

# Semantics In Action For Proactive Policing



**Jen Shorten**

Technical Delivery Architect, Consulting Services



**Jon Williams**

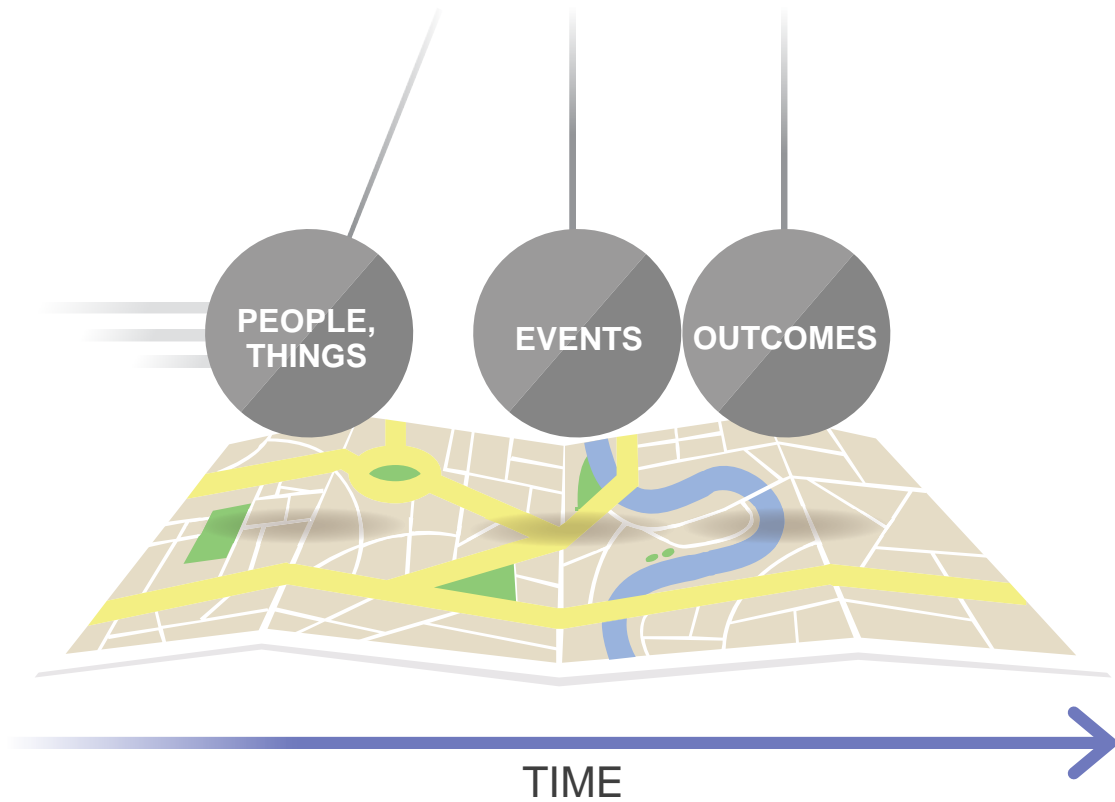
Senior Sales Engineer, UK Public Sector

The background image shows the interior of a police vehicle. In the foreground, a hand is pointing at a small, rectangular screen that displays a green-tinted image, possibly a map or a video feed. The screen is mounted on a dashboard or control panel. To the left of the screen, there are several buttons and a vertical strip of controls. To the right, there are more buttons and a coiled cable. The overall lighting is dim, with some light coming from the screen and the ambient light of the vehicle interior.

# The Nature of Policing Is Changing

“The increasing availability of information and new technologies offers us huge potential to improve how we protect the public. It sets new expectations about the services we provide.”

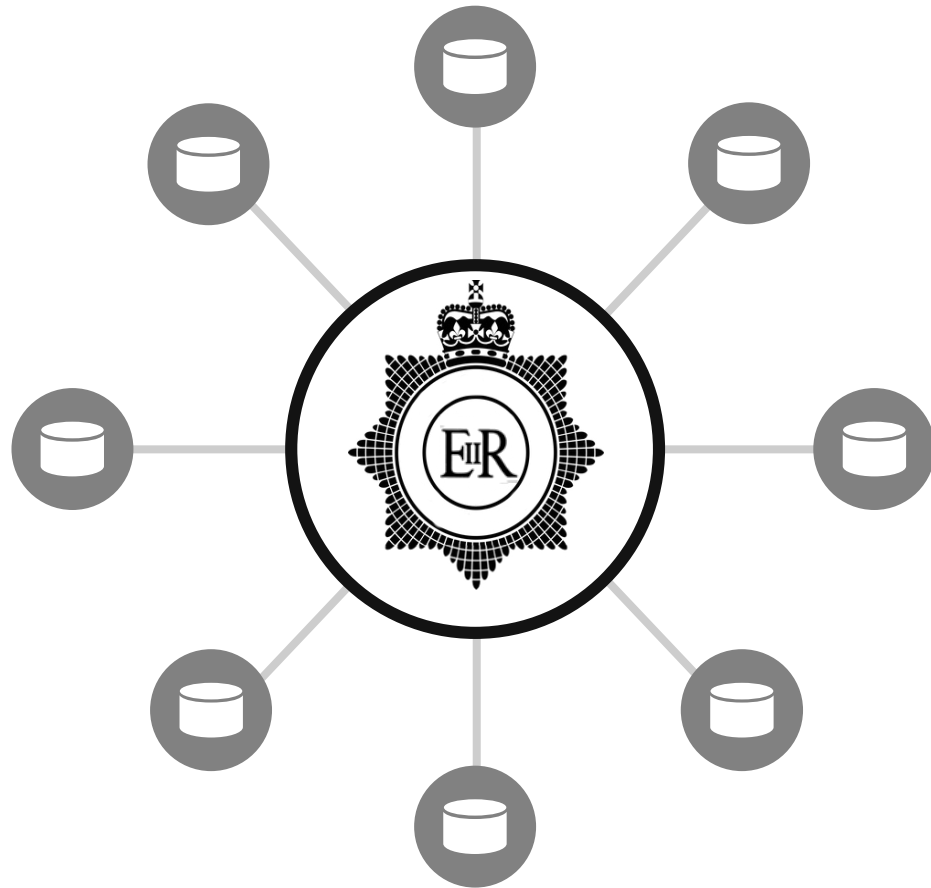
Police IT systems need to adapt to keep up with those changes



