An Application of a Domain-Specific Language Facilitating Abstraction and Secure Access to a Crime and Ballistic Data Sharing Platform

Lukasz Jopek
Cultural and Computing Research Centre (C3RI)
Sheffield Hallam University
Sheffield, United Kingdom
l.jopek@shu.ac.uk

Richard S. Wilson
Cultural and Computing Research Centre (C3RI)
Sheffield Hallam University
Sheffield, United Kingdom
r.wilson@shu.ac.uk

Christopher D. Bates
Cultural and Computing Research Centre (C3RI)
Sheffield Hallam University
Sheffield, United Kingdom
c.d.bates@shu.ac.uk

***Abstract*—Crime investigation requires controlled sharing, secure access and formalised reporting on heterogeneous datasets. This paper will focus on encapsulating data structures and services, whilst exposing abstraction, relevant only to the End-User through the application of a domain-specific language. The language is used for all interactions with the platform, enabling non-technical users to build complex queries. The language also increases the platform's security, by hiding the internal architecture of services and data structures. This solution has been demonstrated to law enforcement communities across Europe as a prototype crime and ballistic data sharing platform.**

Keywords-standardisation;data structures, domain-specific language; law enforcement; public services

I. INTRODUCTION

Odyssey is a research project looking at the difficulties in sharing information about gun-crime between Police forces across Europe [1][2]. The project is building a technology demonstrator to prove the benefits of integrating and sharing ballistic information from states across the European Union. Such sharing will support gun crime investigations and prevention activities where cross-border trafficking of weapons or ammunition is involved. The prototype will enable interoperability between existing systems and ensure compatibility with emerging European standards. The Odyssey Consortium consists of law enforcement agencies such as Europol and NABIS, the National Ballistic Intelligence Service (United Kingdom) which brings standards, broad knowledge and hands-on experience to the project. The Consortium is committed to informing the law enforcement community with the outcomes of its research, standardising data formats, integrating processes and finally prototyping a crime and ballistic data sharing platform.

Currently, law enforcement agencies across Europe rely heavily on in-house systems, which often cannot interoperate on either national, regional or departmental levels [3]. Integration, secure sharing and the use of data is highly limited due to technological boundaries; hence the process is frequently manual and costly [3][5]. The Odyssey project is seen not only as an opportunity to facilitate communication between different agencies, but to also build a common European understanding over current and future needs. The number and diversity of existing technologies, the lack of European ballistic standards and the high reliance on ballistic experts puts Odyssey into a very interesting perspective, whereby new standards, technologies and processes are expected to support current best practices. Technologically, the Odyssey project aims at prototyping a scalable platform, which is interoperable with legacy systems, processes and expertise.

In this paper, we will focus on the process of building a common interface for the system, encapsulating data structures and exposing only those abstractions that are relevant to each user. We will further explore an innovative application of a domain-specific language (DSL) in an area in which one has not been used before. The DSL is used to express every interaction within the system from getting data and defining sharing permissions, to integrating security by hiding internal structures of the platform.

II. DESIGNING A DOMAIN-SPECIFIC LANGUAGE

Domain-specific languages (DSL) express complexity at a particular abstraction tailored to both current and future needs [6]. 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 [7].

A well-designed DSL complies with certain objective qualities. A language perceived as simple, easy, and effective is more likely to be widely accepted, even though these qualities can be subjective. However, these merits are not enough for a good, stable language. The definition and measurement of these qualities will vary in time and from one person to another. Everyone in a team developing a language can agree on where the language meets their expectations, but only when exposed to a wider community for a longer period can the quality of a language be measured. In the world of engineering, users have varied skills, preferences and needs, while usually the primary goal of designing a system is to identify base requirements and build a platform that works.

In the Odyssey project, a DSL was introduced to express the user requirements and solutions in a particular domain. A DSL promotes decoupling between components, modules and software stack layers, making the platform easily extendable and its components highly reusable. In the Odyssey project, a DSL language is used not only to convey the complexity of the domain, but also to facilitate and unify the entire communication across the platforms' components and users (Figure 1).

