Beyond Search in Policing

How Oracle Redefines Real-Time Policing and Investigation:

Complementary Capabilities of Oracle’s Endeca Information Discovery and

WCC’s ELISE
Disclaimer

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.
Executive Overview

Leveraging technology to search and correlate crime data is an absolute necessity in the modern era of *Intelligence-Led Policing*. It plays a key part in Oracle’s own Integrated Policing Platform solution for law enforcement organizations.

*Search* is such a common term, that using it opens the door for ambiguity and confusion over what capabilities are indeed available. In this context, *search* represents a category of capabilities that may range from simple lookups, to sophisticated pattern matching, to analytics-based, weighted searches. In large and complex data environments, multiple search technologies can work in concert to pinpoint results faster and more accurately.

*Beyond search* is offered by combining Oracle’s Endeca Information Discovery with WCC’s ELISE Smart Search & Match to deliver *real-time policing* and *investigation* in Oracle’s solution for law enforcement – the *Integrated Policing Platform*. 
Introduction

Search Evolution

Search has become pervasive to nearly every computer related task we perform. It’s easy to forget that before Google came along, we used to find information through lists and hierarchies of categories. As long as you understood the categorization that someone else had defined, you would eventually find what you were looking for. We now embrace Search Engines, and life would be much harder without them.

It’s not surprising therefore, that search is now an important feature in many business applications. However, to describe every application that provides search as a Search Engine would be to undersell the real value of that application. Google, for example, would actually be better described as a massively scalable indexing and page-ranking engine, designed for the sole purpose of helping people find web sites. You need to look at the full set of capabilities alongside the use case they were designed for to decide the suitability of a search enabling application.

Integrated Policing Platform

Oracle’s Integrated Policing Platform (IPP) is a solution that provides a complete set of functionalities to support the full range of policing operations. The ability to perform searches across different datasets is a core capability of the platform. WCC is a key Oracle partner for the IPP solution with the integration of WCC’s ELISE search technology into the solution. Oracle’s recent acquisition of Endeca with its search technology has led to some confusion about the role of both products in the IPP solution.

Purpose of this Document

The purpose of this document is to look beyond the basic capability of search and to understand how both WCC’s ELISE and Oracle’s Endeca Information Discovery support the requirements of a modern policing platform and their role in IPP.
Policing

Let’s take a moment to consider the challenges faced by policing organizations today. ICT\(^1\) in policing is nearly always fragmented. Specific systems to perform specific functions have grown organically over a number of years so that now it is not uncommon for a policing organization to rely on hundreds of siloed applications to support their operations.

**Police Officers** need to access multiple systems to find answers to commonly asked questions, such as:

- Who is the registered owner of the vehicle caught on CCTV?
- Who do we know that has physically identifying marks similar to those described by a witness?
- Who do these partial fingerprints retrieved from a crime scene belong to?
- Who else can I put in a photo identity line up to see if the victim can identify our suspect?

Information is often missed because:

- It either takes too long to access each system, or
- It is too hard to combine manually with existing information, or
- Information is incomplete, or
- Information is (partially) incorrect.

The police officer on the ground responding to incidents needs to know answers to these questions quickly (in real time) and reliably (to be trusted).

The **Crime Investigator** has more opportunity to develop lines of inquiry and to form a complete picture from all the available data. Forming hypotheses and looking for connections and patterns in the data collected.

Similarly **Policing Management** will take a longer-term view on crime trends and how they are using their resources. They have a need for business intelligence and analytics showing charts and reports to support their decision-making. Increasingly the data from social media is playing an important role in understanding crimes and criminal activity, but the volume and unstructured nature of that information makes it harder to combine with their existing data.

All of these scenarios represent use cases that are driven by smart applications of search.

\(^1\) ICT – Information and Communications Technology
Smart Search Products

WCC’s ELISE

ELISE is the search product from Oracle Gold Partner, WCC. ELISE can store and search any type of data, but WCC’s specialty is in People Centric Search. They recently won the MITRE Challenge\textsuperscript{2} for multicultural name matching, outperforming all established players in that sector.

ELISE is essentially an engine into which you pour data, and run matching algorithms against an API. ELISE has no official user interface of its own. It was designed so that it would be integrated within another user facing application through an API.

ELISE uses a flexible object model to store data of any type, typically these can include

- Biographic Data
  Standard attributes including names, age, height, weight, hair color, ...

• Biometric Data
  Facial scans, Fingerprints (up to 10 per person), Iris scans, DNA, Voice recordings, ...

• Contextual Data

Searches are performed in ELISE by combining multiple matching algorithms with rankings and weightings to return the best result, referred to as **Proximity Matching**. Matching algorithms are interchangeable so that customers can adopt new, improved and/or more affordable matching algorithms as they become available.

ELISE is designed for very fast operation with sub-second (typically milliseconds) response even on very large datasets (millions to hundreds of millions of objects) with very complex objects.

![Figure 2: ELISE Matching Process](image)

WCC have a great deal of experience in selecting and tuning the most appropriate algorithms depending on the search requirements. It is not uncommon for some algorithm implementations to be tuned for performance whilst others are tuned for accuracy.

Through ELISE, WCC can combine both in a single search, in a process they called Intelligent Fusion, so that the fastest algorithm narrows the possible results and then the more accurate algorithm is applied to the shorter result set. The end result being both a fast and accurate result set.
Oracle’s Endeca Information Discovery (EID)

Before the acquisition by Oracle, Endeca’s background was in providing search capabilities for eCommerce web sites. That functionality lives on in Oracle Endeca Guided Search and is being tightly integrated with Oracle ATG Commerce product family. Oracle Endeca Information Discovery (EID) (formerly Endeca Latitude), uses the same core storage engine and search capabilities as before, but is now combined with portal and integration tools to enable user analytics. EID is now a key part of the Oracle Business Analytics product portfolio. It is positioned as the tool for rapid, intuitive exploration and analysis of data from any combination of structured and unstructured sources.