NATIONAL POLICE OBJECTIVES

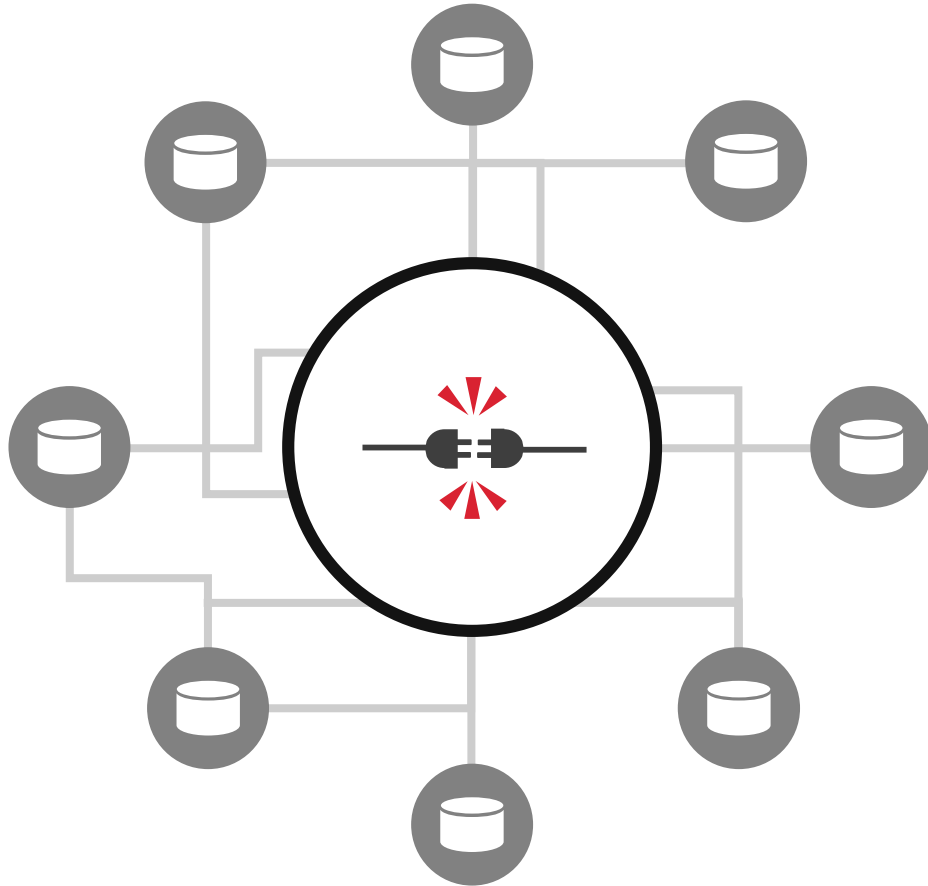
# Digital Transformation of Policing

- Proactive
- Impact led
- Outcomes driven
- Data driven



WHAT POLICE NEED TO DELIVER THAT VISION

## A Unified, Actionable 360 View of Data



THE REALITY

## Data Is In Silos

- Data is spread across disconnected databases
- Data quality issues are significant
- Data collection is a slow manual process

# Current State

SP-20081104-0009

ASSIGNED

MS EMILY SURNAME

MS EMILY SURNAME 23

HIGH ST

RUGELEY

VIOLENCE AGAINST THE PERSON

C10/A

RESOLV

VEH- THEFT OF MOTOR VEHICLE

C50

VEHICLE PURSUIT

P15 (S)

VIOLENCE AGAINST THE PERSON

C10/A

VIOLENCE A

WITNESS I

999

GUARD

Comm

Person of Interest Profile

Name(s):  
First Middle Last (system name)

DOB:

AKA:  
First Middle Last (system name)

SEX:

Address History:

Type	Address	Source/Date
Current Home	Address Eastings/Northings	Source system, Recorded date
Previous Home	Address Eastings/Northings	Source system, recorded date
Contact At	Address Eastings/Northings	Source system, recorded date

[Instructions: ordered by recorded date, newest at the top]

FLUTTE, ADAM JOE

File Edit Search Reports Tools Help

DEL VIEW LIST TOTL

1) DL XNAME ADDR(2)

TIER 1 KNOWN TO

FIRST Adam

AD

ZIP 79134

TOTAL ALIASES: 1 (SEE INVOLVE)

SSN 529-57-8997

FBI FBI9887732

OTHER TEL (234)676-6997

INTERNET

AREA.. NAT. TA

ON RD 5-6 HF 6

SU-L

W-M

AB #

-CH

HWY-

/SAL

external

source Act.

ATION WAY NE-

Names Table

Description

Height Weight Build

Hair colour Eye colour

Marks/tattoos

Type Location Size

Classification Desc

Clothing

Type Desc

Colour Material

Make Style

Features Common name

Median - Microsoft Excel

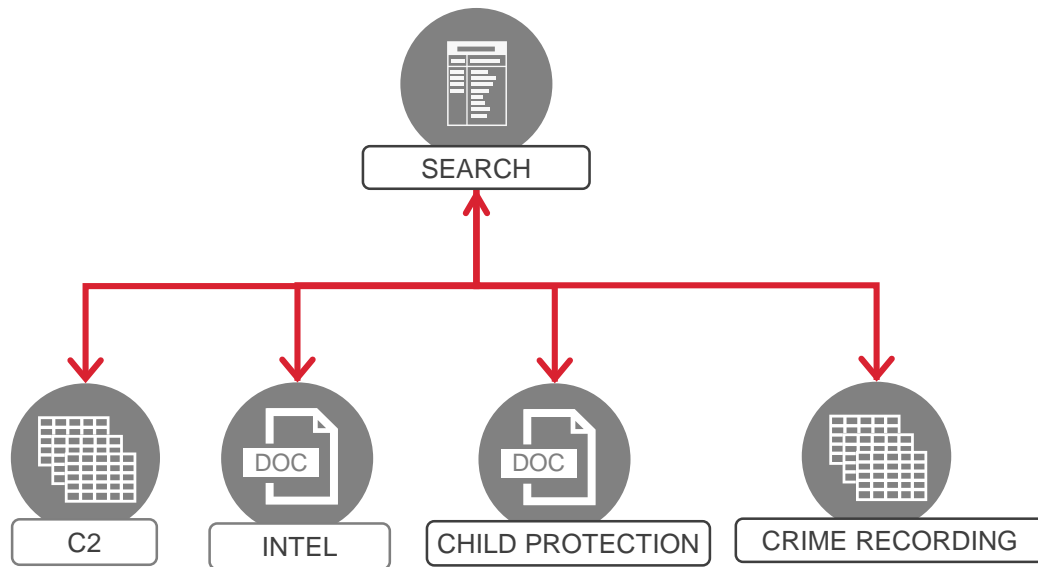
File Home Insert Page Layout Formulas Data Review View Developer

Clipboard Font Alignment Number Styles Cells

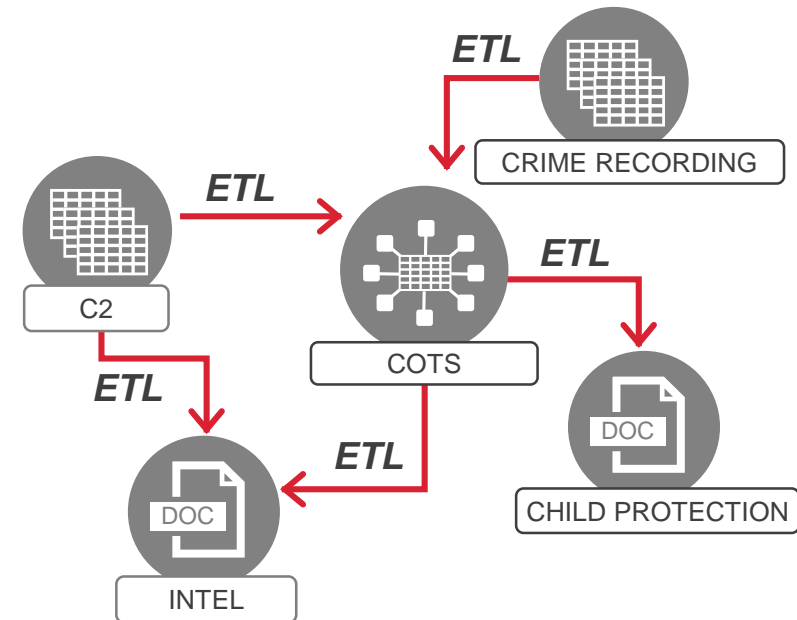
Arrest DOW

A	B	C	D	E	F	G	H
1 Arrest DOW	Arrest	Fingerprint	Days				
2 Sunday	2015/02/21 05:18:00	2015/03/04 09:40:00	11				
3 Sunday	2015/02/20 23:05:00	2015/03/03 10:20:00	10				
4 Saturday	2015/02/20 16:16:00	2015/03/06 11:00:00	13				
5 Friday	2015/02/20 10:48:00	2015/03/10 13:00:00	18				
6 Friday	2015/02/20 02:53:00	2015/03/05 08:40:00	13				
7 Friday	2015/02/20 01:52:00	2015/03/02 11:00:00	10				
8 Friday	2015/02/20 01:44:00	2015/03/10 15:00:00	18				
9 Friday	2015/02/18 20:57:00	2015/03/06 10:00:00	15				
10 Friday	2015/02/18 16:27:00	2015/03/06 08:20:00	15				
11 Wednesday	2015/02/18 04:35:00	2015/03/11 10:00:00	21				
12 Wednesday	2015/02/17 20:16:00	2015/03/03 14:20:00	13				
13 Wednesday	2015/02/17 09:40:00	2015/03/05 13:00:00	16				
14 Tuesday	2015/02/16 19:37:00	2015/03/03 14:00:00	14				

