that protects data privacy and maintains confidentiality. Fortunately, existing identity resolution technologies can combine metadata about data sources while keeping them intact and separate.

Retail Example

Retail is just one example of an industry in which identity resolution can have a significant impact on a company's bottom line. The retail segment generates and maintains massive amounts of data. A modern retailer logs each sales transaction, maintains up-to-date inventory systems, and monitors merchandise returns. In addition, a large retailer will typically employ tens of thousands of people with a high turnover rate, resulting in the need to store a huge amount of HR data.

Suppose an unscrupulous store employee decides to supplement his income by buying quantities of merchandise at employee discounted rates. He then returns the goods for full value at another store in the chain without presenting his employee credentials, violating company policies. When

the employee is asked for identification or to provide his name and address, it's theoretically possible using available HR data to detect that he's an employee and deny the return. This is rarely done. Why not?

Data repositories evolve independently. Each one is normally defined as part of a solution to a specific problem, often with little thought being given to sharing between them (e.g., a retailer's HR system and its returns management system). Different departments own each system and its related data repository, and each one is reluctant to give up control for fear of impacting their own ability to deliver.

Even if cooperation is forthcoming, the repositories may have different data formats and may even run on different hardware platforms. Further complicating matters is the fact that existing systems aren't built with intelligence that can resolve differences between an individual's identity information introduced either through a data entry error or through the fraudulent employee's attempt to deceive.

Identity resolution software augments existing systems to overcome these obstacles, sometimes with striking results. Why shouldn't fraud be thought of as just a cost of doing business for retailers? If you consider a retailer with a relatively high profit margin of 5 percent and a low level of fraud at 1 percent, it's easy to see that preventing fraud can have a huge impact on the earnings and market valuation of the company.

Although this example focuses on retail, other market segments face similar challenges. Insurers, telecommunication carriers, data providers, bankers, and others all have existing applications that aren't equipped to handle identity ambiguity, and they all could improve the bottom line by addressing the problem.

The Impact of Identity Resolution

Widespread interest in identity resolution is driving the development of many new products for commercial and government use. WinterGreen Research of Lexington, MA recently projected the market for "identity resolution middleware" software to grow rapidly for the next several years, topping \$7 billion in 2011.

Commercial and government IT organizations are currently exploring how identity resolution can improve their operations. Finding out "who's who" and "who knows whom" is turning out to be a high priority. Repealing the fraud tax may be almost within our grasp.

ABOUT THE AUTHOR

Robert G. Barker

email: bbarker@infoglide.com

website: <http://www.infoglidesoftware.com>

Voice: 512-532-3500

Robert G. Barker is the acting CTO for Infoglide Software. He has more than 25 years of experience in software development, product management, marketing, and business development. His career includes executive positions with Compuware Corp., Sterling Commerce, and Novell. He holds a bachelor of arts degree with honors from The University of Texas at Austin and completed the first year of his Ph.D. in psychology at Princeton University before joining the software industry. He recently served on the organizing committee for the IEEE International Engineering Management Conference 2007.

© 2007 Align Journal and Thomas Communications, Inc.

All Rights Reserved. Reprinting and distribution without written permission is prohibited.