Building a Terrorism Ontology

Aaron Mannes, Jennifer Golbeck
MINDSWAP
University of Maryland, College Park
awmannes@comcast.net, golbeck@cs.umd.edu

Abstract. The Semantic Web could be a crucial tool to terrorism researchers, but to achieve this potential an accessible and flexible but comprehensive ontology needs to be designed to describe terrorist activity. Terrorist events are complicated phenomena that involve a large variety of situations and relationships. This paper addresses some of the issues the authors have encountered trying to build such an ontology – particularly how to describe sequences of events and the social networks that underpin terrorist organizations.

1 Semantic Web Portals

1.1 Background

Traditional web portals are websites that collect information and links to pages, usually with a common theme or topic. A Semantic Web portal has a slightly different function. Since everything on the Semantic Web is identified by a URI, the notion of linking to files as it is done in hypertext does not translate. Instead, Semantic Web portals collect URIs of files on the Semantic Web, and allow users to interact with the RDF graph of the statements.

In the context of creating a Semantic Web Portal for terrorism, any user would have the ability to submit RDF or the URIs of documents with data.

Using ontologies, the portal can combine statements from multiple files into a single model. Among the implications, this means that users can select sets of statements that reflect their personal interests, even if no one else has had that specific focus. Mapping between concepts to connect items as equivalent also allows statements to be merged into a single model.

Figure 1 illustrates a sample page from a Semantic Web portal. It takes a knowledge model written in RDF and OWL, as described above, and presents it in a coherent way. Section 2 goes further into depth about how this technology is used.

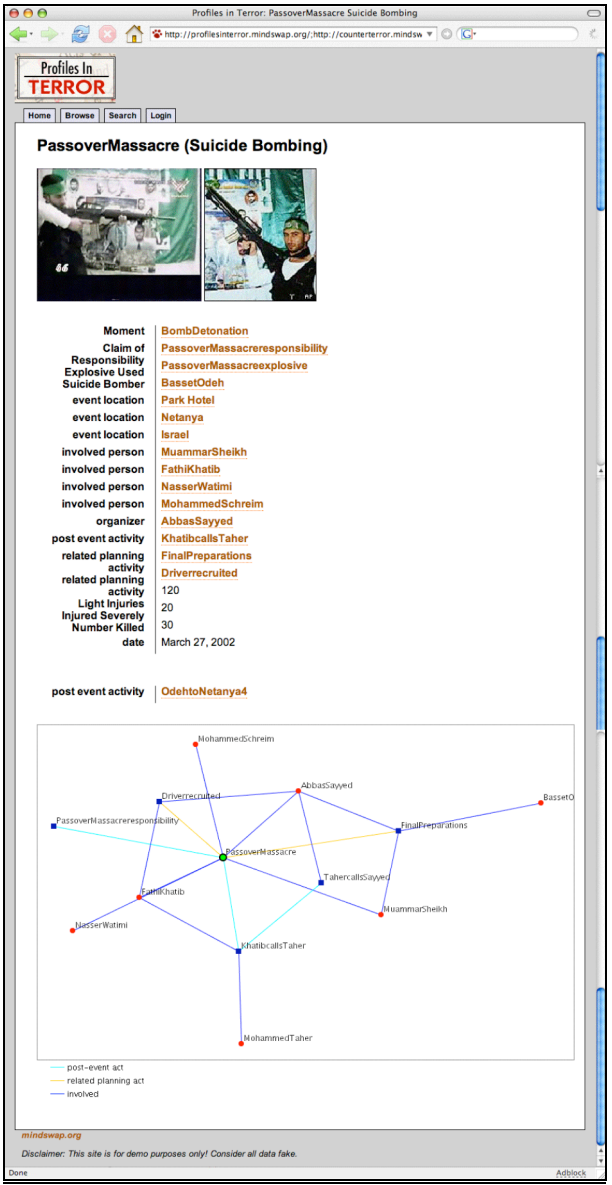

Fig 1. The Semantic Web portal page for a suicide bombing in Israel known as the Passover Massacre. Information about events is often much more extensive than this, but for the brevity of figure size, this more limited example is presented. Please see http://profilesinterror.mindswap.org for further examples.