# Traditional RDBMS Solutions



**FEDERATED SEARCH**



**COTS PRODUCTS**

# Federated Search

---

CRIME REFERENCE IZ/000107/08  
CRIME REPORTED Inflicting bodily injury with or without weapon  
INCIDENT DATE 11/02/2016

SCENE OF CRIME  
9 LYNCOMBE CLOSE  
EXETER, EX4 5EJ

VICTIM BETTY GEORGE  
DOB 02/11/1966  
GENDER F  
OCCUPATION UNEMPLOYED  
HOME ADDRESS  
17 FARM HILL  
EXETER, EX4 2LW

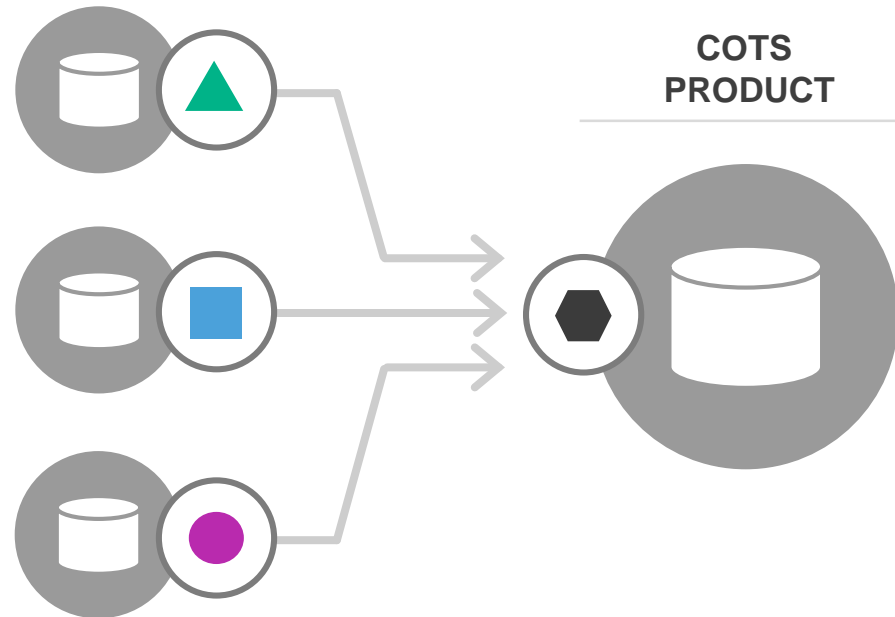
SUSPECTED NIGEL BROWN  
DOB 09/11/1979  
GENDER M  
OCCUPATION UNEMPLOYED  
HOME ADDRESS  
3 PORTLAND ST  
EXETER, EX1 2EG

NARRATIVE  
Victim hit over the head with beer bottle during argument on the road outside of house party on Lycombe Close. Victim and suspect were under the influence of alcohol and drugs at the time.