 ISBN: 978-1-61208-112-0

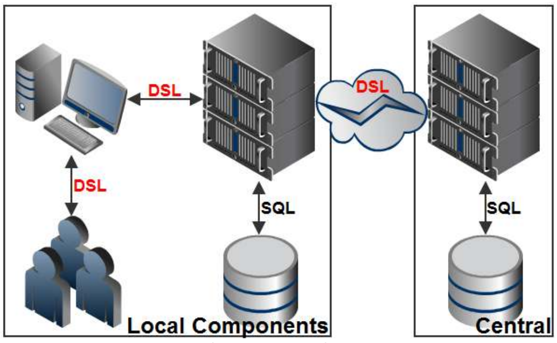

Figure 1. Role of DSL in platform's architecture

A DSL is used to express queries further translated into for example SQL. In the Figure 1, a DSL query is presented to query data from distributed databases by executing the exact statement on separate repositories. The Odysseys DSL enforces unification of queries, which enables standardisation of the results formats and therefore further reasoning. The DSL engine has the capability of retrieving and merging information from diverse data sources by transforming the query into SQL-like languages, but also integrating with for example data-mining result sets or applying reasoning using a built-in ontology.

The particular application of a DSL enhances security by providing information relevant only to the end user encapsulating data structures, abstracting services, and facilitating data manipulation. The Odyssey system enables sharing information without exposure of its structures, providing the requested up-to-date, accurate, relevant and easy to understand information.

Additionally, users are supported with tools to build and visualise DSL statements and their executions, which are also represented in the DSL.

III. THE RESEARCH CONTEXT

The Odyssey project delivers both a user interface and a domain specific language to users who have diverse, but highly specialised expertise. Their high-level cognitive abilities often relate to discovering facts in crime investigations, whilst dealing with uncertainty and different types of abstraction. A mixture of experience from former investigations, combined with an understanding of human nature and circumstances, enable investigators to reason, make decisions and act on them. Dealing with this type of situation requires an entirely different set of skills and builds an impression that the situational information with knowledge and experience is enough to solve any problem. The Odyssey project is committed to fulfilling the end user's functional requirements and expectations, whilst proposing new functionalities, which are not available in current systems due to a lack of data sharing and information exchange abilities.

Odyssey's DSL compromises between the expressiveness of both a formal and flexible semantically-enhanced language. Complexity of syntax was greatly reduced by identification and categorisation of use case scenarios, grouping of functionalities and abstraction of data sources. Moreover, in contrast to, for example SQL, the user does not need to be aware of underlying data structures, nor the platform's architecture, to enable a complex search to be undertaken. The user is asked to define the information of interest and the constraints by which the data will be filtered and sorted. In general, the user queries the system by defining the outcomes and under what conditions. One of the key requirements for the language is to facilitate access to factual information, but without taking the risk of misleading an investigator by presenting non-related data. The system reveals opportunities to the end user by facilitating discovery of new facts and collaboration on possible scenarios. Ultimately, we have identified three main qualities integrated to our solution. They are shown in Figure 2.

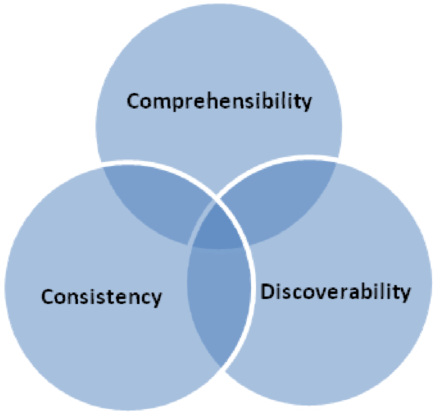

Figure 2. Principles of a scalable DSL design

Comprehensibility: Communication is pivotal to design domain-specific languages. In the domain of law enforcement agencies, a comprehensive, self-explanatory language acts as a bridge between a platform and a non-technical end user. A consistent and well-established syntax builds trust and guarantees time invested in mastering functionality will be applicable across other aspect of the platform in the present and future releases. Furthermore, we identified discoverability as the next most important characteristic of a well – designed language, which in this case means the ease of discovering features based on what we already know and the tools provided.

