National Intelligence and the Integration Gap: A Common National Data Model

By Len Silverston

Terrorism feeds on fear and disorganization. In today's world, small groups of organized terrorists can weave through a powerful nation's defenses and cause havoc, as we have seen recently. Terrorism counts on being able to attack disconnected points in a system so that it cannot be traced. If our government is able to gather assorted data in an integrated fashion, it can counter future terrorist activity more effectively. For example, an invaluable tool for counter-terrorism would be an integrated government database that maintains suspicious profile information from a combination of government agency sources such as FBI watch-list profiles, DEA (Drug Enforcement Agency) criminal activity, NIA (National Immigration Agency) overstayed visa information, CIA public threat warnings, NSA (National Security Agency) suspicious activities, as well as other government and commercial enterprises involved in possible terrorist activity.

Is it possible to create such as database? Can government agencies (and appropriate commercial enterprises such as airlines) share such sensitive information in a collaborative fashion? If government agencies collaboratively created national profiles of people, would the public be outraged over the government having "Big Brother" profiles of people? Does national profiling raise other security issues such as creating potential gold mines of information for technology thieves? Could complete profiles of people be misused and misinterpreted? What parties should be included in a national profiling system? If it is only parties who have had suspicious activities that are linked to possible government threats, how do we define the business rules to identify who should be included within such an integrated information system?

Taking into account the above questions, many would argue that it is unwise to move toward an integrated national profiling system. My perspective is that we, as a nation, are duty bound to serve in the most effective manner possible, and only with integrated information can we make effective decisions. To say that it is too dangerous to have integrated information on people and organizations is to say that we, as a nation, cannot trust our government with valuable information and that our disparate database silos protect us from our own inadequacies of safeguarding and wisely using that information.

Integrating National Profile Information

In order to create integrated national profile information, a data model is needed that defines the appropriate and required information about people and organizations and how that data is interrelated. This data model could provide a common understanding of the data and provide a standard structure for maintaining and sharing data. While there are many data model notations (including object data models and several notations for entity-relationship diagrams), the most important point is that we adhere to standard data structures that a common data model could suggest. If our nation could agree upon standard designs for maintaining common constructs, it would become easier to share and integrate information.

How is it possible to build a common national data model when government agencies (and commercial enterprises) often have widely varied information needs? For example, the DEA maintains information such as drug listings, drug offenders, drug traffic laws, and drug incidents and crimes. The Immigration and Naturalization Service (INS) maintains information such as prospective and past immigrants, applications for immigration, visa statuses, visa violations and immigration laws. Since these agencies maintain such different types of information, does it still make sense for them to share information?

While the specific types of information varies dramatically, there are general types of information that can be shared among agencies to provide dramatic results in helping the government effectively serve and protect. Government agencies maintain a great deal of common information such as information on people, organizations, relationships, contact information, communications between people, transactions, licenses, applications, laws, programs, work effort management, government budgeting and government accounting.

A Possible Model

Figure 1 provides a possible structure that could serve as the basis for common national database integration. This universal data model is certainly not the only way to represent this common data, and there are many other variations that could be used as a standard. The purpose of this column is to illustrate that it is possible to create a common vision for integrating national data. The important point is to form agreement on terminology and data structures between various government and commercial enterprises so that data may be shared more easily.