TRADITIONAL APPROACH WITH COTS

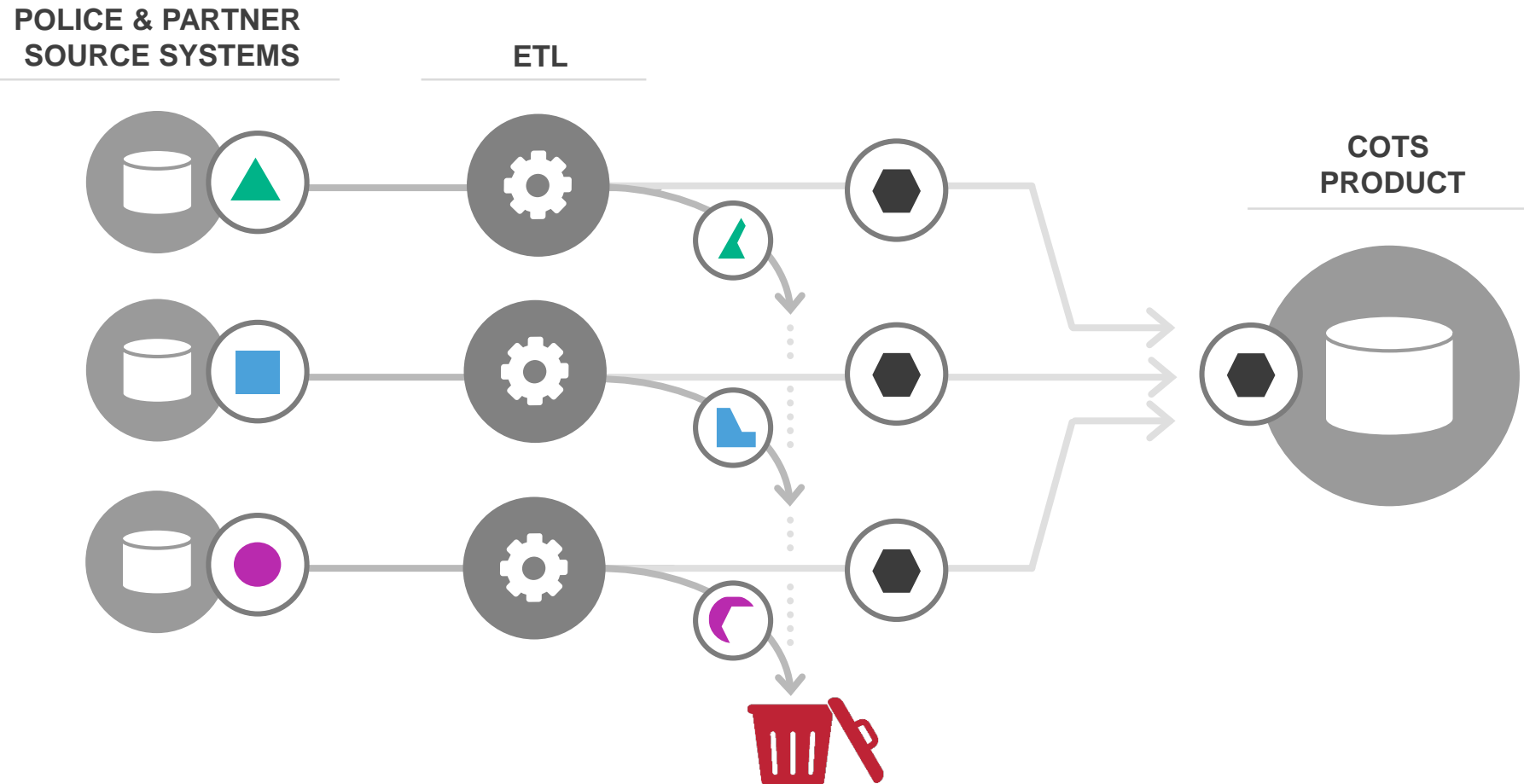
# The Promise: Easy ETL, Low Costs

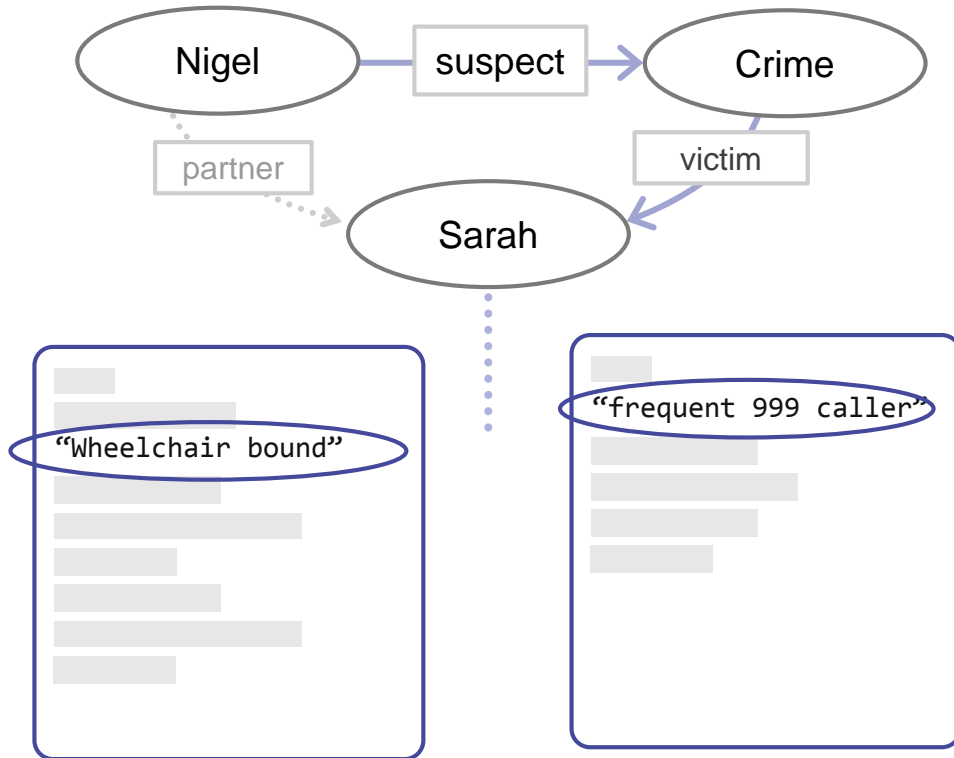
**POLICE & PARTNER  
SOURCE SYSTEMS**



TRADITIONAL APPROACH WITH COTS

# The Reality: Extract, Transform, & Lose





MULTI-MODEL: DOCUMENTS & TRIPLES TOGETHER  
JSON, XML, & RDF

THE IDEAL SOLUTION

## Use All of the Data

- Semantic linking to see relationships between people, locations, events and objects
- Extract context from narrative text
- Build a complete picture by exploiting the value in all of the data

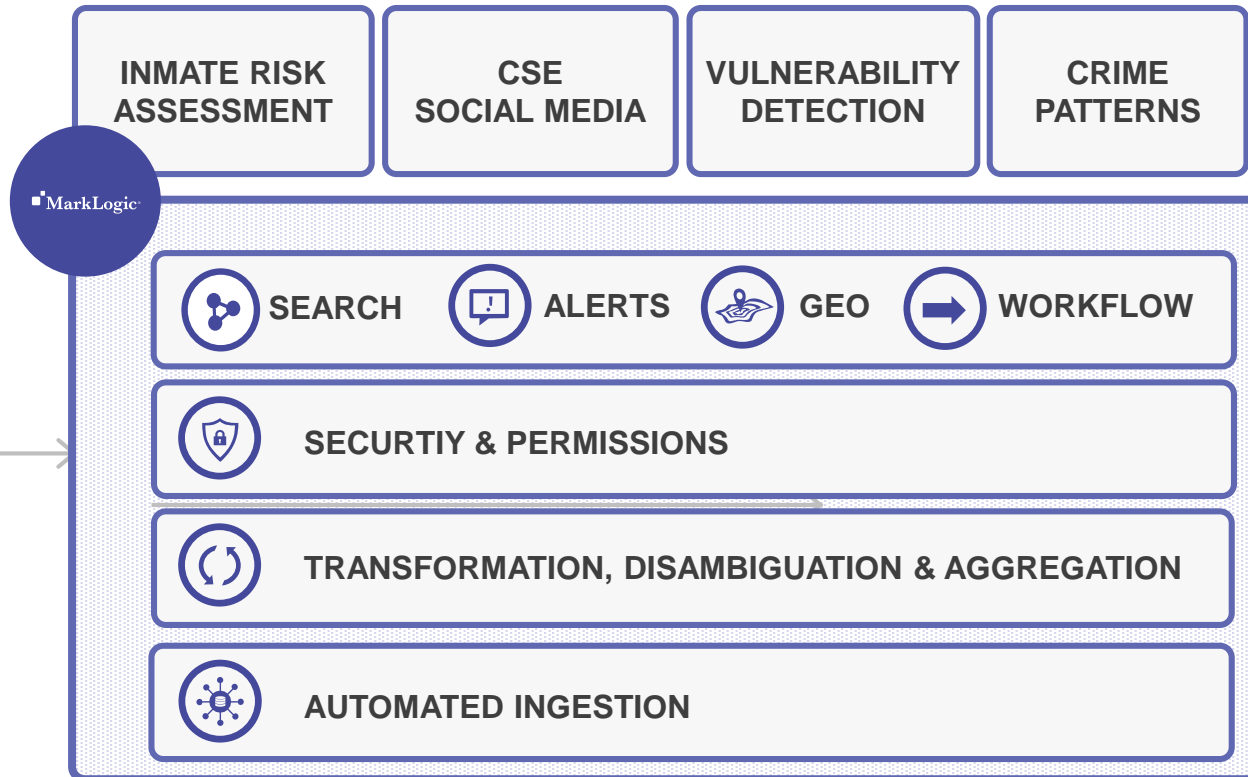
# Police Intelligence Platform

Single point of entry to all force information sources for intelligence and safeguarding

## 10+ Data Feeds

Emergency Calls  
Crimes  
Arrests  
Missing Persons  
Child Protection  
Intelligence  
Mapping & Addresses  
+ MORE