1.2 Scenario

To see how Semantic Web portals may aid in terrorist network analysis

specifically, consider the following scenario.

An analyst for an Eastern European country's counterterror and border control program notices that a day after an Egyptian living in Britain was denied entry into the country for suspicious behavior a Syrian residing in Germany entered the country, stayed in his hotel for a few days and returned to Germany. His odd behavior, never meeting anyone and rarely leaving the hotel, aroused the hotel staff's suspicions to the point that they contacted the authorities and monitored his activity.

There is no obvious link between the two men, but the timing of their travel plans seems too close to be coincidence. Additionally, it was only a few weeks before the Parliament opened and foreign policy was going to be the central issue.

Because the counterterror program has encoded their information with Semantic Web technologies, information about the movements, activities, and associates of both men are automatically pulled from other Semantic Websites facilitating the analyst's search for possible links. Comparing data on the two men's travel showed no evidence that they had ever been in the same place before, nor did they appear to directly know someone in common. It was shown that both men had been affiliated with mosques that were implicated in Islamic activity – but there was no direct link between the mosques – or evidence that any members of the two mosques had met.

Expanding the search for any data in common, the analyst sees that the Egyptian had a former flatmate who had made several trips to Aleppo, Syria under the auspices of his graduate study. The Syrian had been born in Aleppo.

Taking another look at the flatmate, data is pulled in about his travels and past. He had trained in Afghanistan, and pulling data for others who had trained there in the same timeframe, several Aleppo natives turned up. Pulling data on them, a few were from the same neighborhood as the Syrian living in Germany – making it likely that they knew each other.

Checking into the terror-ties and Aleppo connections of the mosque frequented by the Germany-based Syrian revealed that the mosque was linked to a charity that sent funds back to Aleppo (indicating that there were strong connections between the two communities) and that members of the mosque had been arrested for planning to assemble very large explosives.

Putting it all together - the analyst determines that the Egyptian and Syrian had probably not met but were set-up to rendezvous. Because of the importance of the target a large complicated explosive was to be employed - and doing so required two pairs of hands. When his contact did not appear, the Syrian had to call off the operation.

The collected data is easily turned into a report which automatically links to the relevant files. When the analyst contacts German and British intelligence, the report is pulled immediately into their online files. These intelligence officers can then launch an investigation and add their own findings.

2 Terrorist Network Analysis on the Semantic Web

Intelligence work is a little like the unraveling of a knotted skein of wool. You get hold of an end and you have to follow it through until you are near enough to the heart of the knot to see what it consists of.

Stella Rimington, former head of MI-5, *Wall Street Journal*, January 3, 2005, "'Humint' Begins at Home"

If you get a children's magazine… the dots are numbered…. Intelligence is nothing like that. There are no numbers on the dots.

James Woolsey, former Director Central Intelligence, at the American Enterprise Institute, June 3, 2004

Whatever metaphor is applied to untangling terrorist activity, the Semantic Web can be a useful technology, for managing and analyzing data about terrorist activities. Terror operations are conspiracies involving a small group of people carrying out a complex chain of actions and bound by an intricate web of relationships. When investigating terrorism useful information is in short supply, extraneous information is abundant, and separating the two is an enormous challenge. The Semantic Web is not a silver bullet that will identify links that were invisible to analysts. But it can serve as a valuable tool for gathering, organizing, and disseminating information.

After a major crime or terrorist act - and also when potentially interesting intelligence is discovered - diligent investigators examine every possible lead. After 9/11 the National Security Agency's chief of Signals Intelligence, Maureen Baginski set the tone for the investigation, "with an approach familiar to any reader of police procedurals: on a large piece of paper, she wrote the initials 'UBL' and drew a box around them; then she asked her team to come up with any plausible connections, social and otherwise" [7]. The ad hoc databases created from these inquiries become elaborate and are often built around the intuition of the creator or creating team.