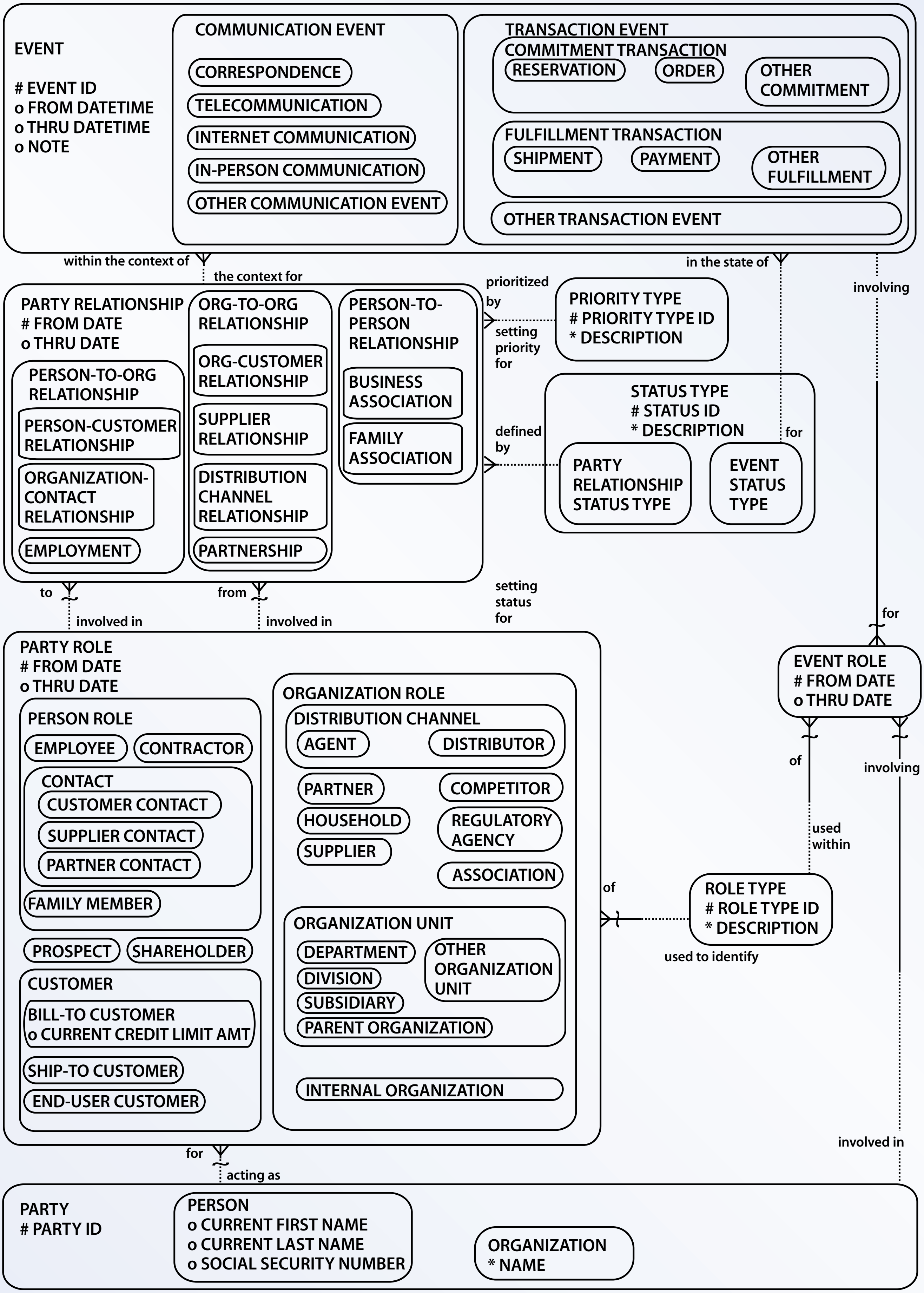

Figure 1: A Universal Data Model for Relationship Development

Data Modeling Notation:

- A crow's foot (three prongs at the end of the relationship line) indicates that there are many occurrences of the entity near the crow's foot for each entity that is not near the crow's foot. For example, each PARTY may be acting in ***one or more*** PARTY ROLES (entity names are shown in caps in this article).
- The dotted line indicates optionality (as opposed to mandatory) for each side of the relationship. Each ROLE TYPE ***may be*** (because this is a dotted part of the line) used to identify one or more PARTY ROLES. Read the other way, each PARTY ROLE ***must be*** of one and only one ROLE TYPE.
- A '#' in front of an attribute indicates that this attribute is a key. A '*' indicates that the attribute is a mandatory attribute. A 'o' before an attribute indicates that the attribute is optional.
- Boxes within boxes indicates subtypes or subentities.
- The '~' on the relationship line represents foreign-key inheritance. This means that the primary key of the entity closest to the tilde includes, as part of its key, the primary key of the entity without the crow's foot.

