A platform for discovering and sharing confidential ballistic crime data.

YATES, Simeon, AKHGAR, Babak, BATES, Christopher, JOPEK, Lukasz and WILSON, Richard

Published version

YATES, Simeon, AKHGAR, Babak, BATES, Christopher, JOPEK, Lukasz and WILSON, Richard (2011). A platform for discovering and sharing confidential ballistic crime data. International Journal of Knowledge and Web Intelligence, 2 (2/3), 202-218.

A Platform for Discovering and Sharing Confidential Ballistic Crime Data

Simeon J. Yates*, Babak Akhgar, Christopher Bates, Lucasz Jopek, Richard Wilson

Cultural, Communication and Computing Research Institute
Sheffield Hallam University
Sheffield, S1 1WB
UK
*Corresponding Author
E-mail: s.yates@shu.ac.uk

Abstract: Criminal investigations generate large volumes of complex data that detectives have to analyse and understand. This data tends to be "siloed" within individual jurisdictions and re-using it in other investigations can be difficult. Investigations into trans-national crimes are hampered by the problem of discovering relevant data held by agencies in other countries and of sharing those data. Gun-crimes are one major type of incident that showcases this: guns are easily moved across borders and used in multiple crimes but finding that a weapon was used elsewhere in Europe is difficult. In this paper we report on the Odyssey Project, an EU-funded initiative to mine, manipulate and share data about weapons and crimes. The project demonstrates the automatic combining of data from disparate repositories for cross-correlation and automated analysis. The data arrive from different cultural/domains with multiple reference models using real-time data feeds and historical databases.

1 Introduction

This paper describes the work undertaken to research and develop a prototype solution to the linking, presentation and analysis of cross-boarder gun crime data within the European Union. This domain is one where technical, policing, national and EU legal frameworks and the behaviours of police forces and criminals regularly change, sometimes dramatically within a short time span. The proposed solution described below has been developed to ensure the system can remain responsive, domain relevant and effective whilst adapting reasonably dynamically to these changes.

Globalisation has been accompanied by a dramatic increase in organised and trans-national crime and terrorism. It takes many forms from homicide, to trafficking of drugs, people and weapons through to the laundering of crime proceeds. The objective of the Odyssey Project has been to develop a prototype intelligence platform for the secure sharing and manipulation of data about ballistic crimes. Ballistic crimes are those that involve the use of firearms and other weapons, ranging from smuggling and the supply of illegal firearms through to homicides (Akhgar, 2009).

Although Odyssey focused on ballistics data, the concept and architecture are immediately applicable to other forensic data sets including DNA, fingerprints, mobile phone records, and explosives analysis. The techniques developed within the project for querying and manipulation could be applied to any domain that involves rich data and personal records.

The platform consists of a series of components including: security, data sharing (data selection and upload, querying, storage of query plans), non-relational data manipulation (semantic querying, data mining and relationship discovery), support for query development (domain-specific query language, intensional support) and an alerting component that executes queries automatically. All of this is built on top of a distributed architecture using message queues to link a range of back-end engines.

The European Commission part funded the Odyssey Project. The project partners are: Sheffield Hallam University (United Kingdom), Atos Origin (Spain), Forensic Pathways Ltd. (United Kingdom), EUROPOL (Netherlands), XLAB (Slovenia), Politecnico Di Milano (Italy), West Midlands Police (United Kingdom), Royal Military Academy (Belgium), An Garda Siochana (Republic of Ireland), SAS Software Ltd. (United Kingdom) and Direzione Centrale Anticrimine - Servizio Polizia Scientifica (Italy).

2 Policing, Legal Frameworks and Ballistics

Requirements for the Odyssey prototype are conceptually driven by the interrelationships of 3 core components. These serving as a canonical set of requirements and cover: Policing; Legal frameworks; and Ballistics. Most importantly these three key sets of requirements are not fixed nor are the relationships between them. Within the time frame of the existing project a range of new technologies for the physical analysis and image matching of ballistic evidence have come to market. Legal frameworks within the EU and member states currently limit the amount and types of data that can be shared across boarders, yet these frameworks are under active review and development. Police forces across the EU are facing a number of ever changing challenges in relation to both criminal and terrorist activity, and available resources. This section details some of the main issues within each of these domains that have provided context for the project.

2.1 Policing

A key aspect of the context in which the Odyssey prototype has been developed is that of existing police information systems. Internationally, there are a large number of bespoke systems including COPLINK, NABIS, HOLMES-2 and I-24/7. COPLINK is an information and knowledge management system aimed at capturing, accessing, analysing, visualising and sharing information between United States law enforcement agencies. COPLINK comprises of two components COPLINK Connect (CC) and COPLINK Detect. COPLINK Connect is designed to integrate disparate heterogeneous data sources, including legacy systems, to facilitate information sharing between police departments.

COPLINK Detect tries to discover associations within police databases. It supports detectives and crime analysts in finding associations between people, vehicles, incidents and locations. The strength of an association is determined through the use of co-occurrence analysis and clustering. The system is able to search for meaningful terms in both structured (database tables) and unstructured (witness statements) data (Chen et al. 2003).