But these ad hoc databases are often on paper, are only intelligible to their creators, and are not maintained in a formal manner. The Semantic Web will facilitate the construction and use of this sort of informal database. But perhaps most importantly, a Semantic Web database can be used to share data electronically and show the process of the investigation – the quality of information, the false leads, and even hunches - to people who were not involved in the database's creation.

Mapping a terrorist organization or event requires a series of steps: gathering data, organizing the data and outlining connections, and identifying holes in the connections and developing theories to fill them. The process is then repeated as theories are tested and new leads are generated. In this process, enormous amounts of seemingly irrelevant data is accumulated, but it will need to be organized into the framework as well because it may become relevant in a later stage of the investigation. The Semantic Web can be a useful tool at each stage of an investigation.

In gathering data, traditional search engines have drawbacks for efficiently searching information. Based on natural language, traditional search engines gather data with too broad a net to be useful for the time-pressed investigator. The bulk of the information accumulated is of limited relevance. By conducting smart searches, the Semantic Web can surmount these weaknesses and maximize the amount of relevant data brought to the intelligence analyst's attention.

Search engines gather enormous quantities of information quickly, but search with out discrimination. Googling al-Qaeda leader Ayman al-Zawahiri brings over 100,000 results – the vast majority of which are irrelevant and repetitive. A particular fact may be buried within thousands of documents. For example responding to a query on Zawahiri's activities in 1993 Google provides 5000+ results. Some of these hits are responses to the copyright date on the document's publication or are an article that mentions 1993 peripherally or chatroom discussions in which Zawahiri was praised and

another discussant mentioned an event in 1993. This is because the search engine, which is based on natural language, is simply matching characters - the information has no inherent meaning to the search engine.

For the Semantic Web, Zawahiri and 1993 would have a specific meaning, so that a search could focus on information relating to Zawahiri's activities in the appropriate timeframe. The marked-up data could then be pulled into the researcher's Semantic Web portal automatically. This ability to aggregate information could save the researcher hours of scanning documents, and, by automatically placing information into a context, possibly reveal connections that the researcher would not have noticed.

A particular area where "smart" searches are essential is on names. In the realm of illicit activity aliases and false identification are commonly employed to deceive law enforcement. In dealing with terrorists from the Middle East this is exacerbated by the different transliterations used for the non-Latin alphabets of the region. Terrorists on watchlists have evaded detection (sometimes unwittingly) by simply using different Latin alphabet spellings of their names. A Semantic Web portal could be encoded to recognize Osama ben Laden and Usama bin Ladin as the same person - or to compare other information, such as birthdate or nationality. A Semantic Web portal could also be encoded to recognize nicknames and aliases, for example Abu Ammar as a common nickname for Yasser Arafat. The utility of this feature is not limited to names. Because terrorists frequently travel on fake passports and use stolen credit cards encoding false information as being connected to a particular individual would be invaluable for tracking individual's movements. This technology could also help avoid false positive identifications. Because the encoded information would have meaning to the Semantic Web, rather than just matching the letters of a name, the system could evaluate other key personal data and recognize whether an individual was the wanted terrorist or simply unfortunate enough to share a name with one.

The Semantic Web also gives the user the ability to shape the information according to changing needs. A Semantic Web portal can be used to examine information from several different angles. In one context it may be useful to examine a suspect's connections to individuals, in another it may be useful to examine that suspect's movements over a specific time period. In a different situation the user may need to examine a terrorist network as a whole. A Semantic Web portal can display data in different configurations and be modified to reflect the user's changing needs. This is particularly important for investigating terrorist activity because data is frequently fragmentary and research needs to be structured around whatever data the researchers possess.