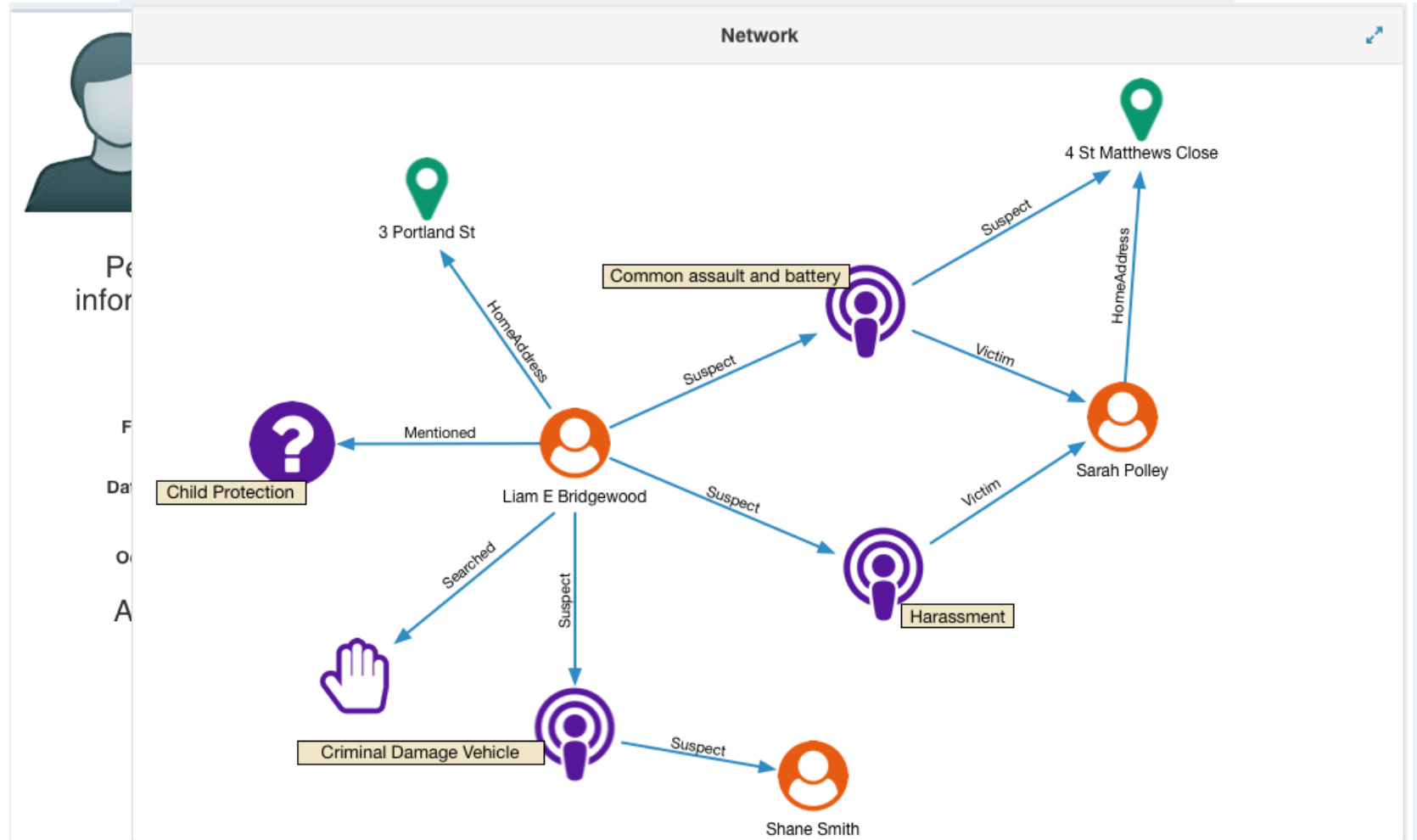
## INTELLIGENCE/ANALYSIS TOOLS



## THE PROJECT

- Single point of entry to all Police information sources for intelligence and safeguarding
- 4 applications built on top of a single unified set of data from 10 different police databases
- 12 weeks of development
- Data quality issues
- Disconnected data

# Use Triples To Find Relationships



# Combine Triples With Documents

## Facebook Search

Upload Archive

facebook-archive-laura-nazmdeh.zip

100%

Analyse

## Options

Export Excel

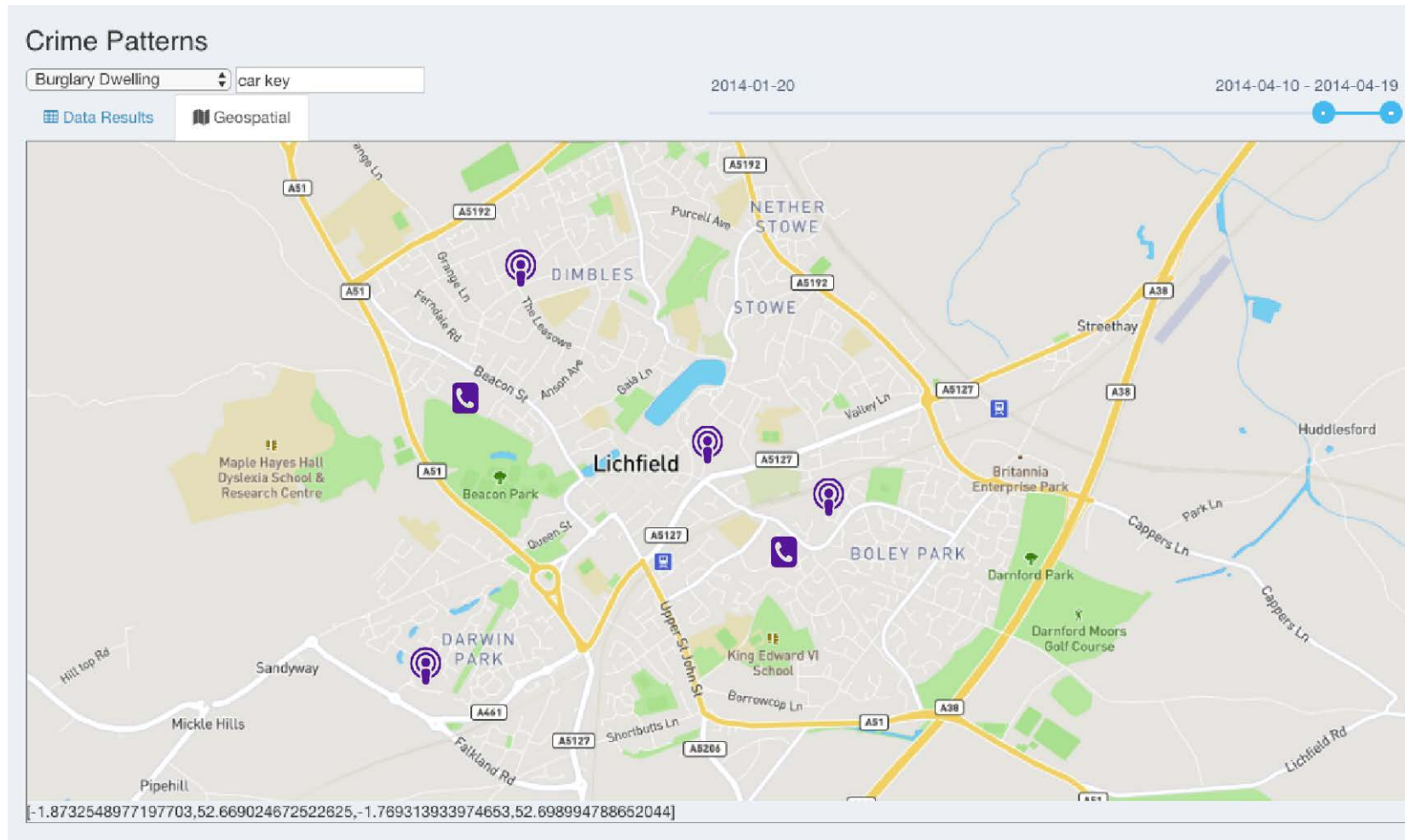
Delete Archive

## User Details

Lovely Naz

Lovely

# Add Geospatial to Triples and Documents






# Statistical Analysis of Linked Data

Vulnerability Report




Available Dates  Available I

Show  entries

Name	
	MAY BROWNLOCK
	Lisa Clown
	Maggie Clown

### Harm Level

1 days:	Count: 2 Score: 84 Hits: 1
7 days:	Count: 2 Score: 84 Hits: 1
30 days:	Count: 6 Score: 225 Hits: 3
90 days:	Count: 8 Score: 240 Hits: 5
365 days:	Count: 8 Score: 240 Hits: 5

10 days	365 days	Persistency	Escalation	
7	7	low	high	
68	98	high	low	
3	4	low	low	

# Load 'As Is'

- Export source data
- Load data as-is as documents – XML/JSON
- Record provenance information – PROV-O ontology
- Harmonize data – envelope pattern
- Canonicalize – POLE model

```
<CRIME>
```

```
<REFERENCE>HW/001900/15</REFERENCE>
```

```
<CODE>MOPI GROUP 2 - ASSAULT A.B.H</CODE>
```

```
<CRIMINCIDENT_DATE>08/11/2015</CRIMINCIDENT_DATE>
```

```
<ADDRESS>
```

```
<TYPE>SCENE OF CRIME</TYPE>
```

```
<STREET>8 HAVEN ROAD, EXETER</STREET>
```

```
<POSTCODE>EX2 8BP</POSTCODE>
```

```
</ADDRESS>
```

```
<PERSON>
```

```
<AUTNPERSONTYPE>VICTIM</AUTNPERSONTYPE>
```

```
<SURNAME>PHILLIPS</SURNAME>
```

```
<FORENAME>MILDRED</FORENAME>
```

```
<DATE_OF_BIRTH>14/01/1927</DATE_OF_BIRTH>
```

```
<GENDER>F</GENDER>
```

```
<OCCUPATION>UNEMPLOYED</OCCUPATION>
```

```
<PERSONALIASLIST></PERSONALIASLIST>
```

```
...
```