UK police forces have access to a number of independent database systems. These databases are used to record, monitor and manage offences in such areas as sex offences, gun crimes and major incident management. NABIS provides ballistic examination services, for twenty UK based police forces, through three hubs, which are based in London, Birmingham and Manchester (Sims 2010). Odyssey is different to the NABIS and I-24/7 systems used in the UK and at Europol because it gives users control of their own data whilst letting them use a range of techniques to query all shared data within the system. The NABIS database is specifically designed to manage data from the examination of ballistic items (Nabis, 2009). I-24/7 has a European-wide dataset that, largely, retains information related to the individual (Interpol, 2007). A gap exists between the systems that collect, store and integrate data on ballistic crime within the EU and those that manage more general data about crimes and criminal activities.

Odyssey tries to narrow this gap by combining data from a wider range of sources than existing systems do. This data will be interrogated using a variety of techniques including relational queries, data mining and semantically based searches. The semantic approach moves querying nearer to the end-users' domain of experience and away from traditional IT. Using context to support the development of queries lets the system find a wider range of results than it would from a more restricted query. Akrivas et al (2002), were able to demonstrate that using semantic structures could expand queries to make them more "intelligent". Odyssey combines expansions of this type with data mining to add weightings to the enhanced results so that users are better able to navigate them.

2.2 Ballistics

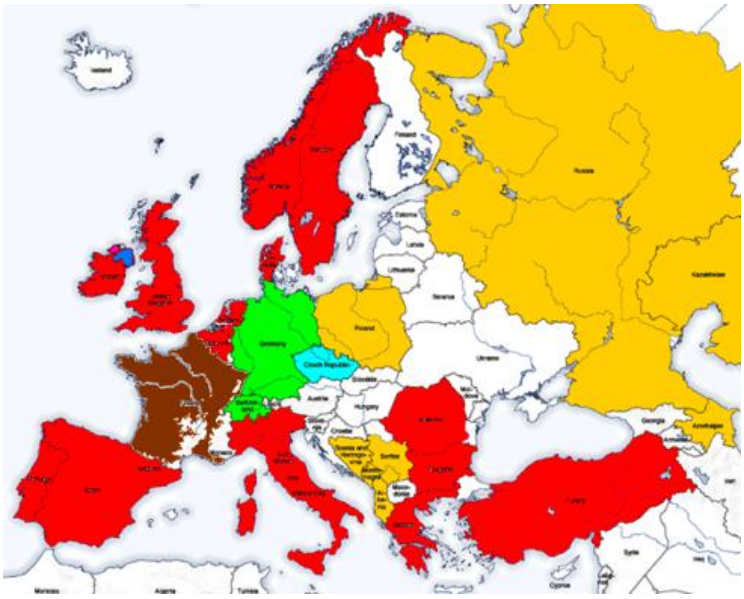
As noted above a number of systems have been developed to support the process of ballistic evidence identification and matching. When ballistic crime is investigated, forensic specialists can compare recovered items such as guns, bullets or cartridge-cases. Test-fired bullets are examined for a range of marks made as they exit the barrel of the gun (Bundeskriminalant 2004). By comparing the marks on different bullets a trained examiner can determine the likelihood that two projectiles came from the same weapon.

Figure 1 provides some indication of the different ballistic matching systems in place across Europe. The range of systems in use makes comparing ballistic evidence across national boundaries both complex and expensive. Items must be transported between sites and re-scanned on the different

systems. The EU and Member States cannot access comprehensive up to date accurate, meaningful assessments and statistical information about the incidence of gun crime and terrorism within, between and among *all* Member States. Neither EUROPOL nor Interpol is able to provide this important information to aid policymaking and decision support. Policy development is therefore piecemeal with no clear methods available to check that the right policy is in place and the right level of resource is being applied. The implications of this situation are best understood through an example case.

This case involved a simple comparison of a bullet recovered in the UK from an Armed Robbery with a case of suspected terrorism in Germany and Belgium. No common pan-European integrated ballistics intelligence information system existed and there was no method of sending a photographic image of the cartridge case or intelligence information related to it electronically to EU forensic laboratories for comparison purposes. The present cost incurred for forensic comparison of one bullet is around €9,000 including travel, accommodation, equipment, logistics and scientific investigations etc. Considering the impact of these costs on the EU it becomes apparent that they almost certainly prevent Member States undertaking too many such investigations.

Figure 1. Ballistic systems in place across Europe



2.3 Legal framework

Both personal and crime data are very sensitive and have to be handled with care. Moving any sensitive data between jurisdictions increases the possibility that it will be compromised. Consequently a range of legislation covers data sharing within the EU. These laws and associated rules place restrictions on law enforcement agencies as they do on individuals or on businesses. Some key foundational issues are detailed in the following sections.

2.3.1 The Swedish Initiative

This is a statement proposing a framework for the simplification of the exchange of information and intelligence between law enforcement authorities. It was adopted it in December, 2006. Nygren (2008) points out that under this initiative the rules governing the cross-border exchange of criminal information and intelligence cannot be stricter than those applying to internal data exchange. In other words cross-boarder data exchange should be equally as open or as closed, and meet the same security standards as within-nation exchange.

2.3.2 Principal of availability

The principal of availability introduces a new form of cooperation in criminal matters within the EU. Law enforcement authorities in one Member State are empowered to grant access to their information

to authorities in other Member States for the purpose of prevention, detection and investigation of criminal offences. Europa (2008) states:

"The principle subjects the exchange of law enforcement information to uniform conditions across the Union. If a law enforcement officer or Europol needs information to perform its lawful tasks, it may obtain this information, and the Member State that controls this information, is obliged to make it available for the stated purpose".