At the core of terrorist activity is a network of personal connections that allows the terrorist organization to function. Consequently, looking at who knows whom and how they know each other is central to understanding the extent of a terrorist cell. The Semantic Web can note the various connections between cell members such as shared residences, communal affiliations, places of employment, and birthplace. By allowing the researcher to focus on these connections and organize information according to them, the researcher will be better able to unravel the web of connections underpinning a terrorist cell.

This flexibility in structuring the data is particularly useful for tracking the movement of money and of suspected terrorists. Intelligence on both of these matters

provides a crucial window into terrorist goals and operations. According to the *9/11 Commission Report* (page 385) "terrorist travel intelligence collection and analysis… has produced disproportionately useful results." Terrorist movements can reflect training needs, assembling for an operation, or planning meetings (crucial for maintaining terrorist networks because operatives try to avoid using communications that can be monitored such as telephones.) The information in a Semantic Web portal can easily be shaped to accommodate this sort of search, showing what links exist between suspected terrorists and a particular place - whether and when they were born there, traveled there, or resided there. Alternately, the information could be quickly re-organized to show suspects's peregrinations so that their routes could be compared. Because the Semantic Web portal gives meaning to the data, the Semantic Web will be extremely useful for tracking movements in time and space.

On a related issue, the Semantic Web could also be used to track terrorist codes. In telephone and email conversations terrorists frequently use simple code words to mask their plans. In one case a terror attack was called a wedding and when one of the speakers asked if the bride was ready he was actually asking about the status of the bomb the terrorists were building. On a Semantic Web portal analysts would could mark up suspicious statements and link them into the context in which they were used - time, place, and the identity and activities of the participants in the conversation.

The Semantic Web could be similarly helpful in tracking and analyzing financial transactions. For example, all the users of a suspect bank account could be noted and then compared for other connections. The Semantic Web can also be used to study patterns of use of stolen credit and ATM cards. A series of purchases of potential explosive components with stolen credit cards, for example, could indicate that an operation was being planned. Because the Semantic Web encodes data it can be an effective tool for sorting through masses of details.

Information comes in numerous forms, not just words and numbers. The Semantic Web's ability to include and annotate different forms of data is critical - because sometimes a picture really is worth a thousand words. Photographs placing individuals together have often been invaluable resources for identifying links between individuals. Photographs of graffiti, which is frequently used to communicate and mark territory by terrorists and gangs, can provide a glimpse into relations between terrorist groups. A written report describing the graffiti will not be as useful as the report may have missed a key detail or had a subjective interpretation. On the Semantic Web, the photograph could be annotated to include notes and theories about its meaning and linked to other relevant information. Maps are another example of a useful image. Simply listing suspect's addresses may not reveal the proximity of their dwellings, whereas a map that could show this might also reveal how terrorist cell members arranged meetings. Photographs of forged documents could be posted and compared for similarities in technique and other crucial details that could reveal their origins. Seamlessly including images in Semantic Web databases vastly increases the user's ability to build models of terrorist networks and activities.

The ability to share and aggregate information electronically is a feature of the Semantic Web that will be invaluable to terrorism researchers. The ad hoc databases created to track terrorist activity are often designed around the immediate needs of the investigator or investigating team and the internal dynamics of the team can mirror the

close-knit unique internal culture of the terrorist cell that is being analyzed. Consequently these databases, often paper files, are not readily intelligible to outsiders. Renowned CIA case officer, Robert Baer, who devoted a substantial part of his career to finding out who was behind the 1983 truck-bombing of the U.S. Embassy in Beirut, describes running a half-dozen Lebanese agents who gathered rumors, public records, political membership lists, old newspaper articles and photos in his book *See No Evil.* He would combine this information with information from the CIA database as well as transcripts from wiretaps. Then, Baer writes:

> *I would spend hours poring over the take, making connections between people, eliminating false leads, adding to my matrices. My makeshift charts started to look like the wiring diagram for a Boeing-747 cockpit.(Baer, 2002) [1]*