# Harmonize People

- Generate a unique ID for the entity instance
- Harmonize element names and data formats
- Generate phonetic versions of names

```
<envelope>
```

```
  <person>
```

```
    <uri>646569e5-5f6c-4667-96c7-09ff84b3e08e</uri>
```

```
    <personType>VICTIM</personType>
```

```
    <surname>PHILLIPS</surname>
```

```
    <surname_dm>flps</surname_dm>
```

```
    <forename>MILDRED</forename>
```

```
    <forename_dm>m1trt</forename_dm>
```

```
    <dob>1927-01-14</dob>
```

```
    <gender>f</gender>
```

```
    <occupation>UNEMPLOYED</occupation>
```

```
  ...
```

# Harmonize Event

- Generate a unique ID for the entity instance
- Harmonize element names and data formats

```
<envelope>
```

```
  <event>
```

```
    <uri>police.uk/event/crime/CMS2_106600</uri>
```

```
    <eventType>Crime</eventType>
```

```
    <eventDate>2015-11-08</eventDate>
```

```
  ...
```

# Harmonize Locations

- Utilize an authoritative reference data source for addresses – e.g. Ordnance Survey
- Record the Unique Property Reference Number (UPRN)

```
<envelope>
  <event>
    <uri>police.uk/event/crime/CMS2_106600</uri>
    <eventType>Crime</eventType>
    <eventDate>2015-11-08</eventDate>
  <location>key:3a001d392f0e819e98810095a542391a96aa177e
</event>
  <address>
    <key>key:3a001d392f0e819e98810095a542391a96aa177e</key>
  </address>
...

```

# Calculate Hashes

- Compute hash codes for dimensional combinations that disambiguate the entity:
  - forename\_dm && surname\_dm && dob
  - forename\_dm && surname\_dm && address
- Leverage MarkLogic's universal index for resolving the keys

```
<envelope>
```

```
<person>
```

```
<uri>646569e5-5f6c-4667-96c7-09ff84b3e08e</uri>
```

```
<personType>VICTIM</personType>
```

```
<surname>PHILLIPS</surname>
```

```
<surname_dm>flps</surname_dm>
```

```
<forename>MILDRED</forename>
```

```
<forename_dm>m1trt</sp:forename_dm>
```

```
<dob>1927-01-14</dob>
```

```
<gender>f</gender>
```

```
<occupation>UNEMPLOYED</occupation>
```

```
<key>key:21a7e3154a71311e07f277d8696262d0bbd1bf94</key>
```

```
<key>key:82b93e911bd18e2612fae64d1c81e889b3858f64</key>
```

```
...
```

# Store Triples

- Record relationships between entities:
  - Person <suspectOf> Crime

```
<envelope>
```

```
<triple>
```

```
<subject>7b9c5fb1-4a49-4d01-8979-a84408da51c5</subject>
```

```
<predicate>suspectOf</predicate>
```

```
<object>police.uk/event/crime/CMS2_106600</object>
```

```
</triple>
```

# Store Triples

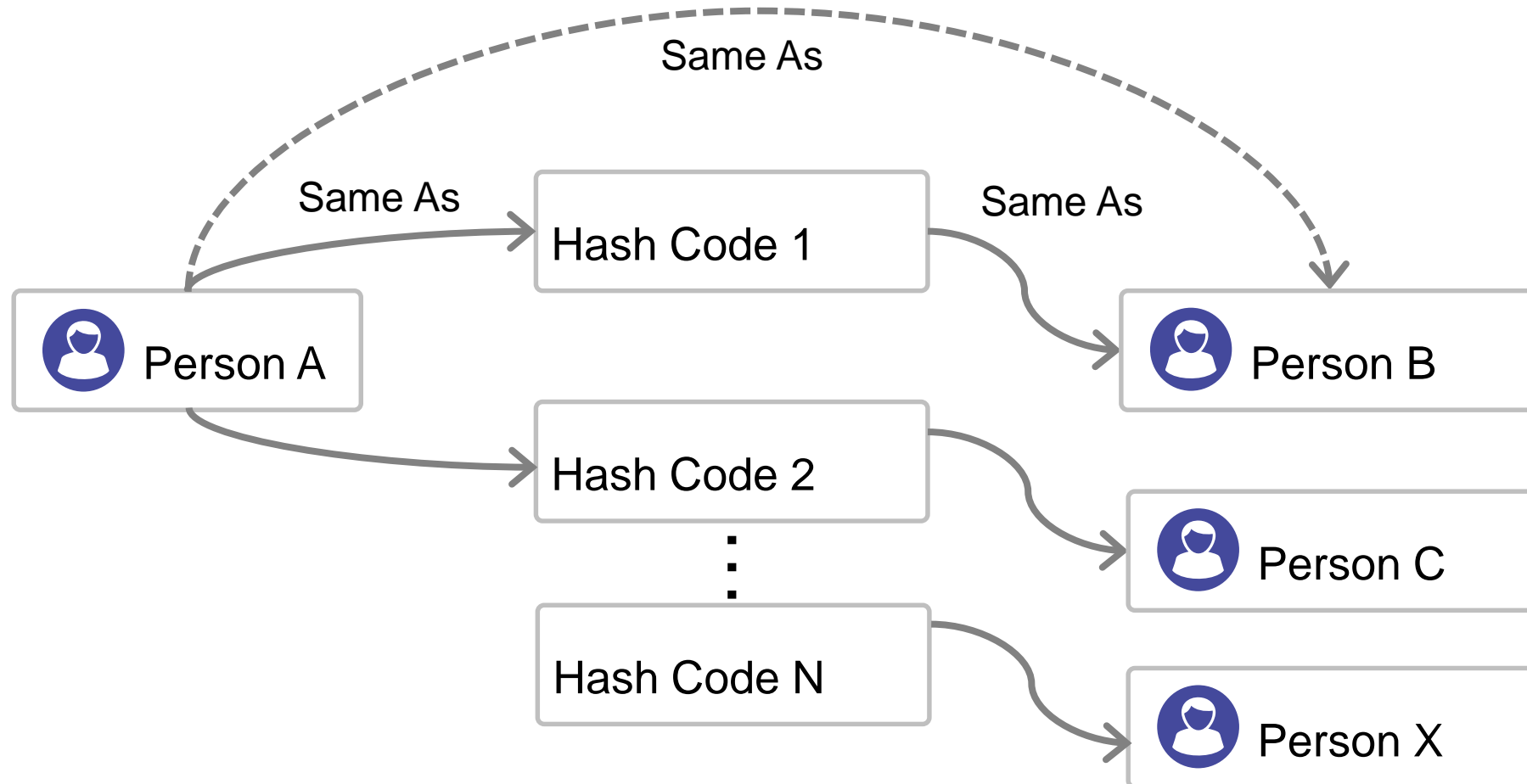
- Record relationships between entities:
  - Person <suspectOf> Crime
- Record relationship between entity instance (i.e Person) and the disambiguation hash code

```
<envelope>
  <triple>
    <subject>7b9c5fb1-4a49-4d01-8979-a84408da51c5</subject>
    <predicate>suspectOf</predicate>
    <object>police.uk/event/crime/CMS2_106600</object>
  </triple>

  <triple>
    <subject>7b9c5fb1-4a49-4d01-8979-a84408da51c5</subject>
    <predicate>sameAs</predicate>
    <object>key:21a7e3154a71311e07f277d8696262d0bbd1bf94
  </triple>

  <triple>
    <subject>7b9c5fb1-4a49-4d01-8979-a84408da51c5</subject>
    <predicate>sameAs</predicate>
    <object>key:82b93e911bd18e2612fae64d1c81e889b3858f64
  </triple>
```