Sharing personal information or information which could be used to identify an individual has always been difficult. Under the principle of availability "the exchange of personal data within the framework of police and judicial cooperation in criminal matters, notably under the principle of availability of information as laid down in the Hague Programme, should be supported by clear rules enhancing mutual trust between the competent authorities and ensuring that the relevant information is protected in a way that excludes any discrimination in respect of such cooperation between the Member States while fully respecting fundamental rights of individuals". Systems such as Odyssey should be built to both encourage the use of personal data where appropriate and to ensure its security at all times.

2.3.3 Prüm decision

A sub set of the EU Member States: Germany, Spain, France, Luxembourg, the Netherlands, Austria and Belgium signed the 'Prüm Treaty' in the German town of Prüm on 27 May 2005 (Prüm 2010). The European Commission supported the German initiative to transform this treaty into an instrument binding on all EU Member States and the Council adopted the Prüm Decision and its implementing provisions on 23 June 2008. The Prüm Decision is described by the EU DG Home affairs as providing for "the automated exchange of DNA, fingerprints, and vehicle registration data, as well other forms of police cooperation, between the 27 Member States".

Given the lack of interoperability, standards and comparable working practice within, between and among EU member states with regard to data exchange on ballistic crime the Odyssey prototype provides a means of responding to this requirement under the Prüm Decision.

3 The Odyssey Prototype

Gun crime is not going to be solved through the use of IT systems, however excellent they might be. Detection remains an essentially human process, (Smith & Tilley, 2005), but one which can be supported through the use of good IT applications.

The detection of pan-national ballistic crime breaks down into a number of complex problems. The first is the realisation that such crime is happening and, for the individual investigator, that her crime might be related to ones which happened across the border. The second problem is to discover the related data. Where crimes occur in different jurisdictions there may be no way in which data about them can be shared readily or easily. Only by sharing data can investigators become aware that two incidents are similar or that they may form part of a larger pattern. The final problem is to share the actual ballistic data: meta-data about bullets or guns, images taken from comparison microscopes or automated imaging systems. The Odyssey platform demonstrates that all of these problems can be addressed using a suitably complex and distributed data management application

When thinking about ballistic investigations one naturally thinks about comparisons of evidence taken from a scene or a victim with images and reference samples. It becomes clear that automating these comparisons is desirable. For a number of technical reasons to do with the ways in which images are taken this is a significantly complex problem. It has been scoped by the Odyssey project but a solution has not been developed.

3.1 Architecture

The prototype uses both local nodes and a central hub with asynchronous communication between them across a message queue. Individual components of the prototype are wrapped in Web Services so that the platform can combine the flexibility and scalability of a modern Service-Oriented Architecture with the robustness and power of a centralised system.

A number of factors impacted upon the architecture including: the need to manipulate data which is distributed across member states; the importance of securing both data and access to it; the use of

different back-ends to manipulate data; the data is likely to be both incomplete and noisy; and this is a distributed system with all of the problems which are typically found in such systems.

Data and processing have to be distributed across locations. The platform's mixture of back-ends would benefit from a single centralised data store containing records of all incidents of gun-crime from across the EU. Such a store would simplify the tracking of weapons or patterns of usage; but as noted above, the use and sharing of crime data is subject to many restrictions, some defined at European level, others set by national Governments. These regulations tend to emphasise the protection of the individual's right to privacy and generally mean that any data that might identify an individual cannot, as a matter of routine, be shared between member states. In developing software and systems for law enforcement this is usually taken to mean that data are always held locally but that individual records may be shared for specific purpose. This presents a difficulty for Odyssey that uses data mining to discover patterns within crime data. To be compliant with European regulations the platform can centralise ballistic data (guns, bullets, etc.) and some data about incidents but nothing that might be used to identify victims, witnesses or perpetrators.