These charts are familiar to any researcher, but pity the investigator who inherits such a file. Unique abbreviations, cryptic notes, and assumptions about the information and the relationships charted characterize these charts. However, such databases created on the Semantic Web could be marked up to show who had entered data, with notes about how and why they came to their conclusions, thereby providing a window into the thinking of the investigative team. Where only a few people can view a folder or chart at once, a Semantic Web portal can be accessed by multiple people from multiple locations. This would facilitate teams made up of members operating from diverse locations, it would also allow for easier collaboration between different teams. But, it would also allow investigators of one situation better access to the data of a related investigation. This would facilitate the intuitive processes - the hunches - that help investigators see patterns. A team looking at a new incident might find something useful in the false leads from an earlier stalled investigation.

A particularly compelling example of the importance of sharing these hunches on a database is found in the *9/11 Commission Report* ([7] p. 353):

> *…In late 1999, the National Security Agency (NSA) analyzed communications associated with a man named Khalid, a man named Nawaf, and a man named Salem. Working-level officials in the intelligence community knew little more than this. But they correctly concluded that "Nawaf" and "Khalid" might be part of "an operational cadre" and that "something nefarious might be afoot."*

The *9/11 Commission Report* goes on to explain how there was information in the NSA's own database and other government databases confirming these suspicions but because of poor inter-agency communications the men were not adequately investigated and their movements were not closely tracked. Ultimately the men reached the United States and linked up with the other 9/11 hijackers.

The 9/11 Commission Report goes on to grant that "it is not likely that watchlisting [these men], by itself, [would] have prevented the 9/11 attacks." The incident also raises issues of organizational culture and procedure far beyond the scope of this chapter. But the Semantic Web could have helped reduce some inherent bureaucratic barriers. A Semantic Web system would have allowed the initial NSA team to post a note to the effect that they thought these men were involved in terrorist activity along with the data - even if it was very limited and fragmentary - that inspired this hunch. Then, if the suspects' activities caught the attention of another analyst there would have been at least

some background information. Equally useful, the second analyst could have seen who made the initial note and follow-up with them.

This sharing would probably not have prevented 9/11. But the U.S. intelligence community consists of dozens of agencies with thousands of analysts between them that sift through petabytes of data daily. Operating on this scale, calling to break down bureaucratic barriers to "connect the dots" is easier said than done. Encouraging more communication may result in analysts and teams drowning each other in data. The Semantic Web can help point researchers and analysts towards the information they need.

Using the Semantic Web's capabilities to monitor terrorist activity and model terrorist networks requires an ontology crafted to express complicated and sometimes contradictory information in an accessible and intuitive way. Terrorists and terrorist organizations engage in a wide range of activities that reflect the variety of human endeavor. An ontology that reflects this must maintain a careful balance between being sufficiently comprehensive while not being so specific that it is tailored to particular situations and cannot accommodate others. The complexity and diversity of situations is such that at some points paragraphs of text are necessary – but this is a last resort because doing so defeats the point of and does not use the complete capabilities of the Semantic Web. At the same time, the ontology must be capable of growing and changing to reflect new aspects of a changing phenomenon carried out by adaptable actors. Creating an ontology that is a useful tool for publicly accessible research purposes is a unique challenge.

There are several terrorism databases that describe terror attacks that are useful starting points for ontology building. The categories of these databases include Date, Location, Group, Casualties, Target Type, and Means of Attack. These categories are useful for supporting statistical surveys, but ultimately they rely on a text description of events for the details and consequently are of limited utility for a full Semantic Web mapping of a terrorist event. One difficulty presented is that a common terrorist tactic (exemplified by 9/11) is to launch simultaneous attacks against multiple targets. The November 28, 2002 attacks against Israeli targets in Kenya in which a trio of suicide bombers drove a truck into an Israeli owned hotel in Mombassa, Kenya and just a short time later two anti-aircraft missiles were fired at an Israeli airliner exemplifies this problem. The attack had two distinct components, the suicide attack and the missiles. These attacks had different means, locations, results, and operatives, but the same leaders organized the attacks.