Figure 1 shows four main entities that form the essence of the information needed to profile people or organizations:

The PARTY entity maintains information about the PERSON or ORGANIZATION such as names, social security numbers, demographics and other information that is associated with a person or organization, independent of the role(s) that they play.

The PARTY ROLE entity maintains information associated with each role that a PERSON or ORGANIZATION plays. This data model shows that a PERSON may be identified by several roles, for example, they may be a WATCH-LIST SUSPECT, a CRIMINAL FELON, a member of a certain organization the government is tracking (ORGANIZATION MEMBER), a recent IMMIGRANT and/or any other role that is important to flag in providing an overall profile of a person. There may be information related to each role such as the immigration status for an IMMIGRANT or the ranking (of how dangerous someone is) for a WATCH-LIST SUSPECT. Similarly, ORGANIZATIONs may be involved in many roles that can provide a more complete profile for each organization.

The PARTY RELATIONSHIP entity shows the affiliations or associations that people and organizations have within the context of the various roles that they play. There are three types of possible relationships: relationships that people have within various organizations (PERSON TO ORG RELATIONSHIP), relationships that people have with other people (PERSON TO PERSON RELATIONSHIP), and relationships that organizations have with other organizations (ORG TO ORG RELATIONSHIP). The model shows some possible relationship subtypes including: the relationship from a watch-list suspect to the agency that maintains that watch list (GOVERNMENT AGENCY WATCH-LIST SUSPECT), a suspected terrorist's relationship to the terrorist organization (TERRORIST AFFILIATION), a person's relationship to a nation (NATIONAL AFFILIATION), a person's membership within a (suspicious) organization (ORGANIZATION MEMBERSHIP), a person's relationship with a business associate or family member (BUSINESS ASSOCIATION/FAMILY ASSOCIATION) and association or rollup structures (subsidiary/parent) that one organization has with another organization (ORGANIZATION ASSOCIATION/ ORGANIZATION STRUCTURE). Having access to a more complete picture of each person's and/or organization's relationships can be instrumental in identifying suspicious individuals or organizations that may be harmful to national safety.

The EVENT entity maintains various activities that occurred within the context of each PARTY RELATIONSHIP. This can be used to track key COMMUNICATION EVENTs (i.e., correspondence, phone calls, e-mails, meetings or other types of communications that occurred between two parties) or TRANSACTION EVENTs (funds transfers, airline travel, package deliveries, arrests, criminal activities or suspicious transactions) that could be analyzed to gain insight on potential future harmful events. The EVENT ROLE entity maintains the various PARTIES involved in each captured event and ROLE TYPE that each party plays in the event. For example, "John Smith" may have played the event role type of "package sender" in a TRANSPORTATION TRANSACTION event that was flagged as suspicious and therefore maintained within this data structure.

There is a tremendous amount of valuable information that could be combined and maintained with this model. For practicality reasons, the data within the model should be limited by business rules determining what types of people, organizations, roles, relationships and events are needed to analyze national information.

There are many more entities and attributes associated with this model that are not shown in Figure 1 and to find out more details about additional entities, attributes, and database objects, please refer to The Data Model Resource Book, Volume 1 (Wiley, 2001).

The Challenges of Maintaining National Profiles

While a data model is a critical component of integrating national information, there are a great number of other issues involved in gathering national profile information. Following is a brief explanation of some of the potential challenges.

