Semantics In Action For Proactive Policing

Jen Shorten

Technical Delivery Architect, Consulting Services

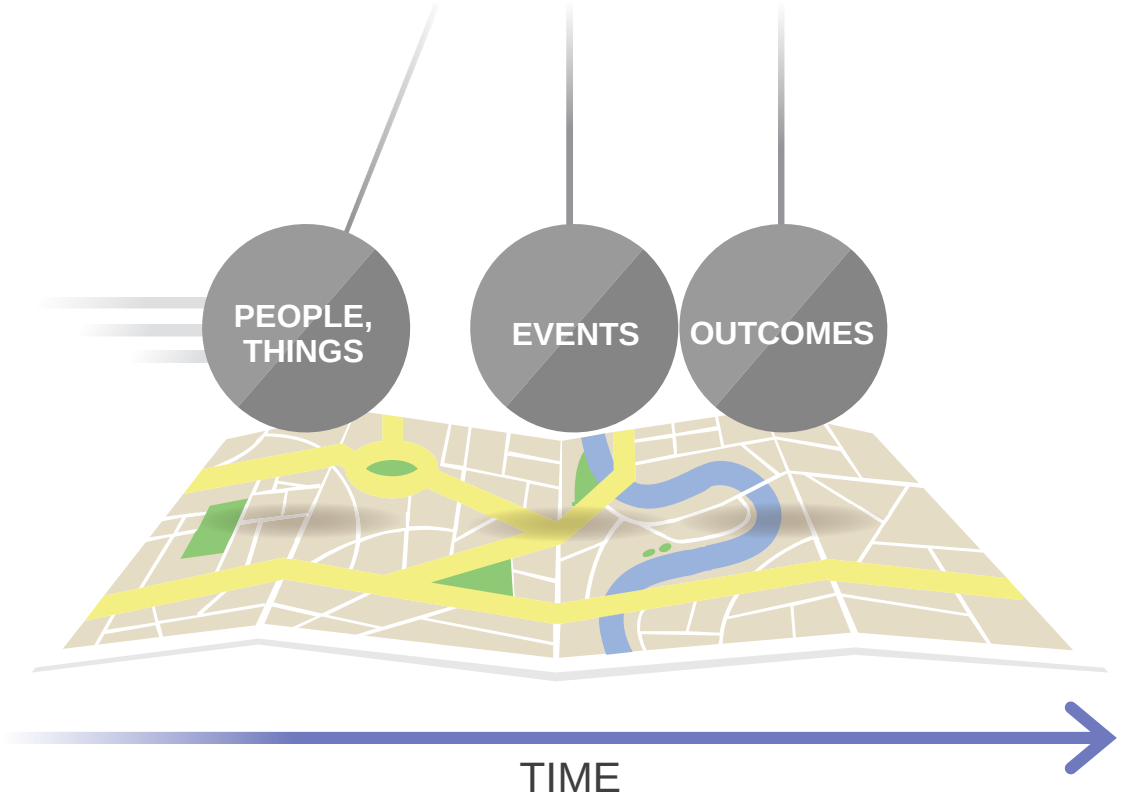
Jon Williams

Senior Sales Engineer, UK Public Sector

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

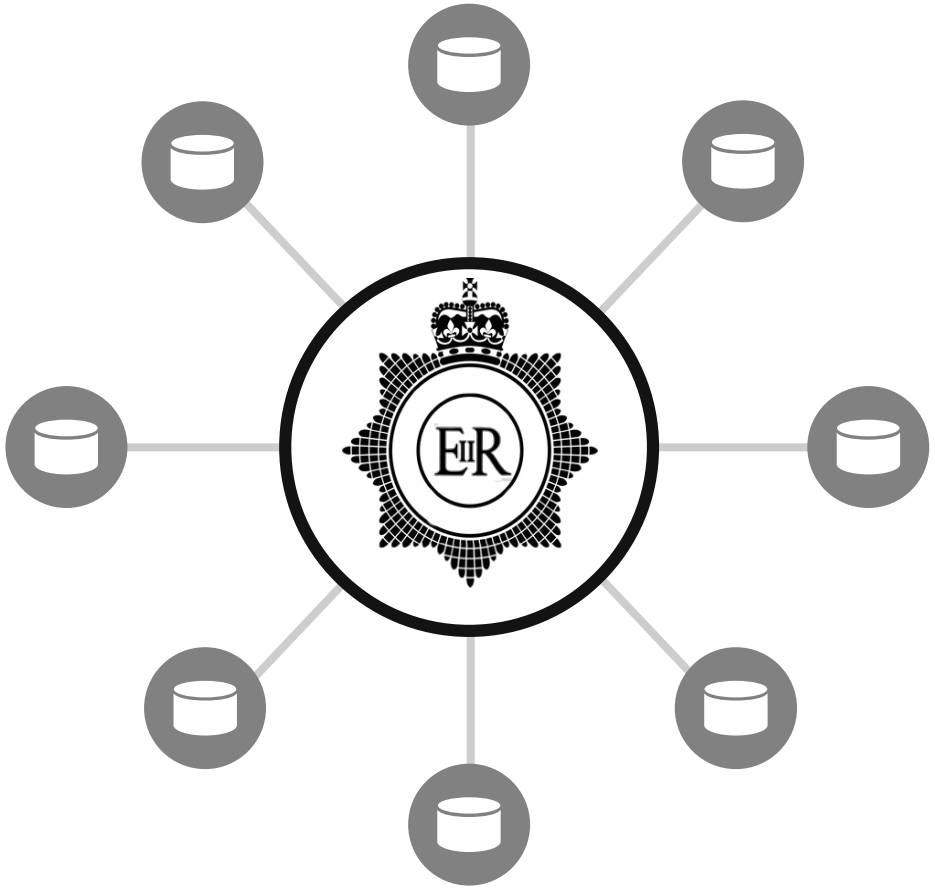


NATIONAL POLICE OBJECTIVES

Digital Transformation of Policing

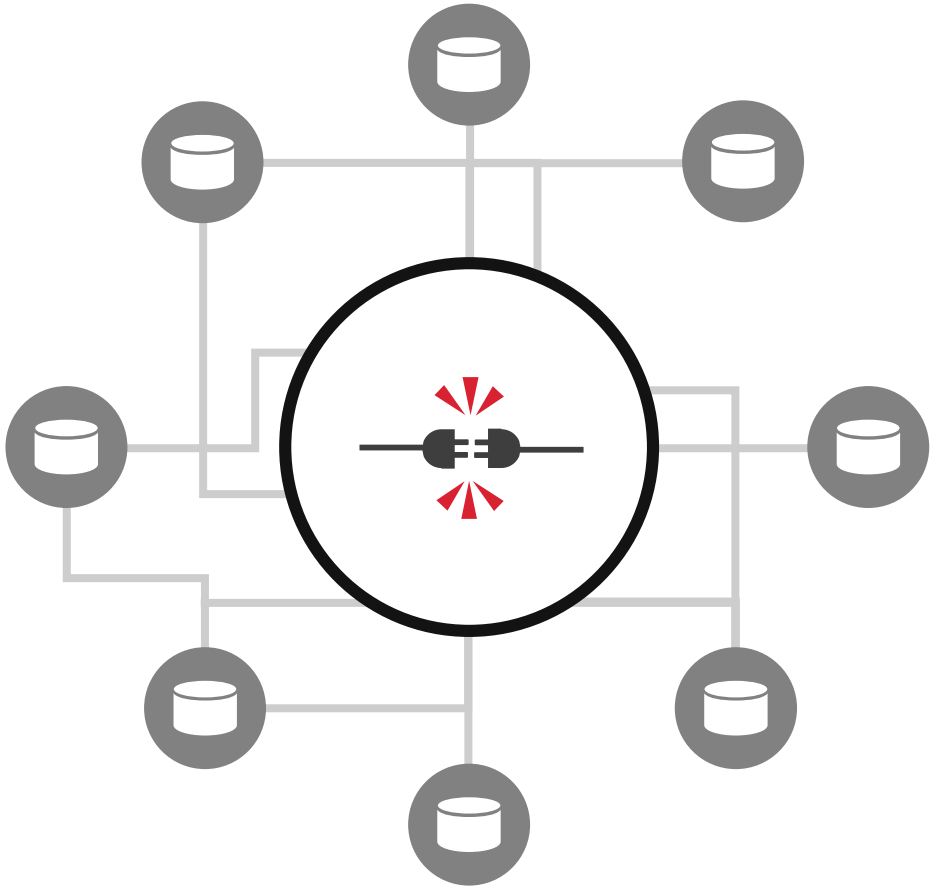
- Proactive
- Impact led
- Outcomes driven
- Data driven