Consistency: The language and its controlled syntax encapsulate functionality, architecture and may even determine the entire system's performance, by for example, optimising queries and merging results. These algorithms are developed for abstracted use case models and a user is not allowed to make any modifications or optimise queries per case. The great advantage at this stage was the active involvement of the user community (especially West Midlands Police, United Kingdom), who carefully gathered both user requirements and developed an iterative process of language evaluation. Each partner in the Odyssey project has a different perception of the problems we are addressing and has contributed to the design of the platform in different ways. Sheffield Hallam University representatives visited Northern Ireland Police and West Yorkshire Police in England identifying needs and getting a hands-on experience of the current state of the art crime and ballistic ICT systems.

Discoverability: In addition, these visits provided an insight into daily activities and processes, the Odyssey project is dedicated to improve. Satisfied at this stage, the Odyssey Consortium proposed a language that would enable the modelling of crime investigations and play a

 ISBN: 978-1-61208-112-0

key role in creating data access rules and the enhancement of security to the entire platform. The user requirements stated explicitly that we should develop a language which provides a simplified method to retrieve data and to operate on datasets, but to also facilitate access to services, enable management of users' roles and maintain data sharing rules. We require a language that would facilitate every interaction, with the entire system, in a controlled, structured and concise way.

IV. ODYSSEY SEMANTIC LANGUAGE

We propose the development of a syntax and a semantic language which supports modelling of active crime investigations by operational detectives that will link general crime 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. Furthermore, it was developed in an open-source grammar development environment, ANTLR. A structural Piggyback [8] design pattern was introduced to facilitate transparency between languages and services being a hosting base to which the DSL is translated. The DSL was designed on the top of a SQL bearing in mind abstraction, data mining and process modelling capabilities.

The ANTLR output is further integrated with the NetBeans Rich Client Platform (RCP) and the combination produces a fully-featured editor that seamlessly integrates with the graphical representation of search and results. The features correspond syntax colouring, error highlighting, code completion, etc.

The example below (Figure 3) presents a query expressed in Odyssey Semantic Language (OSL) that retrieves firearms with a twenty-two calibre (0.22 inch):

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

Figure 3. Selecting a firearm of a specific calibre

With a very similar query structure we can apply sharing rules on a set of data (See Figure 4 below).

ALLOW firearm **WHERE** calibre **HAS VALUE** 0.22

Figure 4. Applying sharing rules on a dataset

In the diagram below, we introduce a few entities from the Odyssey database structure that will be used in the next example to present how OSL abstracts and simplifies access to relational datasets. The database itself contains over 50 tables to model crime and ballistic evidence or retains user accounts and their roles. The diagram shows how a location, a central concept of the database structure, can be linked to a firearm. A location is only one of an incident's characteristics; other descriptors include documents or other related incidents. An incident effectively links locations with ballistic items that are a generalisation of firearms, cartridge cases, bullets, ammunition, projectiles, and other categories of ballistic and crime evidence.

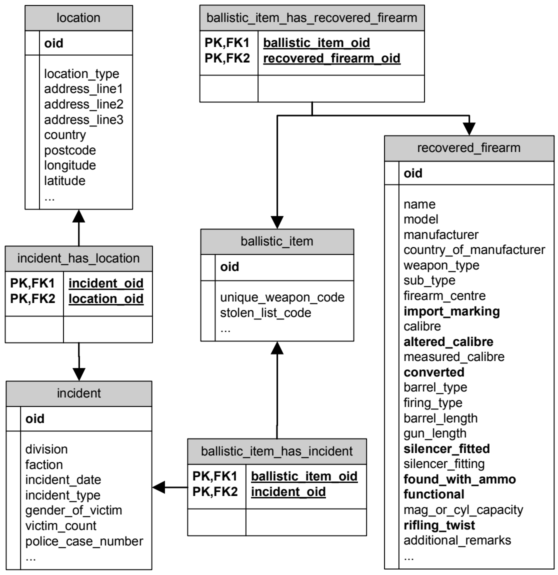

Figure 5. A partial E-R diagram presenting how a firearm and a location can be linked together

In Figure 6, we show how easy and straightforward it is to build a complex SQL-like JOIN across multiple tables from the above using the OSL. In fact, the task is almost effortless and does not require any understanding of the structure. The user does not need to be aware of a number or even classes of the tables that require joining.