This only scratches the surface of the issues needed to reflect a terrorist attack. In the Mombassa attack, the missiles fired at the Israeli airliner were believed to be from the same batch that included missiles fired at a U.S. military plane in Saudi Arabia and other found near the Prague Airport when Israel's Foreign Minister flew in. The type of weapon is important, but the weapon's history – how it was acquired – is essential to unraveling a terror attack. Background on other items involved in the attack, such as the car and the explosives are also important to investigators. The bomb's primary ingredients and design are of obvious interest. Putting together the history of the bomb's construction might be a key part of the investigation (British experts have referred to the bombs left on the subway that failed to detonate as "forensic bingo.") The vehicle's history is also important to investigators seeking to examine the terror attack. The cell

that bombed the World Trade Center in 1993 was traced through the remnants of the rental truck used in the bombing.

Investigators and analysts also need to be able to break events down into the exact sequence of events. Describing a suicide bombers search for an a target, or the events in which a plane was hijacked – without simply resorting to blocks of text is essential to building a useful Semantic Web tool for examining terrorist activity.

The actual attack is only the tip of the iceberg, the final link in a long chain of events including the reconnaissance and planning, and the assembling components and personnel. The 1998 attacks on the U.S. Embassies in Africa and 9/11 both required several years of careful planning. Further, the individuals that carry out attacks are often bound by a complex web of relationships that extend over the course of several years.

3 Modeling Terrorism in OWL

3.1 Conceptual Modeling

As the above examples demonstrate, reflecting events and sequences of events in time is a central challenge to building a successful terrorism ontology. A class of events to describe terror attacks in which properties included location, date, casualties etc. based on the existing terrorism databases was created, but, as described above it was insufficiently specific. The problem of describing multiple attacks, was solved with the Russian stacking doll solution – that is creating a sub-attack property, which has the range of terrorist attack so that several events can be linked as part of one mega-attack. This solution proved useful in many other places as well.

The problem of describing minute to minute events was solved by creating a "Moment" class to describe specific moments in time. In a multiple suicide bombing, a moment can be created and attached for each bomber. In other complex attacks, such as airline hijackings, moments can be used to reflect the shifting situation and the different aspects of the hostage negotiation. For certain, specific and relatively frequent Moments, such as suicide bomb detonations, a subclass was created with unique properties.

In order to describe the meetings and movements that go into planning a terrorist attack several subclasses of event were created. "Contacts" class describes meetings between operatives. Subclasses were created to describe different types of contacts such as meetings, telephone and internet communications, and information delivered by messenger. Important properties, besides location and participants, reflect what was transacted in the contact. It can include instructions, advice, orders, and approval of plans. Items can also be transacted; weapons, explosives, identification documents, and cash have all been transferred in meetings. Special subclasses had to be created to describe financial transfers that were done through financial institutions and for explosives construction.

Another important type of event that is important to tracking terrorist activity is travel. A class for travel events was created. Because a journey often involves numerous stages, the stacking doll solution was applied and a property called "Travel Segment" with the range of "Travel Event" was created.

Events have properties of "Related Planning Activity" and "Post Event Activity" which have the range of event so that Travels and Contacts can be attached to each other and to attacks.