MarkLogic



WHAT POLICE NEED TO DELIVER THAT VISION

A Unified, Actionable 360 View of Data

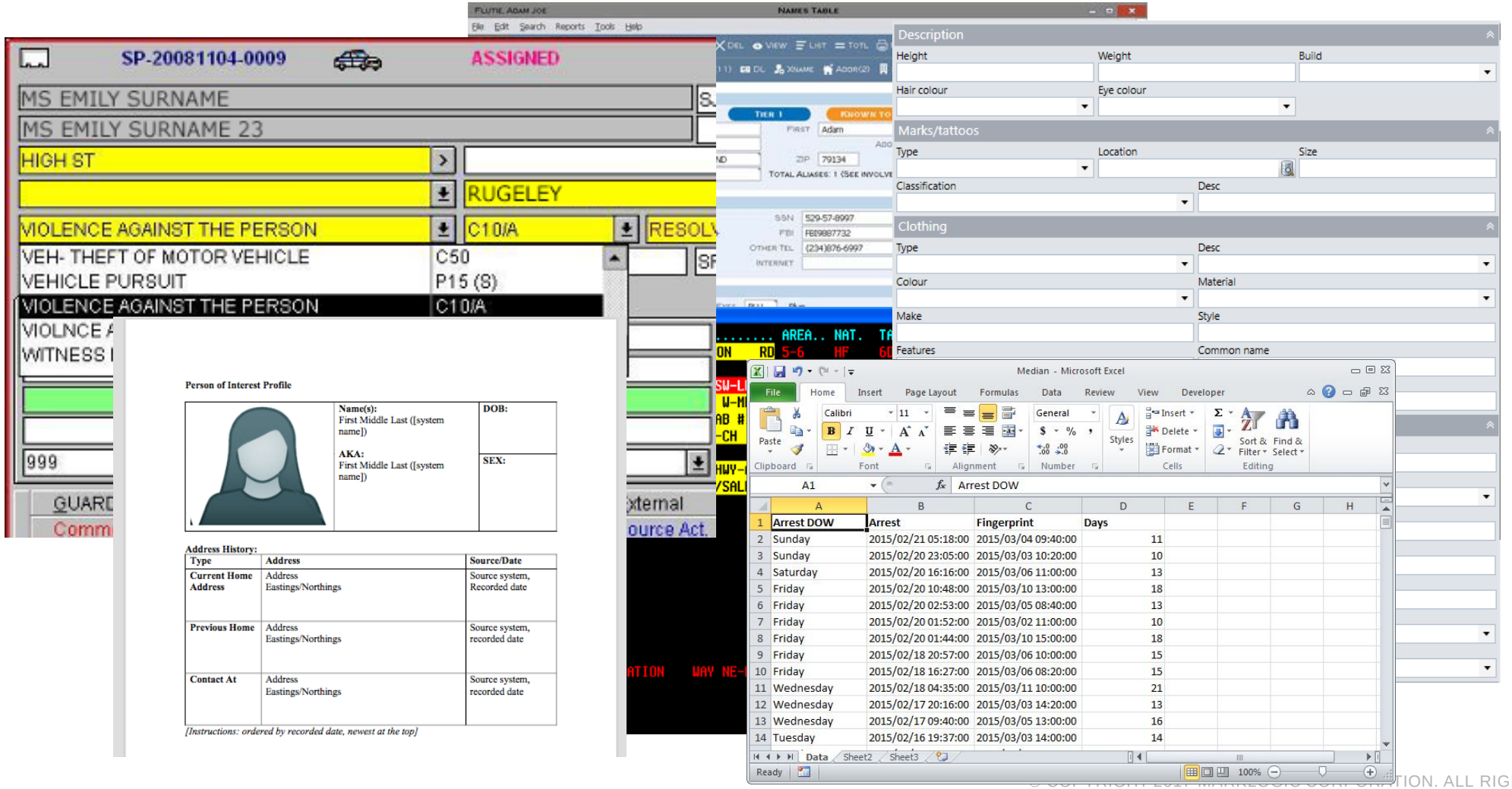


THE REALITY

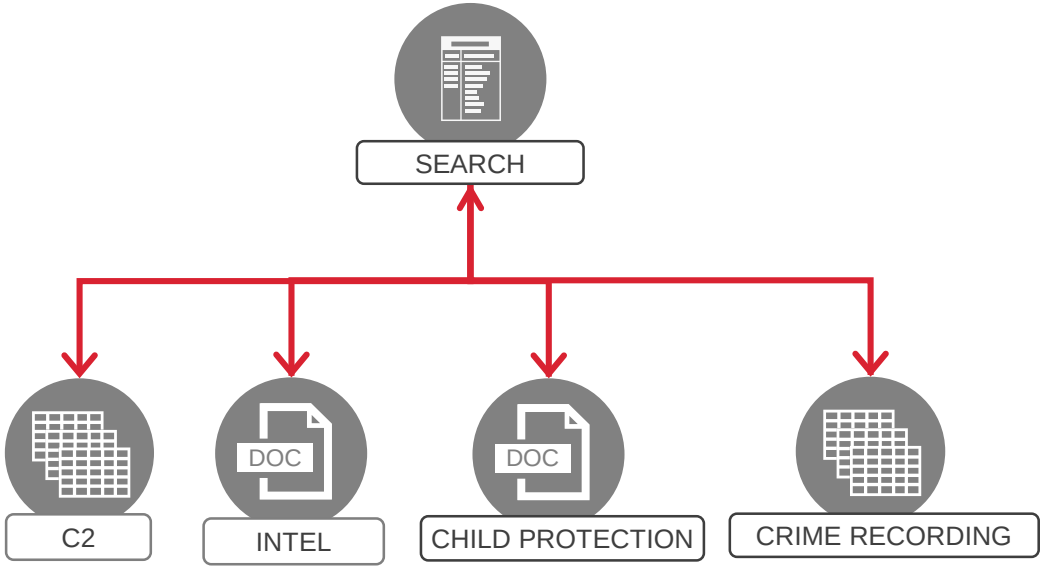
Data Is In Silos

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

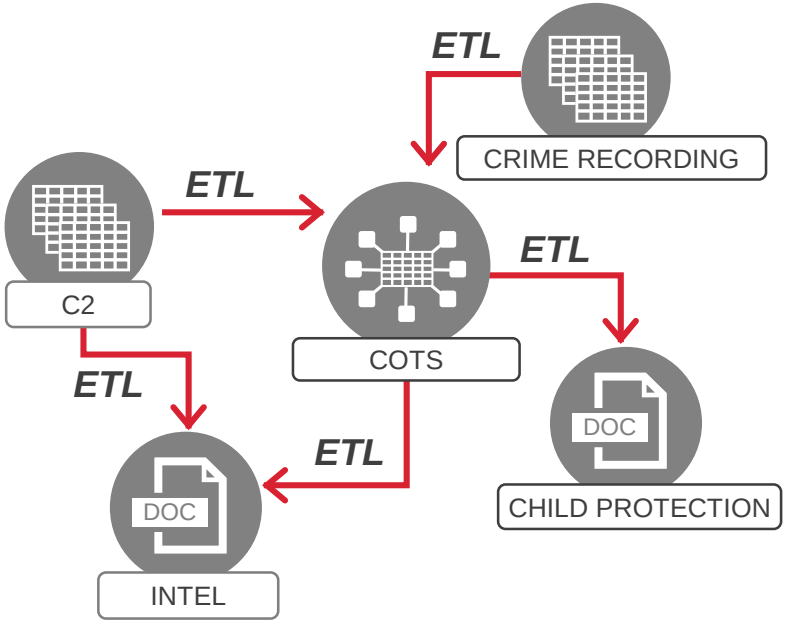