# Collapsing Entities Using Semantics



# Advantages of the Multi-Model Approach

---

## Fast

- Fast search - limits joins as entities are documents, relationships are triples
- Fast ingest - disambiguation effort is performed at query time
- Fast disambiguation - sem:transitive-closure operation is very quick

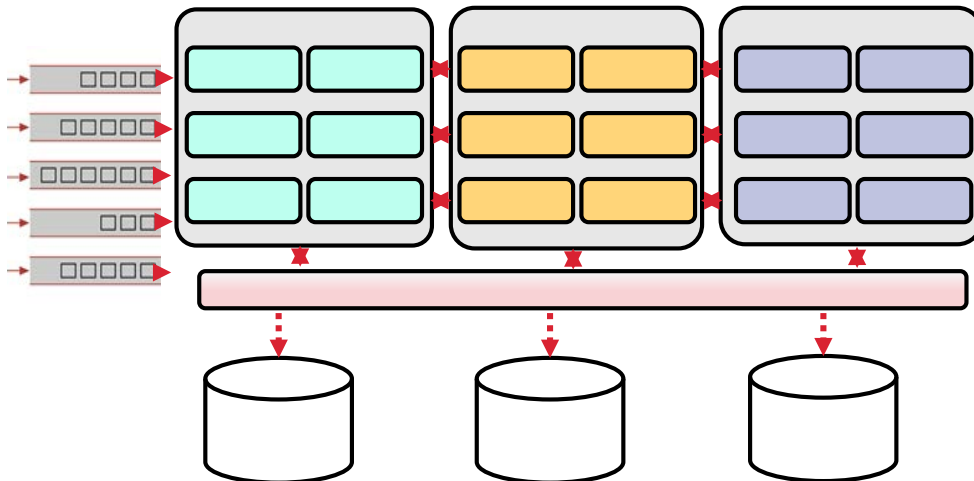
## Flexible

- Query time disambiguation allows rules to be changed or applied on a per user basis
- Use different predicates for use-case sensitive deduplication and different degrees of confidence