Is safeguarding our homeland (as well as other parts of the world) more important than safeguarding our information privacy rights? There are numerous privacy laws (i.e., Gramm, Leach and Bliley Act, Health Information Privacy Accountability Act) that protect human privacy rights, set standards for acceptable data transfers and require notification and acceptance regarding collecting certain information. In addition to these laws, many people feel that the government should not be entitled to private information on people. The national data model could only include suspicious individuals who represent a threat; however, who is to make the determination regarding whom is considered suspicious and how is that determination made?

Can we trust our government and/or commercial enterprise (such as airlines) to have more integrated information on people and organizations without misusing this information? There is no shortage of examples of misuses of information such as Watergate, Clinton's illegal summoning of FBI information about Kathleen Willey,[1] inappropriate commercial selling of private information, countless cases of credit card fraud, stolen identity stories and many corrupt legal convictions where innocent victims have been fraudulently convicted of crimes based upon erroneous information. Note that a great deal of information about parties is already available. It is just not available as quickly or as easily to be practical because it is much more difficult to gather and reconcile the various pieces of information about a person scattered throughout disconnected and inconsistent data sources.

How does the integration process know that a person's or organization's records from different databases are actually the same party and belong in the same profile? The names, addresses, contact information and other identification information are often entered inconsistently. There are many challenges in correctly matching, combining and reconciling records from different databases and bringing them into a common database structure. While the problem is complex, there are numerous solutions designed to perform pattern matching using probabilistic algorithms (based upon the data, the software provides probabilities that the disparate records actually represent the same party). Another solution provides integrated databases of personal information extracted from many sources using approximate string matching algorithms to combine disparate records. The cover of the book, Authentication: From Passwords to Public Keys by Richard E. Smith[2] illustrates the difficulty of identifying people when it shows a cartoon of two dogs interacting on the Internet as they exclaim "On the Internet, no one knows you're a dog."

There are numerous technology challenges and possibilities for maintaining common profiles including smart card systems, virtual schemas (as opposed to a common physical database), XML schema standards and biometric identification systems such as retina scanners, fingerprint identification systems, face image matching systems and x-ray devices designed to anatomically identify people.

Do government agencies (and/or commercial enterprises such as airlines and transportation companies) want to share their information? As pointed out in the last article in this column, "Terrorism: A Call for Integration," if people and/or organizations do not want to share, information will not be shared.

Conclusion

With the current barriers and political agendas in place between various government agencies (as well as between commercial enterprises), I believe that we are not going to completely integrate people and organizational information from all sources any time soon. However, it would help to make incremental steps towards the goal of having integrated information more easily accessible. We need to prioritize and choose subsets of critical data that, if integrated, could make a difference for our nation. For example, if FBI watch-list information and CIA suspicious activity information were combined and available throughout various government and commercial enterprises (such as airlines), perhaps it could help immensely in detecting potential terrorist activity. In order to facilitate information sharing, we need a common national data model, not necessarily the exact model in this article, but any variation of it that captures the key information needed to serve and protect our country and that is widely accepted on a national level.

These times call for cooperation. Cooperation towards integration could very well prevent our disintegration.

References:

1. Yost, Peter. "Clinton Found To Violate Privacy." Washington Associated Press. March 30, 2000.

2. Smith, Richard E. Authentication: From Passwords to Public Keys. Addison-Wesley Publishing Co. October 1, 2001.

Len Silverston is a data management consultant with more than 20 years of experience in helping enterprises integrate data. Silverston is the author of the best-selling The Data Model Resource Book series (Wiley, 2001, http://silverston.wiley.com), which describes more than 230 integrated, reusable generic and industry-specific data models. He has developed extensive software versions of these data models, some of which are now licensed worldwide by Microsoft and some that are available for licensing directly. Silverston's company, Universal Data Models, www.universaldatamodels.com), provides consulting, training and software to jump start data modeling and data warehouse design efforts while increasing design quality and facilitating data integration. He can be reached at lsilverston@univdata.com.