Current State



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

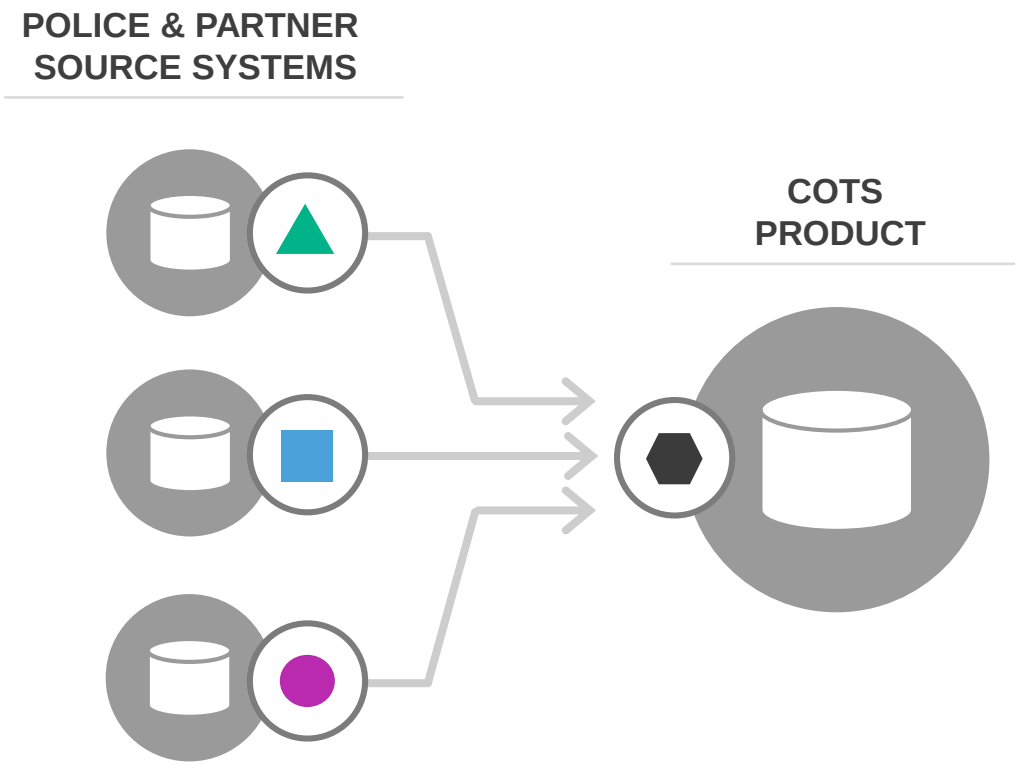
VICTIM BETTY GEORGE
DOB 02/11/1966
GENDER F
OCCUPATION UNEMPLOYED
HOME ADDRESS
17 FARM HILL
EXETER, EX4 2LW

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.