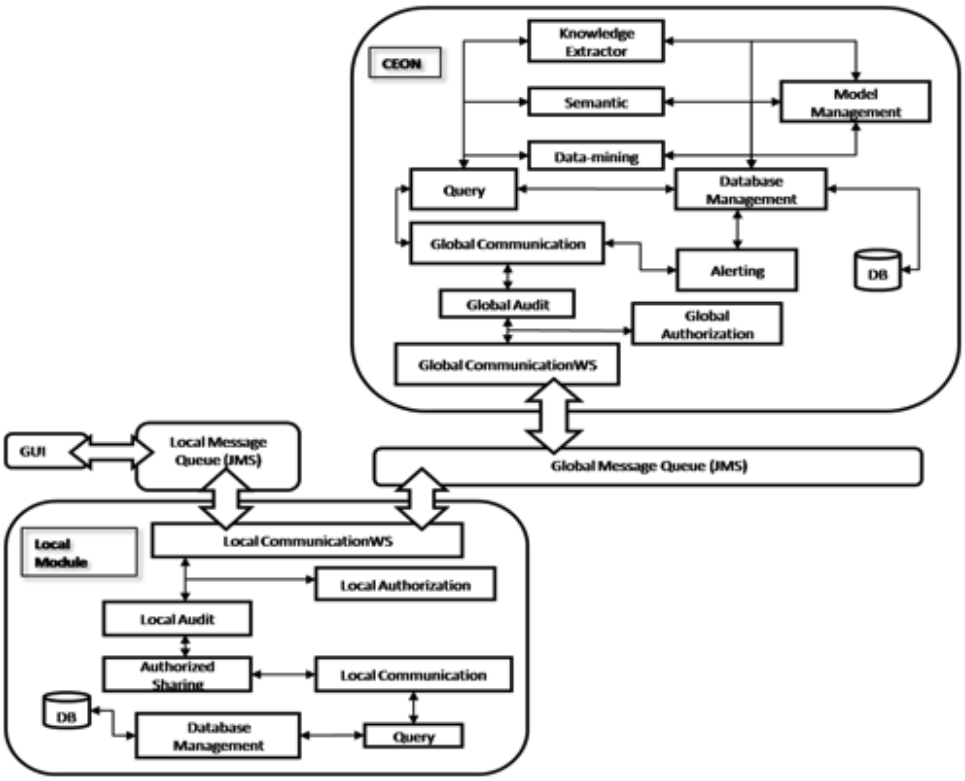
Security is an important requirement for any system used by law enforcement agencies. The data that Odyssey stores and manipulates is sensitive because it often relates to on-going criminal investigations. The architecture has to balance the competing need to keep data secure and the need to share data with colleagues who, since this is a pan-European system, may work in different jurisdictions. Odyssey has a fairly standard security scheme in which users must authenticate on to the platform with an ID and a password before being given access to data and processing based on their role and location. Messages moved across the queue are encrypted using a public key infrastructure whilst the queue itself runs over a VPN.

The platform has three different data processing modules. There is a standard relational database that holds bulk data and handles queries in which target records are known or can be easily identified. There is a data mining system that is used to discover patterns within the data. Such a pattern may be a set of records which appear to be related to a particular record of interest but which do not have direct connections in the relational data, or changes to the data as when a new type of weapon enters the market and is seen to move across Europe. Finally, we have an Intensional Querying module that through understanding the data helps investigators formulate better queries (Giacomo, 1996).

There are no standards defining the data that are gathered during investigations. Each country uses its own approach – individual organisations within the same country may even gather different data. Often the data is incomplete because officers do not have the time or expertise to enter it correctly into a computer system. Data is also incomplete because investigations are live processes. As an investigation proceeds more data is gathered and new relationships are created and existing ones modified or removed. The platform has to handle these changes and make them explicit to investigators.

The prototype has had to be designed to handle some of the more common problems of distributed systems. Processing queries can take a long time, especially when they rely on mining of large data sets. The central system has to be able to handle multiple concurrent queries that may be resident on the server for long periods. Clients cannot remain connected to the server whilst their long-running queries execute. The architecture has to be built so that clients receive responses to their earlier queries when users authenticate onto the system. This can be achieved in many ways, on the Odyssey platform it is done through the use of an asynchronous message queue. The Odyssey platform is built from three separate modules: a local node, a Central Odyssey Node (CEON) that has richer functionality and a message queue.

Figure 2. Odyssey architecture



3.2 Local nodes

The local node is the primary repository within the platform. The local node has a PostgreSQL relational database that holds data about ballistic items and crimes within a particular jurisdiction. The database is accessed through a local message queue and an endpoint that parses incoming requests and translates them into SQL commands that are then applied to PostgreSQL.

The local node routes communication between agency to CEON, and agency-to-agency. Using the Odyssey platform local authorities are able to share secure messages including queries and their results. But its function is also an encryption of all messages, decryption and verification of all incoming messages, auditing of communication, access to local database through the IDatabaseComponent interface, interfacing with GUI components through ICommunicationComponent interface, interfacing with JMS broker, and authorizing data to be sent to CEON.

Each police force or other authority runs its own local node. When the platform is fully operational there are many local nodes running but all are independent of each other. The Odyssey desktop client gives users access to their local node but not to any of the other nodes in the system thus avoiding problems of trans-jurisdictional access to data. A node can be any size. Some may hold data for an entire nation whilst others might contain just the data for a particular area.

Using only the local node has few benefits over using existing Police information systems since any results are based on data that are likely to be in those other systems. The power of Odyssey comes from combining local and central results.