There are three major pieces to Oracle EID:

- **MDex Storage Engine**
  The storage engine uses a very simple and flexible data model with data about each entity being held in one very wide row. Data items are often denormalized and allowed to repeat for performance. The result is that the need for complex Extract-Transform-Load (ETL) is removed, even when data is combined from multiple sources. It also makes deployment very rapid as the need for very complex Extract-Transform-Load (ETL) processes is removed. Searches performed on the data are returned along with “Facets”, a summary of all the additional attributes that match the search results. It is these facets that enable a user to keep refining their search and to explore the data. Data can also be aggregated and joined using a powerful SQL-like query language.

- **Integration Suite**
  This suite of tools enables the rapid unification and augmentation of diverse sources of
information. It includes (1) the Content Acquisition System for gathering content from file systems, content management systems, and websites, (2) the Integrator which provides out-of-the-box ETL tooling for integrating and enriching enterprise data, and (3) the open Web Services API which allows for direct data integration from other tools, such as Oracle Data Integrator and Hadoop. It also provides integration components for analyzing unstructured blocks of textual data to perform Entity Extraction, Sentiment Analysis and Document Summarization, thereby enabling the combination of structured and unstructured data analysis.

- **Discovery Portal - Studio**
  A web based user interface for end users to be able to interact with the data. The portal is supplied with a number of UI components that are EID aware for presenting data in many different ways, including charts (bar, pie), maps, tag clouds and tables. There is also a component for displaying and enabling user interaction with the facets returned with a search.

With EID Search is used to find an entry point into to data and from there the rest of the analytic capability is used to actually understand that data.

Large police departments are combining data from Social Media with existing departmental data to identify hot spots, correlate deployment of thousands of police officers, initiate operational orders and retrospectively identify potential witnesses. One such proof-of-concept was created using data from the Chicago Police Department during the May 2012 NATO Summit in that city. As well as the 7000 visiting dignitaries, the summit brought with it tens of thousands of protestors. The demo shows the use of EID and its unstructured data analysis capabilities – melding Social Media information with departmental 911 and force deployment information. With this analysis, officers could be rapidly redeployed to preemptively address evolving threats. This scenario was featured in Oracle President Mark Hurd’s keynote presentation at Oracle Openworld.³

Beyond Search

Introduction

Both ELISE and Endeca have search as a core capability but both use search to enable very different types of functionality. To decide which tool is the right one for a particular use case, you need to consider each product’s broader functionality and categorize them differently. The diagram below shows the key differences in functionality beyond search.

Operational Matching

WCC’s ELISE is aimed at the systematic and automated matching of structured objects. The knowledge and skills required to rapidly match multiple complex data objects is built into the ELISE deployment so that it can be used by a non-expert user or by other application system components. It can therefore be described as both an Expert System and Matching Engine. In summary, the expertise of the experts is implemented in matching rules to allow non-experts to perform a simple search delivering expert results.

In policing, this relates to the frequently used, operational type of searches performed by police officers - such as person identification by physical descriptions and fingerprints, or vehicle searches by partial registration numbers and vehicle descriptions.

Discovery & Analytics

Endeca is aimed much more at user driven scenarios where there is a combination of structured and un-structured data sets, from which someone is trying to gain understanding. It enables someone with the right knowledge and skills in that domain to gain greater insight from that data. The name Information Discovery describes this usage perfectly, as would investigate and explore.
From a policing point of view, this relates well to use cases of crime trends analysis and to complex and organized crime investigation, which might be performed by trained investigators or analysts.

Combined Use Case

It’s clear then that Oracle’s Endeca and WCC’s ELISE support quite distinct use cases. However there are cases where the two products would work in perfect concert.

Consider a scenario where a robbery has been committed and only a partial fingerprint could be taken from the scene of crime. In this case ELISE’s matching algorithms would still be able to create a short list of people whose fingerprints share characteristics of the partial print. This search could be performed by a regular officer at the crime scene using a mobile device. The results could be passed back to an investigator and would form a short list of suspects. The investigator is then able to focus the rest of the investigation around a much smaller number of suspects. The investigator would use Endeca to analyze previous crime reports, witness statements and social media to discover if there is information linking any of these suspects to that crime.

Figure 5: WCC ELISE and Endeca EID used in concert
Conclusion

Whilst both WCC’s ELISE and Oracle’s Endeca Information Discovery are generally described as Search tools, it should now be clear that both tools provide a lot more functionality than just pure Search. It should also be clear that both have unique capabilities that address quite different Policing use cases and scenarios. Both WCC’s ELISE and Oracle’s Endeca Information Discovery have key roles to play in the IPP solution. The combination of both these products in the IPP solution is a clear differentiator against other solutions and provides real value to our customers.

Figure 6: Oracle’s Integrated Policing Platform – Functional Architecture
Contact Information

Oracle

Hong-Eng Koh  Industry Lead  Email: hong-eng.koh@oracle.com
David Shepherd  Industry Expert  Email: david.shepherd@oracle.com
Martin Jarvis  Industry Solution Architect  Email: martin.jarvis@oracle.com
Martin Walters  Endeca Expert  Email: martin.walters@oracle.com

WCC

Tarvinder Sembhi  Industry Lead  Email: tsembhi@wcc-group.com
Mark Wegman  ELISE Expert  Email: mwegman@wcc-group.com