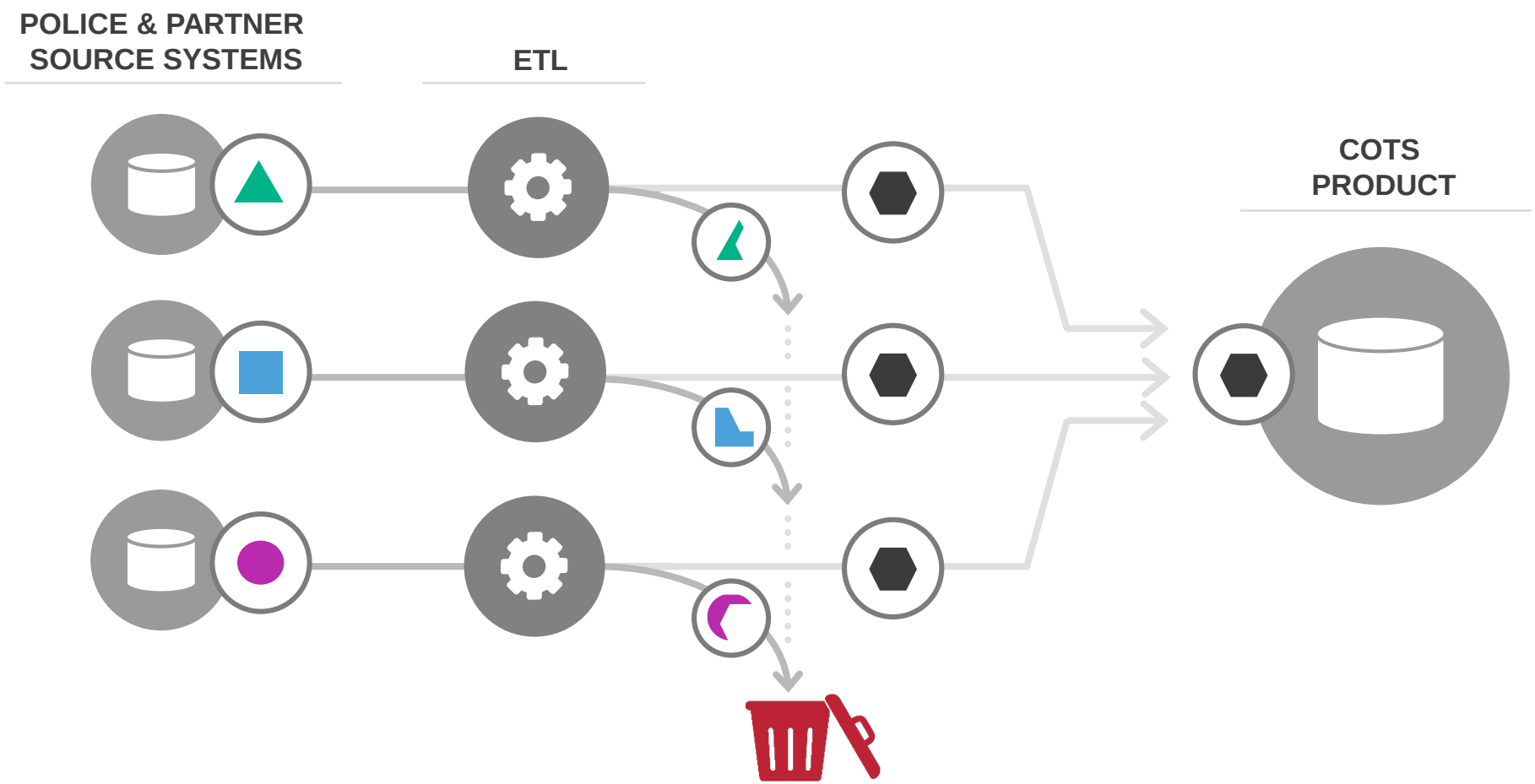
TRADITIONAL APPROACH WITH COTS

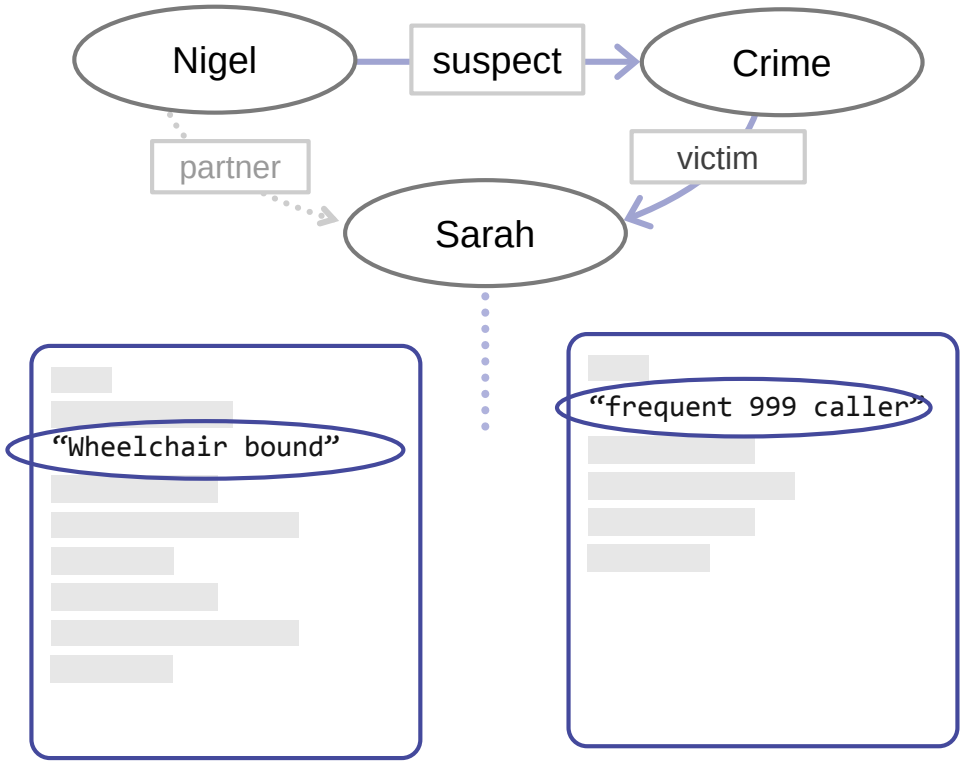
The Promise: Easy ETL, Low Costs



TRADITIONAL APPROACH WITH COTS

The Reality: Extract, Transform, & Lose





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

THE IDEAL SOLUTION