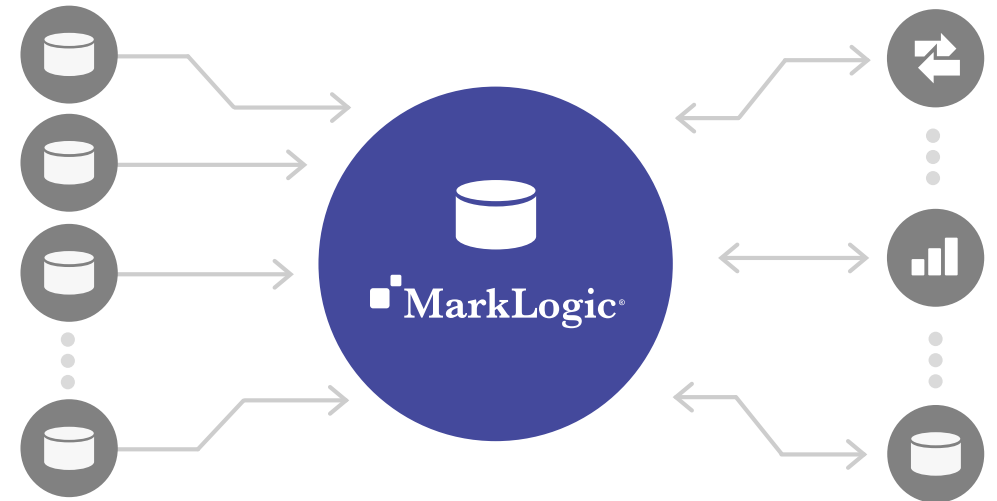
TWO OPTIONS

# Multi-Vendor Approach vs. MarkLogic Approach

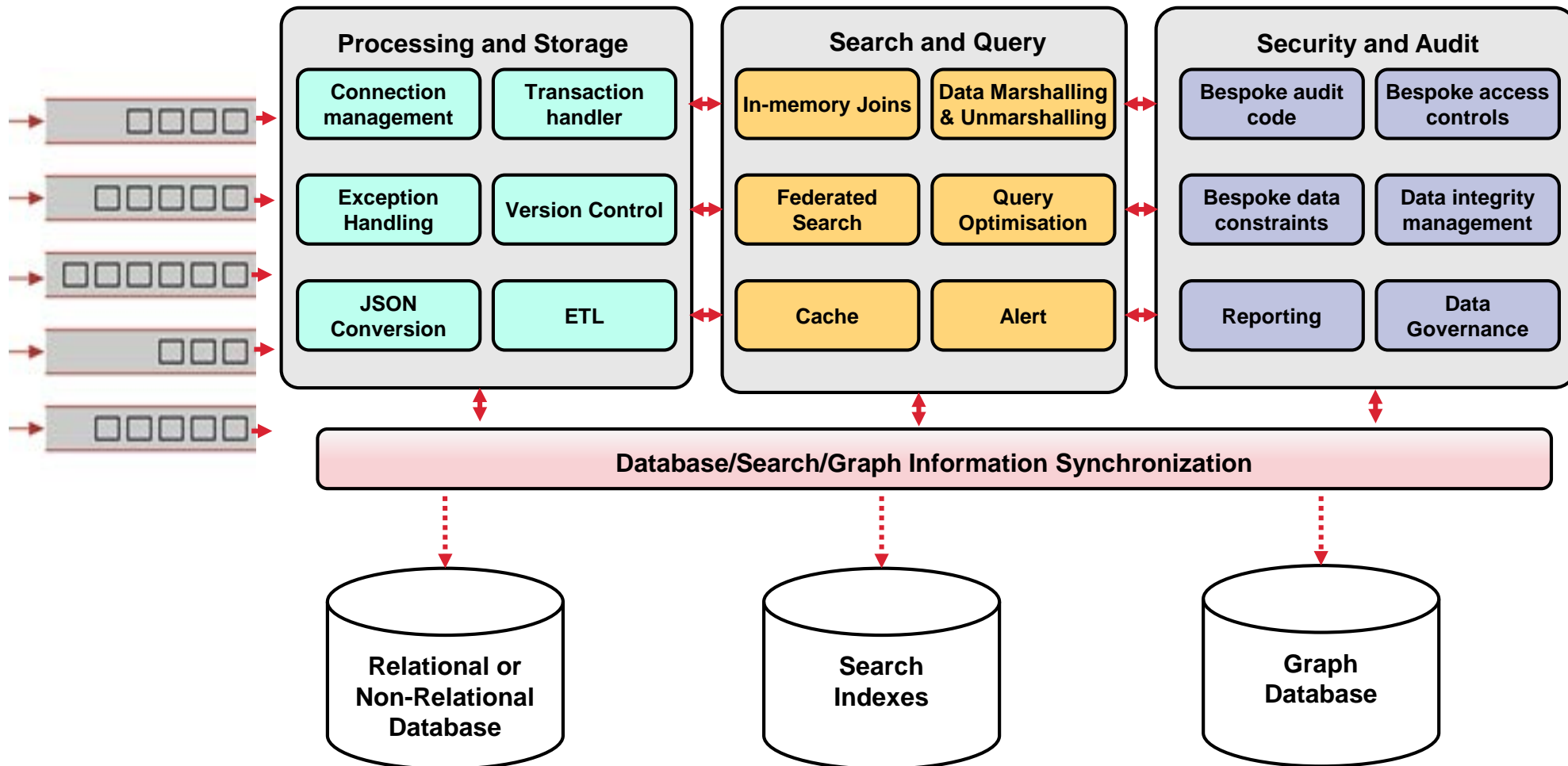
MULTI-PRODUCT,  
MULTI-DATABASE STACK



MARKLOGIC  
OPERATIONAL DATA HUB

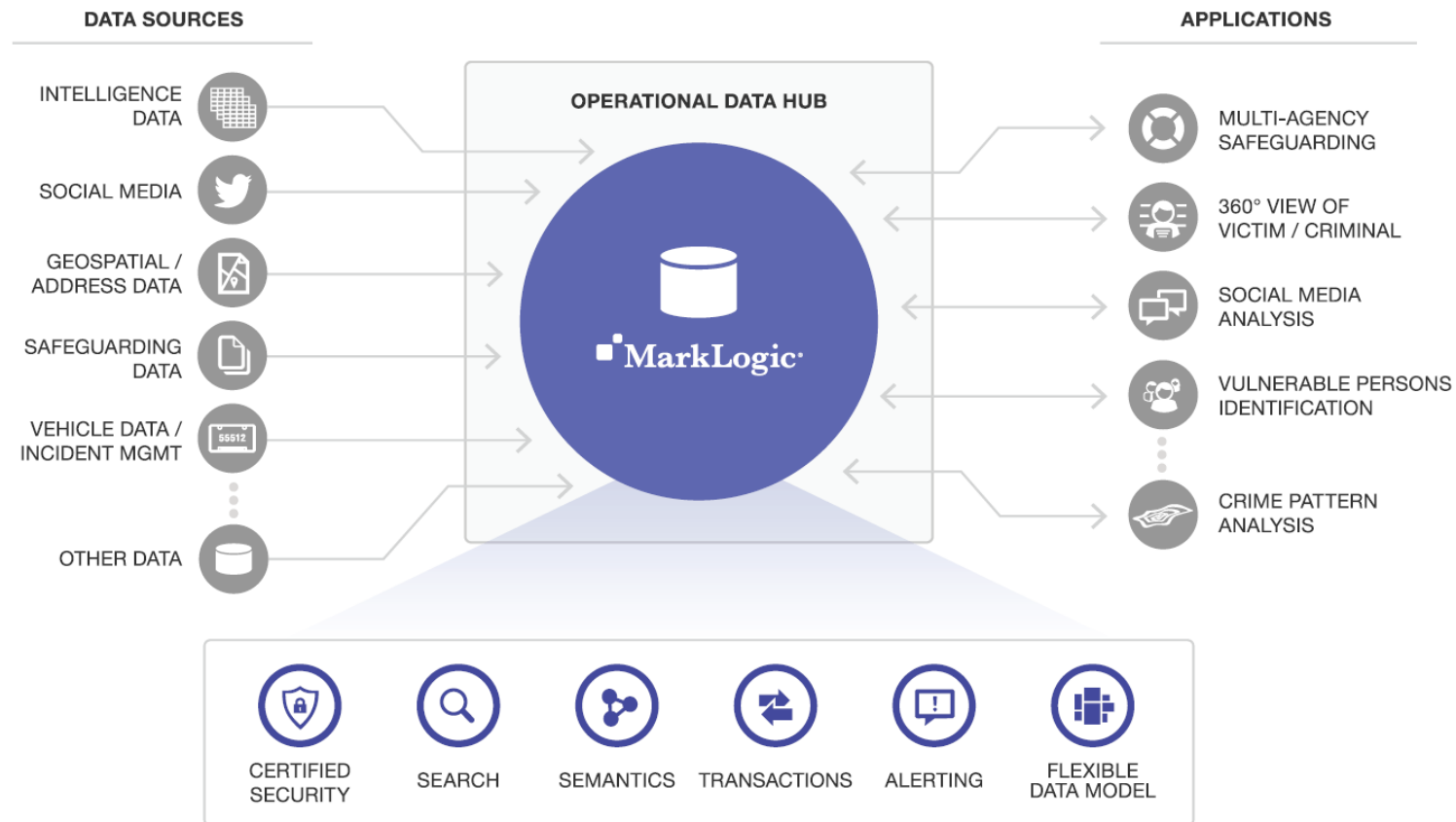


# Multi-Product, Multi-Database Stack



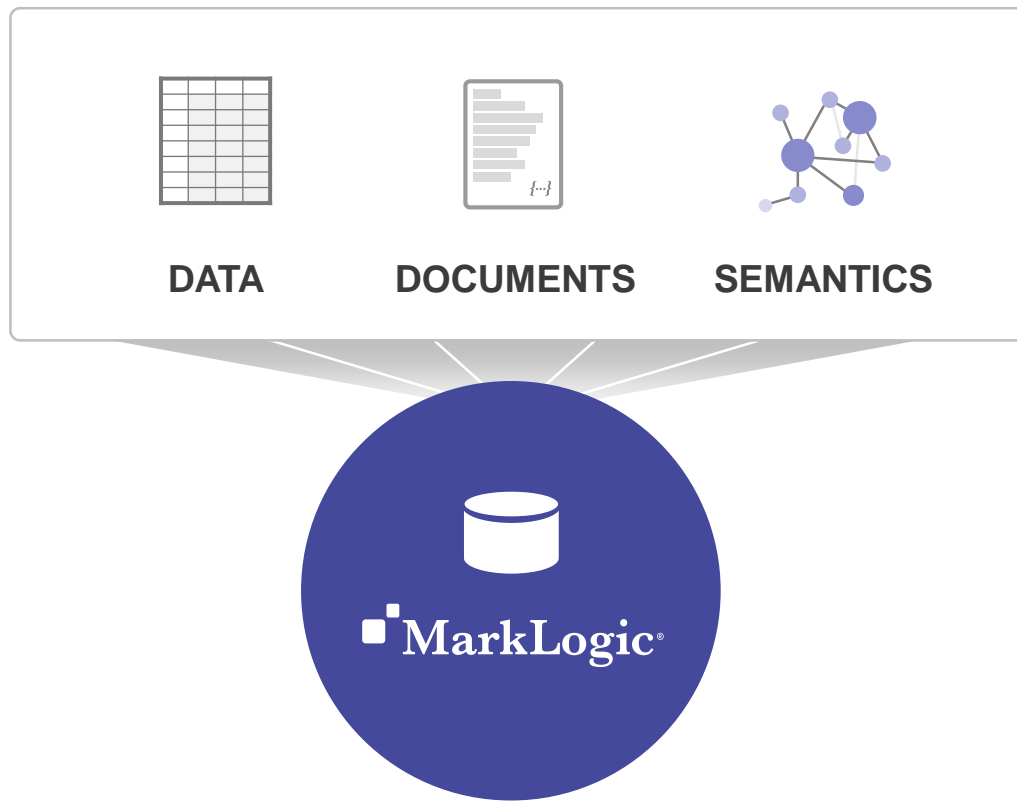
OUR SOLUTION

# The Operational Data Hub



An example of the MarkLogic Operational Data Hub supporting multiple data-driven applications for intelligence and safeguarding

# Benefits of a Document + Triple Store



## ***All the benefits of each, plus:***

- Docs can contain triples, Triples can annotate docs, Graphs can contain docs
  - Faster data integration using semantics as the glue
  - Ideal model for reference data, metadata, provenance
  - Ability to run really powerful queries
- Massive speed and scale
- Simplicity of a single unified platform
- Enterprise features (security, HA/DR, ACID transactions,...)

## Why MarkLogic?

---

True Multi-model Database



For All of Your Data



Enterprise-Ready



Agility and Faster Time to Results



Simplified Architecture



# Q&A