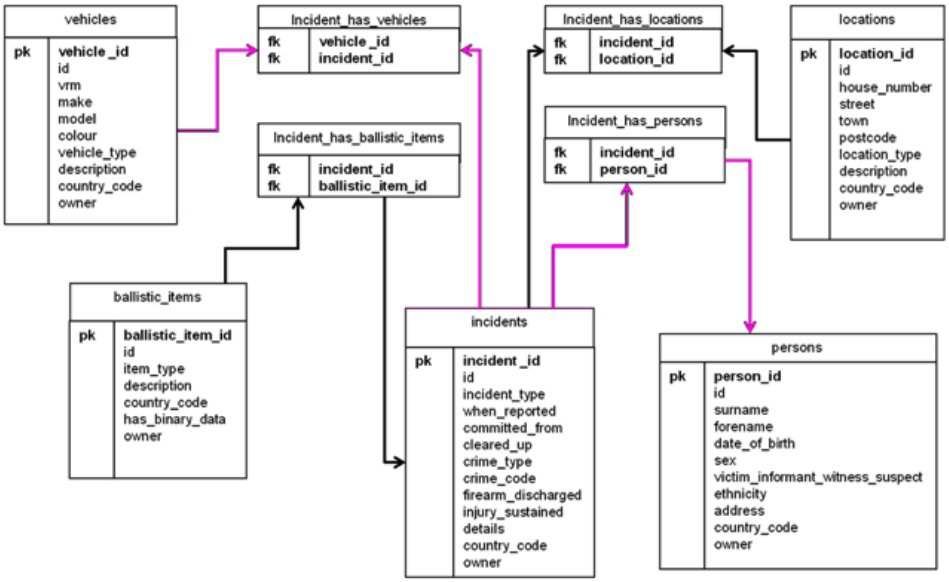
3.3 CEON

CEON, is at the heart of the platform. CEON has exactly the same queue endpoint as the local node and a PostgreSQL database that has exactly the same structure as the local one. CEON also has connections to an Intensional querying system and to a data-mining application, SAS 9.2. The platform has an experimental Semantic Web engine which tries to provide a richer querying interface through domain-derived taxonomic structures.

3.4 Relational Database

The main data store in the platform is a relational database developed using PostgreSQL 8.4. The database structure reflects the types of structure used in systems such as COPLINK, NABIS and by some of the databases used at EUROPOL. Most of its tables hold metadata with relatively few tables required to store the details of incidents and investigations. Figure 3 shows a fragment of the structure. The database structure is replicated at each local node. Each Local authority includes only its own data in its local node. Any data, which it wishes to share with other authorities, is uploaded to CEON where the same database structure appertains.

Figure 3. Partial database schema



4 Manipulating the Data

4.1 Intensional Querying

The application of data mining techniques to extract useful knowledge from datasets has been researched over a number of years, (Nath 2006). Implementations have been tried in a number of Police information systems, notably COPLINK, (Chen et al. 2004). By mining frequent patterns from repositories, it is possible to provide the investigators with partial, and often statistically-supported, results. However, such results can never be guaranteed to be completely accurate and may send investigations in the wrong direction by suggesting the wrong line of enquiry.

The Odyssey platform uses the uncertainty of data mining to give investigators *implicit* knowledge from the repositories and to use that knowledge to formulation more effective queries (Strohmaier et al, 2009). When a user faces a large and complex dataset for the first time they will not know its features. Frequency patterns provide a way to understand what is contained in the dataset. Summarizing the vast integrated dataset shared by different EU Police Organizations can increase the quality of results, accessing the most promising results for a given query. To this end the Intensional Querying module has been developed. We envisaged two possibilities for the use of *approximate* knowledge:

- The user directly queries the association rule base.
- The user queries the Odyssey repositories, but also receives an approximate answer.

In both cases the user will be provided with some useful *general knowledge* related to the mode of investigation. In the following trivial query, expressed in Odyssey's querying DSL):

```
WHAT ABOUT Incident Person WHERE country_of_crime = 'UK' AND
gender='m' WITH CONFIDENCE 0.9
```

The statement will trigger the intensional knowledge system to return any information about the listed elements given the defined conditions. Thus every association rule containing:

- (at least) attributes from the relations translated from the keywords in the `WHAT ABOUT` list (for example `Incident, Person`)
- in which elements satisfy the conditions (for example `country_of_crime = 'UK' AND gender = 'm'`)
- having confidence more or equal than the stated value (for example `0.9`)

The results are sent back to the intensional system for further processing such as ordering. The completed result set is returned to the client where it acts as a prompt, or set of prompts, to the user to help them either refine or widen their search criteria.

4.2 Data Mining

The CEON component includes a full SAS data-mining system which is used to manage data uploads through its excellent GUI tools and to mine the repository looking for patterns and hidden structures. The data-mining and knowledge extraction modules need to pre-process the database data in order to extract information for its later use. In particular, SAS data-mining solution requires for a de-normalised version of the data (Wilson et al, 2010). Processes to load any data that has changed into SAS and add it to the de-normalised structure are triggered periodically to keep it up-to-date. Mining queries may then be re-executed. The reason that Odyssey has a central database is so that it can mine data. The benefit of centralising and sharing is that much richer results can be obtained. When a data-mining query discovers data it actually returns only record IDs. The middleware sends these IDs to the CEON instance of Postgres where they are used in SQL SELECT statements to retrieve complete records. These records are returned to the user who initiated the query.

4.3 Semantic Querying

The final backend that is available to users is a Semantic engine. One of the first acts of the Odyssey project was to define the taxonomy of ballistic items and ballistic crimes. Inputs to, and outputs from, the platform must be structured according to this taxonomy.

Organisations using local nodes are able to share data by uploading it into CEON. Typically they will upload a subset of their local database composed of records that they have permission to share. Most of the data held in Odyssey can be shared without encountering problems of privacy or confidentiality. For example, the details of a used cartridge case are not likely to be confidential. Data about crimes and possible crimes are more sensitive since from these it might be possible to identify people. Where data is sensitive in this way the platform lets authorities share those columns that will not conflict with data management legislation.

The kinds of queries that investigators ask are conceptually rich and include a lot of uncertainty (De Bruin et al. 2006). In Odyssey these queries are handled using a semantic engine that runs at CEON. Queries are converted into SPARQL and applied to the data through a Jena engine. Both the semantic engine and SAS are used to automate and simplify the process of discovering similar data to that which is being investigated. This gives detectives the opportunity to find hidden relationships within trans-national datasets that they would otherwise never find.