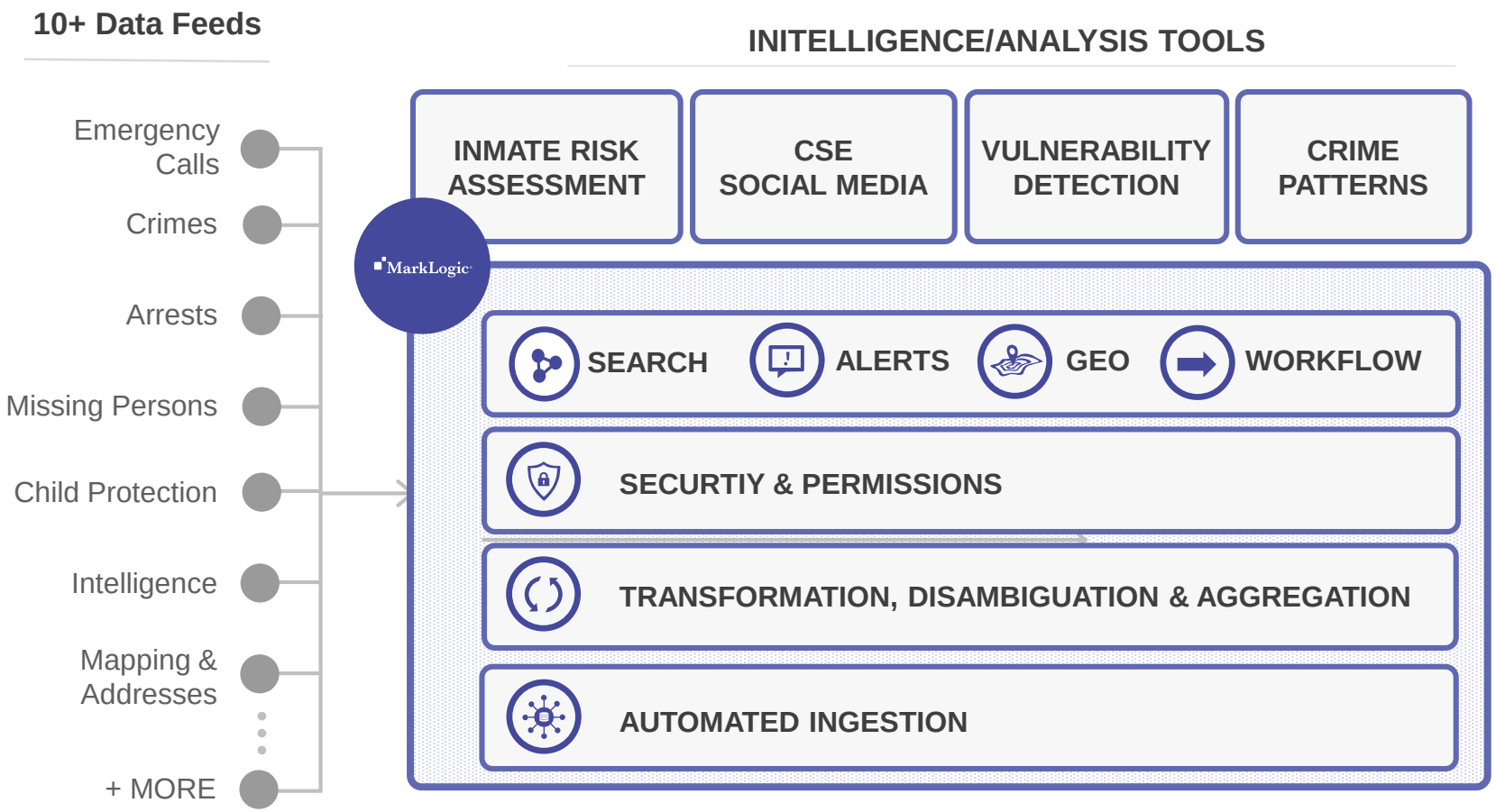
Use All of the Data

- 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

MarkLogic

Police Intelligence Platform

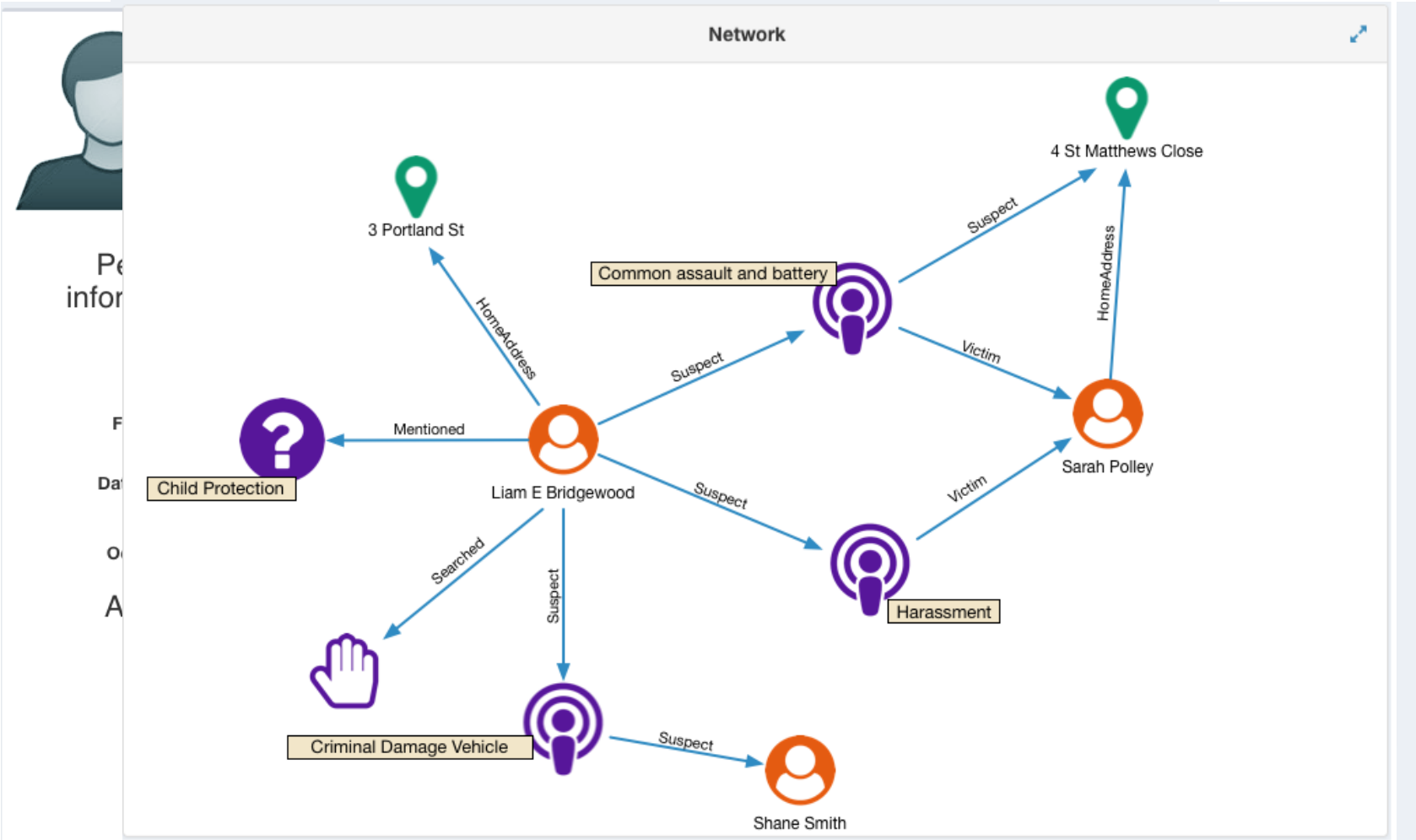
Single point of entry to all force information sources for intelligence and safeguarding