QUERY firearm location **WHERE**
calibre **HAS VALUE** 0.22 **AND**
country **HAS VALUE** United Kingdom

Figure 6. Selecting a firearm linked to a location

What the user is asked to specify are concepts representing types of data and the constraints wants to apply onto the dataset. Hence, joining tables and merging resources is performed without the user's attention. In addition, what the OSL implementation enables is the integration of results from various data sources and services. This means a user can perform even more complex tasks with very similar effort; for example overlaying data with data-mining results. This level of abstraction creates a very powerful environment for non-technical users interacting with the system.

In contrast to Figure 6, the example below is of the similar expressiveness, but represented in a pure PL-SQL. According to the E-R diagram from Figure 5 the query would look like this:

SELECT rfa.oid, rfa.*, loc.oid, loc.*
FROM recovered_firearm rfa
LEFT JOIN ballistic_item_has_recovered_firearms bit_rfa **ON**
(bit_rfa.recovered_firearms_oid = rfa.oid)
LEFT JOIN ballistic_item ba **ON**
(ba.oid = bit_rfa.ballistic_items_oid)
LEFT JOIN ballistic_item_has_incidents bit_inc **ON**
(bit_inc.incident_oid = ba.oid)
LEFT JOIN incident inc **ON**
(inc.oid = bit_inc.incident_oid)
LEFT JOIN incident_has_locations inc_loc **ON**
(inc_loc.incident_oid = inc.oid)
LEFT JOIN location loc **ON**
(loc.oid = inc_loc.location_oid)

WHERE rfa.calibre = 0.22 **AND** loc.country = "United Kingdom";

Figure 7. SQL representation of the example

In the project, we prototyped a standalone client based on NetBeans RCP, enriched with visual features of the embedded Visual Library. This implementation fully supports the OSL and provides a graph-based visualisation facilitating search, browsing and what is more, reuse of search results in further investigations. We provide a user with a set of functionalities to visually manage multiple searches and results on a single screen at the same time. Besides, there is a text view available to the user, which is a document-based representation of graph content that seamlessly integrates with the visualisation.

Figure 8 below illustrates a search, whereby a user is looking for incidents linked to at least one of a previously identified person, under conditions such as location, timeframe, or an incident type.

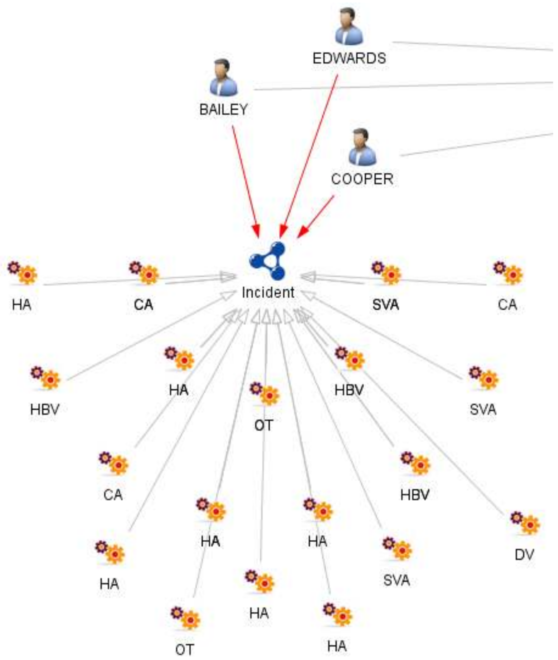

Figure 8. Visual representation of search and results with dependencies set between previous results and search

There are two types of widgets, with the first one used to build an OSL search query (labelled "Incident" in the example above) by setting properties of the class of data and dependencies from previous results (in this example, representing persons). The second type of widget represents results, which are data entries in the system, for example persons and incidents in the diagram above.

This graphical information retrieval and visualisation tool was introduced to guide and assist the end user in building, editing and executing OSL queries. The client provides a windows-based user interface that the end user is familiar with. Moreover, it offers a visualisation which is used to search, browse, but also receive alerts on new entries or updates in the database.