The semantic engine lets users build queries that are dependent on their role. A crime analyst may want to ask different questions to those which a detective asks - they may be more strategic or intelligence-led, whilst the detective is focussed on operational matters. Such roles are not static. The same user may sometimes require intelligence data and at other times require operational information. Vallet et al, (2007) note that "users may have stable and recurrent overall preferences, not all of their interests are relevant all the time. Instead, usually only a subset is active at a given situation, and the rest can be considered as noise preferences". The platform has to take into account the changing context within which a user queries the system.

5 Query Language

Users of the Odyssey system will be experts in the gathering and analysis of complex, incomplete data. Detectives and crime analysts or other civilian support staff are experts in the understanding of crimes through the use of rich data such as statements or observations (Smith & Tilley, 2005). This intellectually complex work requires a clear cognitive focus and well-honed skills. The Odyssey platform is a very complex piece of software. Users cannot be expected to know that their queries are being applied to different back-ends or what data structures are used within the system. Indeed their use of the platform should, wherever possible, be natural so that the system supports and enhances their usual working practices.

A domain-specific language, DSL, is an artificial computer language that is used to describe solutions to constrained problems. A DSL provides a natural and effective interface between a complex system and its users,(Fowler & Parsons, 2010) which can be more expressive than operations constructed purely through a GUI. Domain-specific languages express complexity at a particular abstraction tailored to both current and future needs (Yu, 2008). A DSL lets non-technical people understand the overall design of a platform and interact with it, using an understandable notation that reflects their particular perspective (Bonino et al. 2004).

The DSL that was created is called the Odyssey Semantic Language (OSL). It supports the modelling of active crime investigations by operational detectives and facilitates the linking of generic crime features to ballistic data. Its innovative features are associating data retrieval techniques with data-mining results and encapsulating multiple services. Moreover, the language facilitates modelling of investigation processes and is an integral part in the platform's security.

5.1 Defining the DSL

OSL is a formal language specified by a context-free grammar. The OSL grammar was structured to make use of tokens taken from the English language in such way that the resulting constructions, that is, those sentences considered valid by the grammar, resemble the natural language of investigators so as to facilitate their construction and interpretation, (Jopek et al, 2010).

The grammar is defined in the Extended Backus-Naur Form within the ANTLR framework, a language recognition tool that simplifies the construction of a parser and lexical analyser pair from the grammar definition, as well as allowing for additional embedded code - in this case in Java. This simplifies the creation of a translation into the languages needed for the subsystem modules that are mainly SQL.

The language has relatively few keywords. Most keywords actually occur inside meaningful phrases as shown below:

GET CHARACTERISTICS returns a taxonomic structure: GET CHARACTERISTICS Person returns all the fields that describe a person (gender, ethnicity, age, etc.)

IS IT TRUE THAT returns "Yes" if the condition is true otherwise "No". For instance: `IS IT TRUE THAT Vehicle HAS PROPERTY VehicleMake WITH VALUE 'Saab'`

SHOW STATISTICS gives simple statistical information such as average value, standard deviance and variance about records matching certain criteria. For example: `SHOW STATISTICS ON PersonEthnicity WITH VALUE 'white'`

SHOW SIMILARITIES BETWEEN: `SHOW SIMILARITIES BETWEEN Person WITH VALUE 1 AND 13`

SHOW QUERY / SHOW ALL both declare a simple retrieval from database. The difference between them is that SHOW QUERY creates normal joins between tables whereas SHOW ALL does a full outer join between tables.

WHAT ABOUT executes an intensional query: `WHAT ABOUT Person Vehicle WHERE VehicleMake = 'Ford'`

SHOW SIMILAR returns all records that are share the value of at least the given number of columns with the given instance: `SHOW SIMILAR Person WITH VALUE 1 HAVING 4 EQUAL COLUMNS`

CONFIDENCE: the value specified affects the number of results returned to the user. The higher the confidence the smaller the returned result set. For example: `WHAT ABOUT Person WHERE ethnicity = 'white' WITH CONFIDENCE 0.5`

The example below presents a query expressed in Odyssey Semantic Language (OSL) that retrieves firearms with a twenty-two calibre:

QUERY `firearm` **WHERE** `calibre` **HAS VALUE** 0.22

Typically requests into the system begin with QUERY. This term was chosen because there are so many possible terms (SEARCH, GET, FIND) that we needed one that was neutral and meaningful. OSL is used to upload data, share it and modify it that is why all operations need to begin with a keyword that identifies the operation (QUERY, UPLOAD, MODIFY, ALLOW).

In the example, `firearm` is used to identify the database table that is going to be searched. Users are never told that this is a table. They interact with a set of objects that come from their domain, from detective work. These include firearm, cartridge-case, bullet and incident. All queries are assumed to return a set of records that are presented to users as domain-level objects rather than as records, although that set may be empty or may contain just a single item.

Queries are retrieved from the message queues by a layer of middleware that parses the OSL and converts it into one of SQL, SPARQL, SAS, ProcSQL or into an intensional query. The choice of backend language depends upon the nature of the query. Queries for the PostgreSQL database begin with the keyword query, those for the SAS data mining system with SIMILAR and those for intensional with WHAT ABOUT.