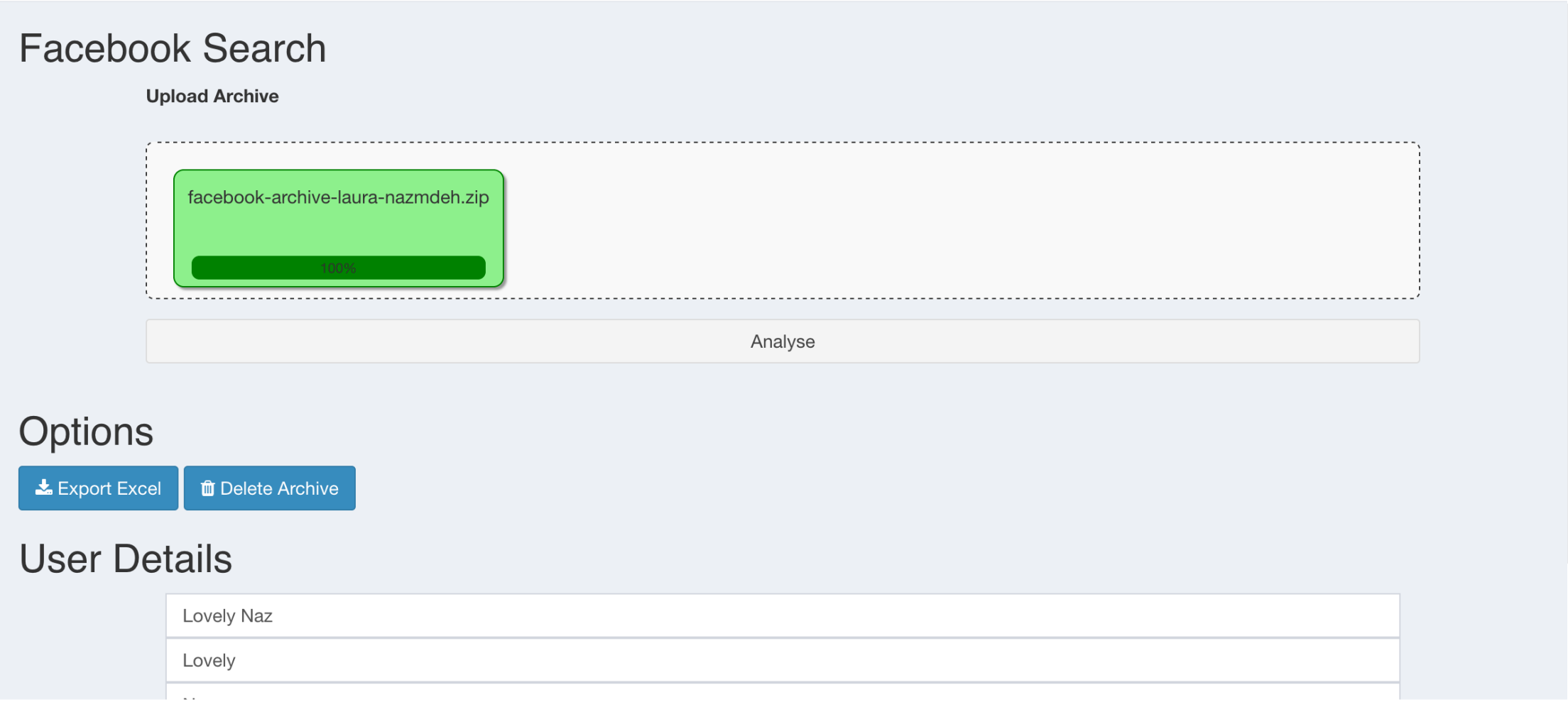
THE PROJECT

- 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

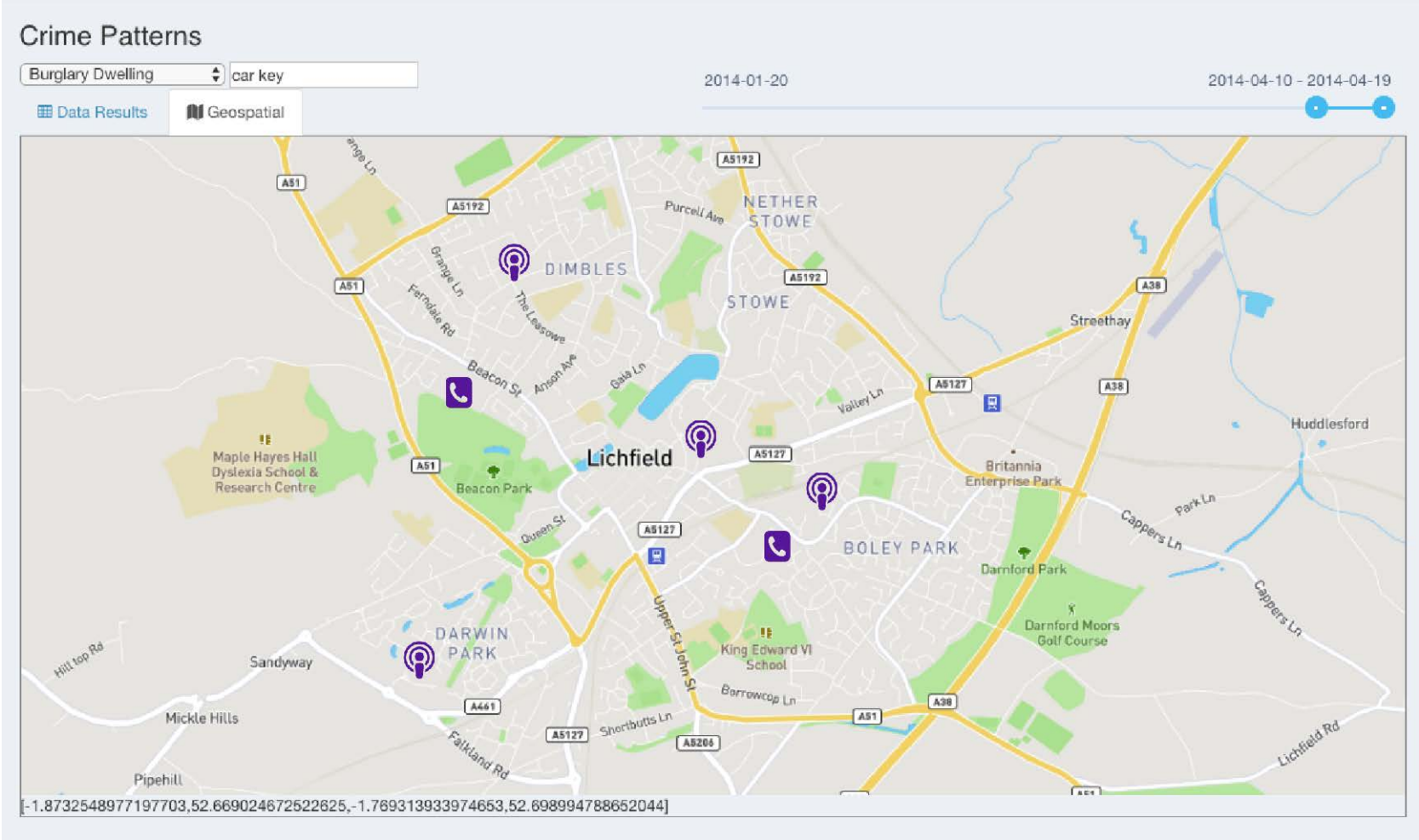
Use Triples To Find Relationships



Combine Triples With Documents



Add Geospatial to Triples and Documents



Statistical Analysis of Linked Data

Vulnerability Report

Available Dates 2014-08-20 **Available I**

Show 10 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

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

MarkLogic

DATA PROCESSING

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>
...
```

MarkLogic

EXTRACT ENTITIES

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>mltrt</forename_dm>
    <dob>1927-01-14</dob>
    <gender>f</gender>
    <occupation>UNEMPLOYED</occupation>
...
```

EXTRACT ENTITIES

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>
...
```

EXTRACT ENTITIES

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>
...
```

MarkLogic

APPLY MATCHING RULES

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>mltrt</sp:forename_dm>
    <dob>1927-01-14</dob>
    <gender>f</gender>
    <occupation>UNEMPLOYED</occupation>
    <key>key:21a7e3154a71311e07f277d8696262d0bbd1bf94</key>
    <key>key:82b93e911bd18e2612fae64d1c81e889b3858f64</key>
...
```

MarkLogic

RECORD RELATIONSHIPS

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>
```