V. FUTURE WORK

Formalisation and standardisation of the OSL specification is one of the key areas which will be addressed after the language is presented to the user community and approved by the Odyssey Standardisation Committee.

Additionally, we will be exploring further possibilities of data mining techniques in order to extract and index information for further processing and analysis. Existing technologies recognise a vast potential in text-mining of personal statements and other plain-text documents gathered during an investigation. This would lead to further modification and new extensions to the language, which could not be addressed in this paper. For example, text-mining could result in entities such as person, location and vehicles extraction and mapping of the results on a timeline for further investigation and incident sequential analysis.

The OSL is currently under development and at this stage does not entirely cover all of the user needs and requirements; ideally, the user would be able to model processes and sequences of events that lead to or follow a crime. This is not a usual use of a domain-specific language and it might even seem to contradict with common practices. In general, DSLs model per-case solutions and do not explore the benefits of sequencing actions, events or outcomes. Therefore, usual DSL implementations are limited to a linear communication with a system rather than enabling the user to reason on data and automate the interaction with a system. Currently, the platform does not model nor visualises sequential data, but the need was widely discussed with the end user community. Furthermore, a solution based on mapping of crime and ballistic incidents on a timeline was proposed.

In summary, the future work will focus on formalisation and standardisation of solutions and practices described above, such as the DSL and processes the language is compatible with. Moreover, we will also investigate the potential of text-mining in the domain of crime investigations, which could potentially lead to changes in the OSL. Finally, we will research on how sequential data can be used of benefit and expressed in the OSL, in order to enable modelling of crime investigation processes and modelling of crime cases as such.

VI. CONCLUSIONS

The Odyssey project key result areas are the standardisation of data collection, storage and sharing, the facilitation of interoperability between existing systems and the provision of an infrastructure to both securely collaborate on cross national investigations and also extracting information through various data-mining and knowledge extraction techniques. These objectives of the Odyssey project lead to cost saving and increased efficiency, but also promote collaboration between law enforcement agencies across Europe through the use of information and communication technologies (ICTs).

In this paper, we presented how a domain-specific language can facilitate access to a platform by encapsulating data structures, enhancing security, but most importantly, enabling a non-technical user to interact with the platform through the use of a language suitable for the field of expertise.

We have designed a language according to the user requirements and prototyped a platform that makes the full use of its features. The OSL is used to access and manipulate data from multiple sources, collected by various techniques and of different investigation value. Additionally, the OSL manages access to the user permissions and the sharing of data. The presented solution enables the end user to interact with the platform seamlessly switching between the OSL text- and the graph-based editor.

[1] Chau, M., Atabakhsh, H., Zeng, D., and Chen, H. (2001), 'Building an Infrastructure for Law Enforcement Information Sharing and Collaboration: Design Issues and Challenges', University of Arizona

[2] Su, S., Fortes, J., and Kasad, T.R. (2005), 'Transnational Information Sharing, Event Notification, Rule Enforcement and Process Coordination', International Journal of Electronic Government Research (IJEGR), pp. 52-62, 2005

[3] Travis, J. (1998), 'Informal Information Sharing Among Police Agencies', National Institute of Justice, December 1998

[4] Redmond, M. and Baveja, A. (2001), 'A data-driven software tool for enabling cooperative information sharing among police departments', European Journal of Operational Research, pp. 660-678, June 2001

[5] Chen, H., Schroeder, J., and Hauck, R. (2002), 'COPLINK Connect: information and knowledge management for law enforcement', University of Arizona, Decision Support Systems, pp. 271-285, 2003

[6] Yu, L. (2008), 'Prototyping, Domain Specific Language, and Testing', Engineering Letters, February 2008

[7] Mernik, M., Heering, J., and Sloane, A. (2005), 'When and how to develop domain-specific languages', ACM Computing Surveys, pp. 316-344, December 2005

[8] Spinellis, D. (1999), 'Notable design patterns for domain-speci®c languages', Department of Information and Communication Systems, University of the Aegean, pp. 91-99, December 1999

ISBN: 978-1-61208-112-0