The conversion from OSL into a query language gives heavily optimised queries with the minimum effort from users. The following example shows how a simple statement becomes a query across three tables with a series of optimised joins.

```
QUERY ballistic incident WHERE weapon_manufacturer HAS VALUE Sig
Sauer AND victim_gender HAS VALUE female
```

```
SELECT * FROM odyssey.ballistic_incident

     LEFT JOIN ballistic_incident_has_recovered_firearm ON
     (ballistic_incident_has_recovered_firearm.recovered_firearm_oid
     = ballistic_incident.oid)
     LEFT JOIN ballistic_incident_has_recovered_firearm ON
     (ballistic_incident_has_recovered_firearm.recovered_firearm_oid
     = recovered_firearm.oid)
     LEFT JOIN ballistic_incident_has_case ON
     (ballistic_incident_has_case.ballistic_incident_case_oid =
     ballistic_incident.oid)
     LEFT JOIN ballistic_incident_has_case ON
     (ballistic_incident_has_case.ballistic_incident_case_oid =
     case.oid)

WHERE case.gender_of_victim = "female"
AND recovered_firearm.manufacturer = "Sig Sauer";
```

5.2 Hiding the DSL

The Odyssey platform returns results as sets of linked objects. These are displayed in a desktop application. The user is able to see graphs of objects and, by manipulating their properties, can build new queries easily and quickly. Query plans can be saved so that the query can be re-executed later. These plans are simple OSL statements that can be shared between users, for example on email.

The GUI does not present a differentiation between queries intended for the semantic, relational or mining back-ends. Queries are executed across all of the querying systems unless the user edits the OSL to prevent this. Results from all of the back-end systems is integrated into a single graph.

Figure 4. The Prototype GUI

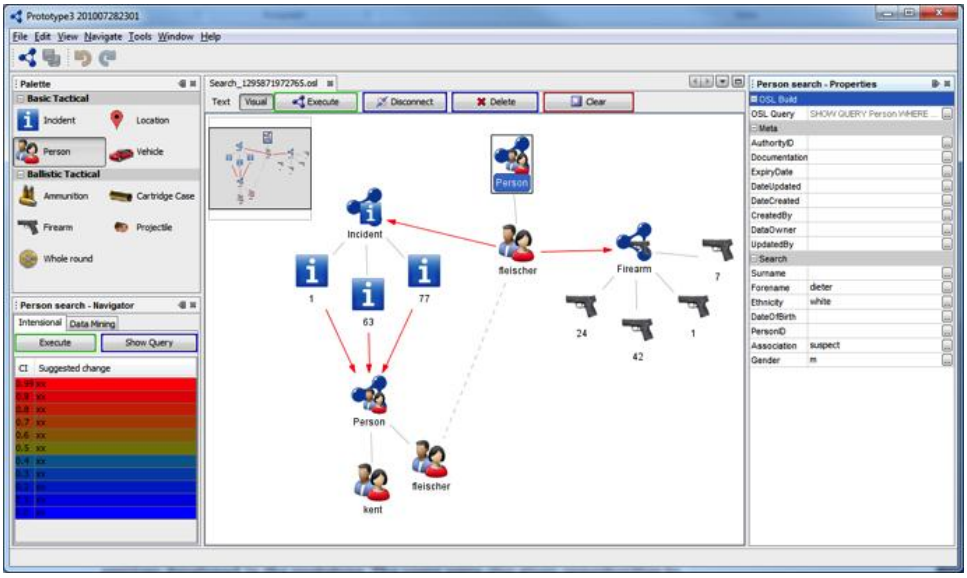


Figure 4 shows how the graphical user interface of the Odyssey platform facilitates search and browsing across the entire crime and ballistic dataset. It takes the full advantage of inductive and deductive approaches so that the end-user can inductively find relevant information and deductively identify values while browsing and narrowing down the possibilities based on the information presented. The interface enables building advanced queries while hiding the complexity of the underlying data structures from the user. The output of the intensional module is shown on the left of the figure. Different colours are used to indicate the strength of association that the module has discovered. The user may choose to modify their query using the changes that are suggested here. Simply selecting a suggestion does this - the GUI automatically re-writes the query for the user.

Presenting result graphs and using them to build new queries is an established GUI technique. The Odyssey project validated the approach through extensive testing with users. The project's validation process included a demonstration of the applications and services developed in the prototype. Users were also given opportunities to interact with the prototype. This allowed the Consortium to review the high level objective of the Odyssey platform, whilst evaluating the Stakeholders continued expectations and needs. In line with the adapted research method the lesson learned during the validation process was elaborated into new set of requirements for the third validation cycle. A third validation process will take place near the end of the project in Year 3 to allow users another opportunity to 'test' the latest version of the prototype and give their feedback on its usability.