MarkLogic

RECORD RELATIONSHIPS

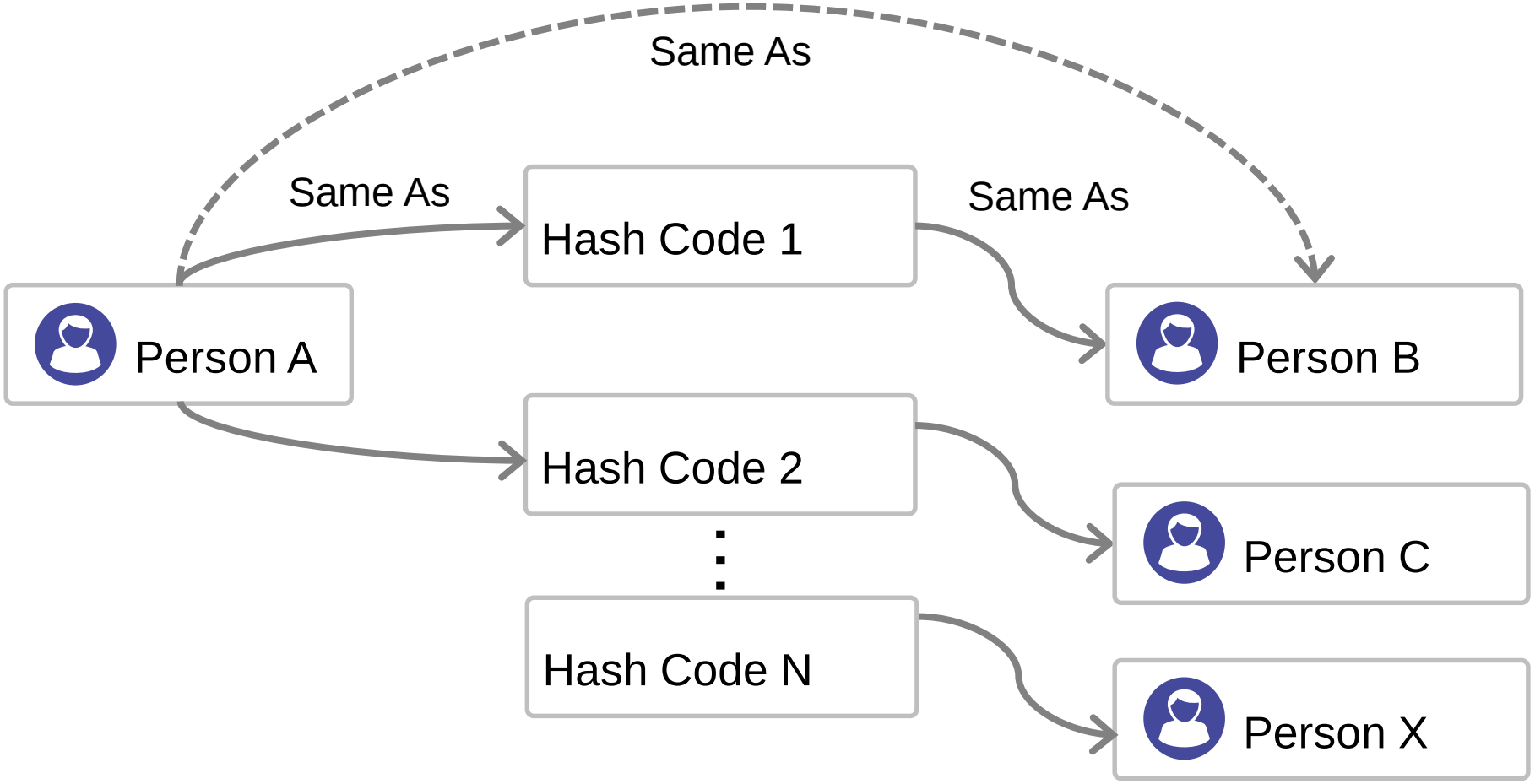
Store Triples

- 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



MarkLogic

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

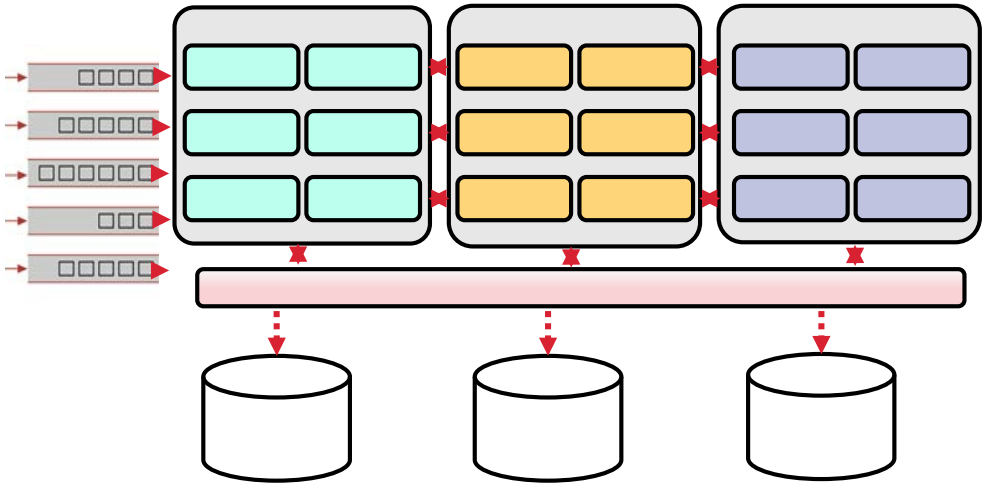
Flexible

- 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

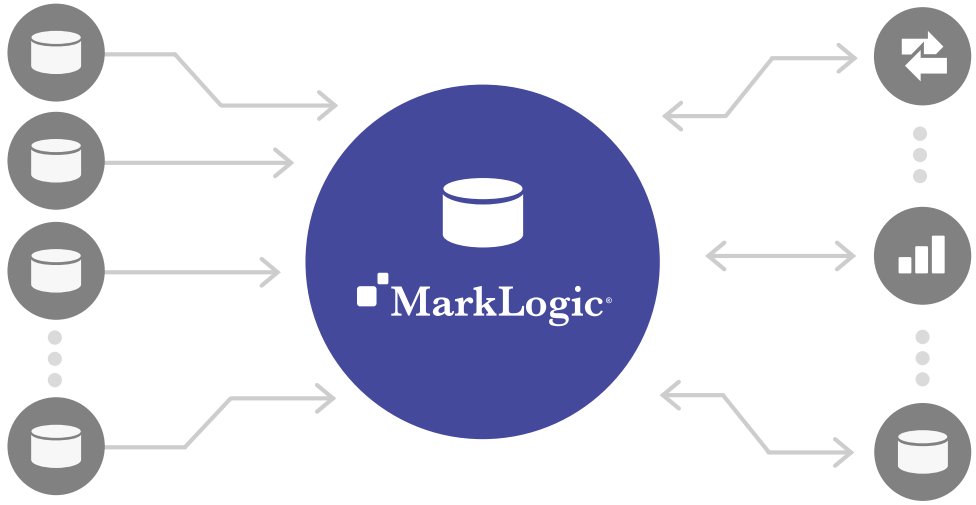
TWO OPTIONS