One property consistently shared by every event is "Involved Person," which has the range of "Person." Similar to terrorist events, there are databases, such as the FBI's Most Wanted List, that provide basic information on individuals such as height, physical description, date, and place of birth. There are a few subproperties of "Involved Person" such as "Suicide Bomber" but for the most part the Involved Person's role is defined by their participation in "Related Events" and "Moments." However, personal connections are at the core of terrorist activity. Former CIA operative and forensic psychiatrist Prof. Marc Sageman, in his excellent *Understanding Terror Networks* writes, "…social bonds play a more important role in the emergence of the global Salafi jihad than ideology. Friends and relatives of identified terrorists need to be pursued and investigated wherever they reside. Especially important are those who were friends of a terrorist just before he started acts in furtherance of the jihad…" ([6] p.178)

Events described, be they meetings between people or terror attacks, described specific action. Meetings in which specific information and items were transferred are important to understanding terrorism but equally important are the ongoing relationships that cement the small cliques that become terrorist cells. To reflect this, a different class for relationships was created. This class allows two or more involved persons and allows events to be attached. In particular, marriages have a unique event – the wedding. Since weddings are attended by the bride and groom's closest friends the attendees can include future co-conspirators. Active members of an underground organization are often married in secret and anyone attending is likely to be affiliated with the organization. Other important relationships include housemate, co-worker, and organizational affiliation.

Similarly, individuals have permanent and changing characteristics. Height and date of birth are permanent (although there may be conflicting information about them) but an individual's employment status, for example, is subject to change. Paralleling the relationships class, a personal status class was created with subclasses including "Employed," "Incarcarated," "Military Service," "Student," "Hospitalized," "Resided," and "Terrorist training." However, the number of possible activities of an individual are enormous so a catch-all category for "Organizational affiliation" was also created. This could encompass volunteer work with a political party, attendance at a religious institution, or a wide variety of other activities.

3.2 The Ontology

The resulting terrorism ontology is on the web at
www.mindswap.org/dav/ontologies/terrorism.owl

The ontology is written in the Web Ontology Language OWL in OWL DL. It contains seventy different classes, and 173 properties (71 DatatypeProperties and 102 ObjectProperties). The only OWL DL feature used is the unionOf property. The topics include events, relationships, resources, locations, and the relationships among them.

This ontology is used within the Profiles in Terror portal at http://profilesinterror.mindswap.org/. The website collects information about different terrorist people and events and provides the user with a way of browsing among them.

The ontology is also used to highlight terms within terrorism related news stories and link them into concepts described with the ontology. We envision this website as a starting point for implementation of the concepts described above.

4 Conclusion

The Semantic Web holds tremendous promise as a useful resource for students of the phenomenon of terrorism. But for this promise to be fulfilled a successful ontology is necessary. The ontology must be specialized enough to describe some of the unique aspects of terrorism - such as explosives - but readily comprehensible to the non-expert. For ease of use, it is essential that the ontology remain on a manageable scale and not expand to hundreds of classes so that the average user can understand and manipulate it.

Developing a workable terrorism ontology could have many other benefits. Using an ontology to describe real human activity presents many challenges. The ways in which people travel, communicate with each other, and change over time makes developing an ontology to describe them more of a literary endeavor than a science. It is no stretch of imagination to apply a terrorism ontology to studying other criminal activity. But if key aspects of human experience are successfully described by the ontology there are many other possible uses - from sociological research to describing travel experiences to charting the behavior of characters in a soap opera. A successful and accessible terrorism ontology could be an important step to making the Semantic Web a real tool for the broader public.

References

1. Baer, Robert, *See No Evil*, Crown Publishers, New York: 2002.
2. Becket, Dave (ed.), *RDF/XML Syntax Specification (Revised)*, W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/rdf-syntax-grammar/>
3. Brickley, Dan (ed.), *RDF Vocabulary Description Language 1.0: RDF Schema*, W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/rdf-schema/>
4. Dean, Mike, Schreiber, Guus (eds.), *OWL Web Ontology Language Reference,* W3C Recommendation, 10 February 2004, <http://www.w3.org/TR/owl-ref/>
5. National Commission on Terrorist Attacks, *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States*, W. W. Norton & Company, New York: 2004.
6. Sageman, Marc, *Understanding Terror Networks*, University of Pennsylvania Press, Philadelphia: 2004
7. Walsh, Elsa "Learning to Spy: Can Maureen Baginski save the F.B.I.," *The New Yorker*, November 8, 2004.