6 Conclusion

The Odyssey platform incorporates the use of advanced data mining techniques enriched with semantic technologies. It extracts information from various data sources and indicates how the information will be used next. Moreover, it creates an ontology-driven knowledge repository that enables the analysis of information in a more abstract way, which gives an advantage of being able to illustrate global tendencies or crime patterns. Odyssey platform uses a novel approach for incorporating dynamic user requirements into system realisation (i.e. OSL). The repository is used to operate and investigate real cases using logic reasoning and knowledge interference. Additionally, the platform is able to generate unified graphical results and clearly demonstrate the outcomes of complex analysis. Finally, the platform operates on a very specific domain, which enables the concentration of explicit problems, constantly evaluating outcomes, and suggesting the most promising solution. The platform is set to fill a major gap in the cross-national investigation and security systems. National police forces will be able,

once the platform will be running, to increase their investigation potential by accessing the refined data and graphically represented data patterns. Moreover, the Odyssey platform is structured as a framework that could be easily replicated for other forensic data sets as well as applied to different domains, thus re-defining the standards of information exploitation for large data sets. The latter provides a major millstone for truly integrated and pan-European law enforcement knowledge management Systems.

7 References

Interpol, 2007. Connecting Police: I-24/7. Available at: http://www.interpol.int/Public/ICPO/FactSheets/GI03.pdf. [Accessed February 3rd, 2011]

Europa, 2002. Proposal for a Framework Decision on exchange of information under the principle of availability. Available at: http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/05/367&format=HTML&aged=0&language=EN&guiLanguage=en [Accessed January 25, 2011].

Prüm, 2010. Prüm Decision. Available at: http://ec.europa.eu/home-affairs/policies/police/police_prum_en.htm [Accessed January 25, 2011].

NABIS, 2009. National Ballistics Intelligence Service. Available at: http://nabis.police.uk/database.asp [Accessed January 25, 2011].

Akhgar, B et al., 2009. A Pan European Platform for Combating Organized Crime and Terrorism (Odyssey Platform). In Centeris Conference on Enterprise Information Systems. Ofir, Portugal.

Akrivas, G, Wallace, M, Andreou, G, Stamou, G, Kollias, S, Context-Sensitive Semantic Query Expansion, in *Proceedings of IEEE International Conference on Artificial Intelligence Systems (ICAIS'02. p*p 109. 2002 .

Bonino, D, Corno, F and Farinetti, L, 2004. Domain specific searches using conceptual spectra. In *16th IEEE International Conference on Tools with Artificial Intelligence*. ICTAI 2004, pp. 680-687.

Bundeskriminalant, 2004. Firearm Type Determination. Available at: https://www.forensic-firearms.bund.de [Accessed January 25, 2011].

Chen, H. et al., 2004. Crime data mining: A general framework and some examples. *IEEE Computer*, 37(4), pp.50-56.

Chen, H. et al., 2003. COPLINK: managing law enforcement data and knowledge. *Communications of the ACM*, 46, pp.28–34. Available at: http://doi.acm.org/10.1145/602421.602441.

De Bruin, J. S. et al., 2006. Data mining approaches to criminal career analysis. In *Proceedings of the Sixth International Conference on Data Mining*. Sixth International Conference on Data Mining. pp. 171-177.

Fowler, M and Parsons, R., 2010. *Domain-specific Languages*, Addison Wesley.

Giacomo, G, 1996. Intensional query answering by partial evaluation. *Journal of Intelligent Information Systems*, 7:4, pp 205-233. Published by Springer Netherlands, Nov. 1996.

Jopek, L., Wilson, R., and Bates, C, 2010. An application of a domain specific language facilitating abstraction and secure access to a crime. In *Proceedings of IARIA 2010.* pp 29-33, Lisbon, Portugal, October 2010.

Mernik, M., Heering, J. & Sloane, A.M., 2005. When and how to develop domain-specific languages. *ACM Comput. Surv.*, 37, pp.316–344. Available at: http://doi.acm.org/10.1145/1118890.1118892.

Nath, S. V., 2006. Crime pattern detection using data mining. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. International Conference on Web Intelligence and Intelligent Agent Technology. pp. 41-44

Nygren, F, 2008. The Swedish Initiative. http://www.daten.european-police.eu/2008/nygren.pdf [Accessed January 25, 2011].

Sims, C, 2010. National Ballistics Intelligence Service Update Report. Available at: http://www.west-midlands-pa.gov.uk/documents/committees/public/2010/12_PerfandOps_22April2010_National_Ballistics_Report.pdf.

Smith, M. and Tilley, N. 2005. *Crime Science: New Approaches to Preventing and Detecting Crime*, Portland, USA: Willan Publishing.

Strohmaier, M., Kröll, M. & Körner, C., 2009. Intentional query suggestion: making user goals more explicit during search. In *Proceedings of the 2009 workshop on Web Search Click Data*. WSCD '09. New York, NY, USA: ACM, pp. 68–74. Available at: http://doi.acm.org/10.1145/1507509.1507520.

Vallet, D, Castells, P, Fernández, M, Mylonas, P, and Avrithis, Y, 2007. Personalized Content Retrieval in Context Using Ontological Knowledge. *IEEE Transactions On Circuits And Systems For Video Technology*, 17:3. MARCH 2007.

Wilson, R., Jopek, L., and Bates, C, 2010. Sharing Ballistics Data across the European Union. In *Proceedings of IARIA 2010.* pp 8-13, Lisbon, Portugal, October 2010.

Yates, S., et al., 2009. Semantic Interoperability between Ballistic Systems through the Application of Ontology. In *IADIS WWW/ Internet Conference*. pp. 153-157.

Yu, L., 2008. Prototyping, Domain Specific Language, and Testing. *Engineering Letters*.