Multi-Vendor Approach vs. MarkLogic Approach

MULTI-PRODUCT, MULTI-DATABASE STACK

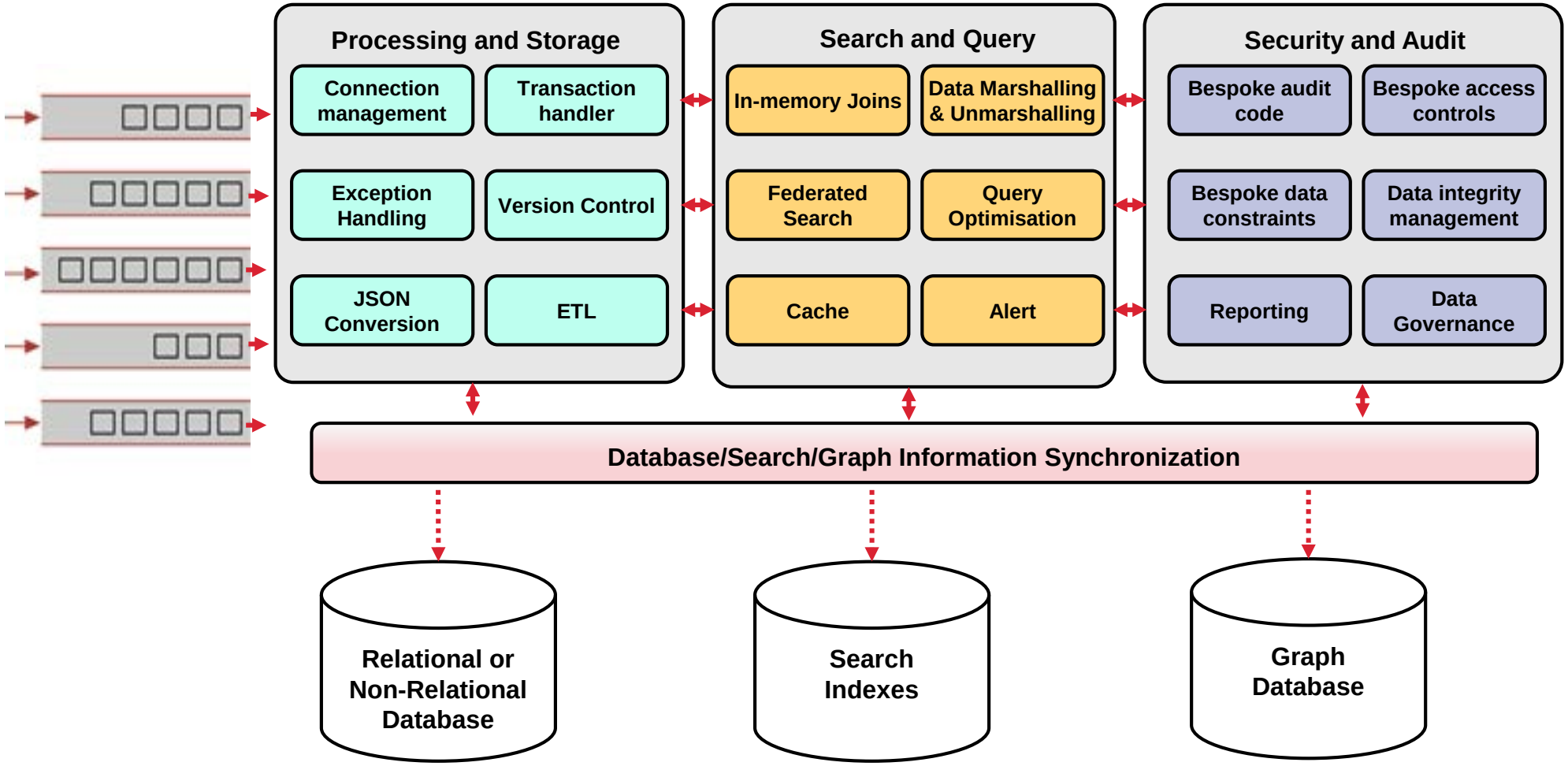


MARKLOGIC OPERATIONAL DATA HUB



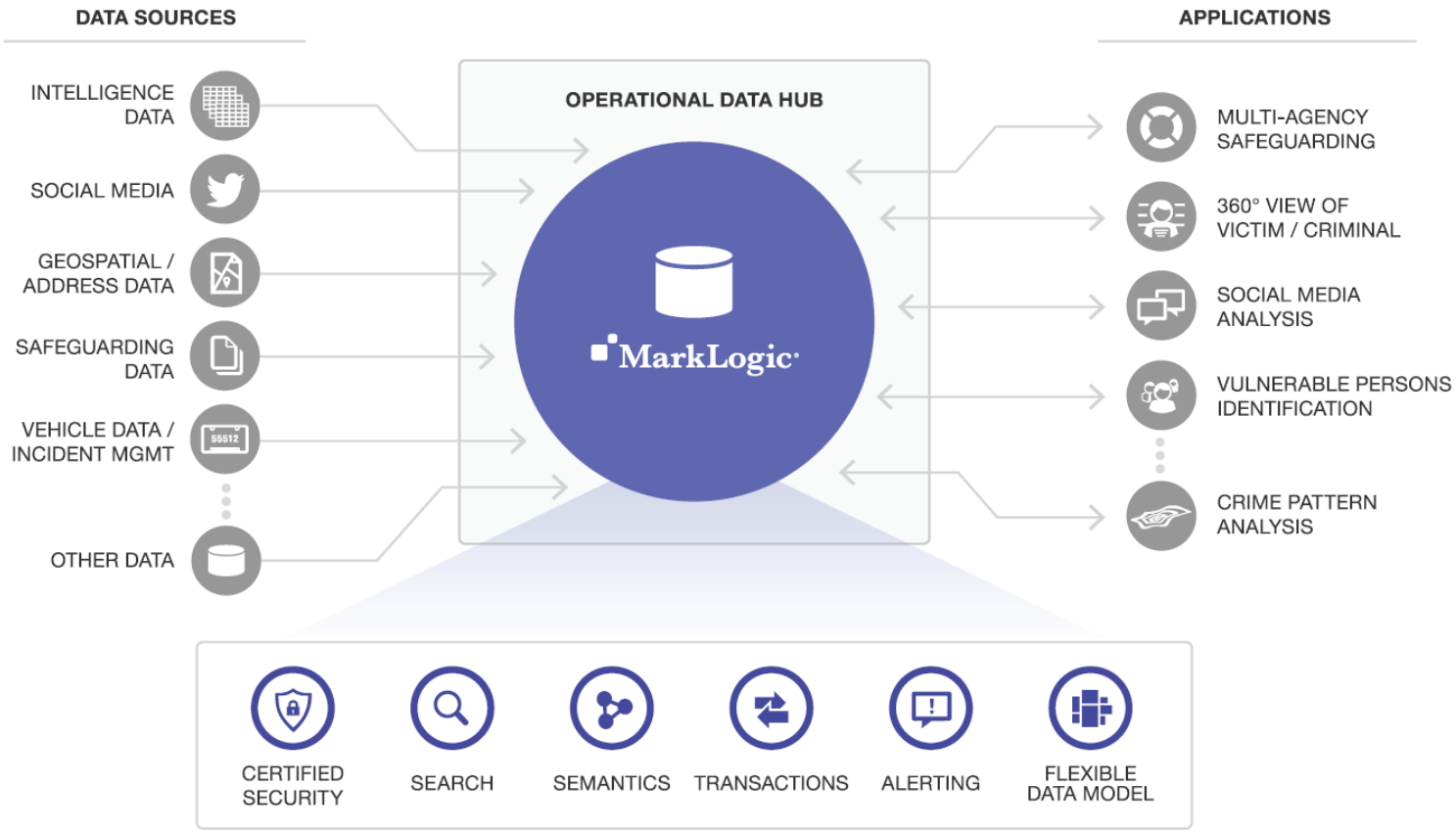
MarkLogic

Multi-Product, Multi-Database Stack



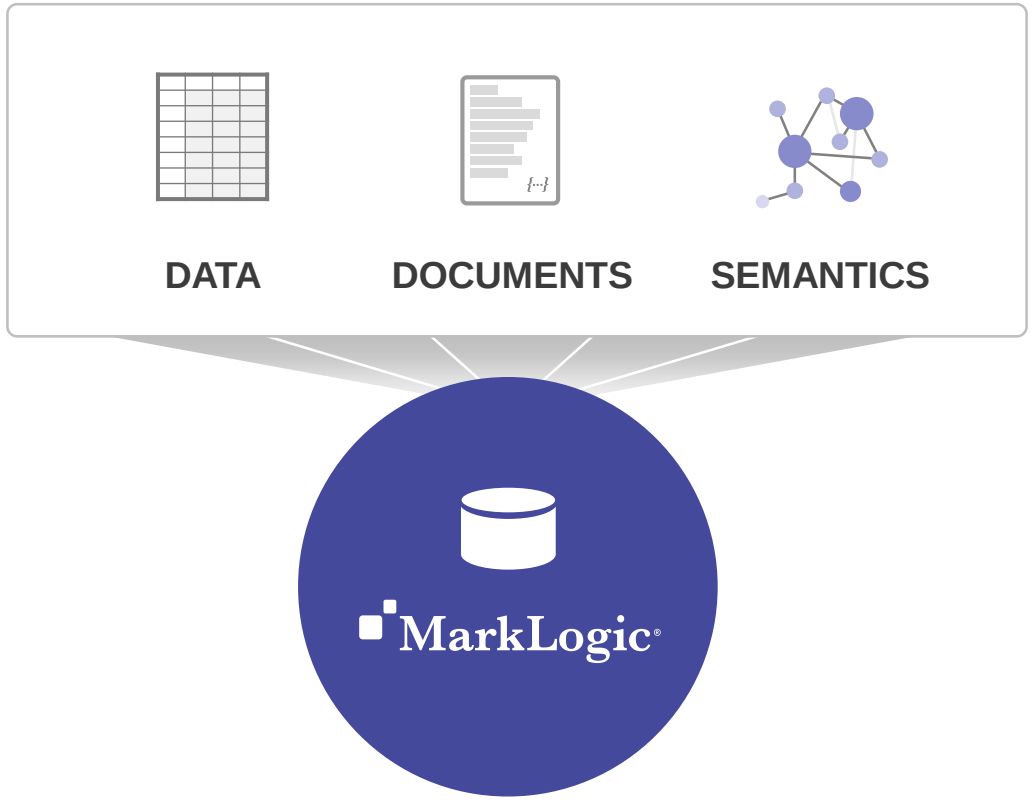
OUR SOLUTION

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

MarkLogic®

Q&A