

Mastering Information for Government



Master Information Needs



Intelligence & Law Enforcement

- ✓ *National Security*
- ✓ *Homeland Security*
- ✓ *Border Security*
- ✓ *Municipal & State Police*
- ✓ *Fusion Centers*
- ✓ *Airline Screening*
- ✓ *Immigration & Customs*



Monetary & Tax

- ✓ *Tax Fraud & Evasion*
- ✓ *Financial Crimes*
- ✓ *Cyber crime*
- ✓ *Anti Money Laundering*
- ✓ *Watch List Checking*
- ✓ *Insider Threat*
- ✓ *Identity Theft*



Social Services & Healthcare

- ✓ *Verify citizen identity*
- ✓ *Determine eligibility for social programs*
- ✓ *Reduce improper payments*
- ✓ *Excluded Individuals and entities (deny services/payment)*
- ✓ *Detect member/provider fraud*

**Predict , Mitigate & Prevent Risk
Create Trusted View (Manage)**

Know your Threat

Know your Citizen

Know your Client

Information Challenges



Multiple Systems of Record

- Data silos exist internal and external to agency or country
- Relevant data not shared
- Value of one piece of data in unknown without context



Privacy Requirements

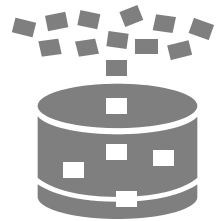
- Minimize exposure of Personally Identifiable Information
- ...while still locating persons with nefarious plans



Public Pressures and High Stakes

- Successfully correlating information means lives are saved.
- Failures to correlate information can be catastrophic and highly visible

How Can Governments Meet These Challenges?



Data Overload

- Less "signal" to much "noise"
- Individual systems over populated
- Incomplete and inconsistent records



Multiple Types of Data

- Different formats
- Different purpose (e.g., about people or things)



Cost Control

- Productivity lost on manual data correlation tasks
- Increasing staff costs to integrate flood of data

Master Information to...

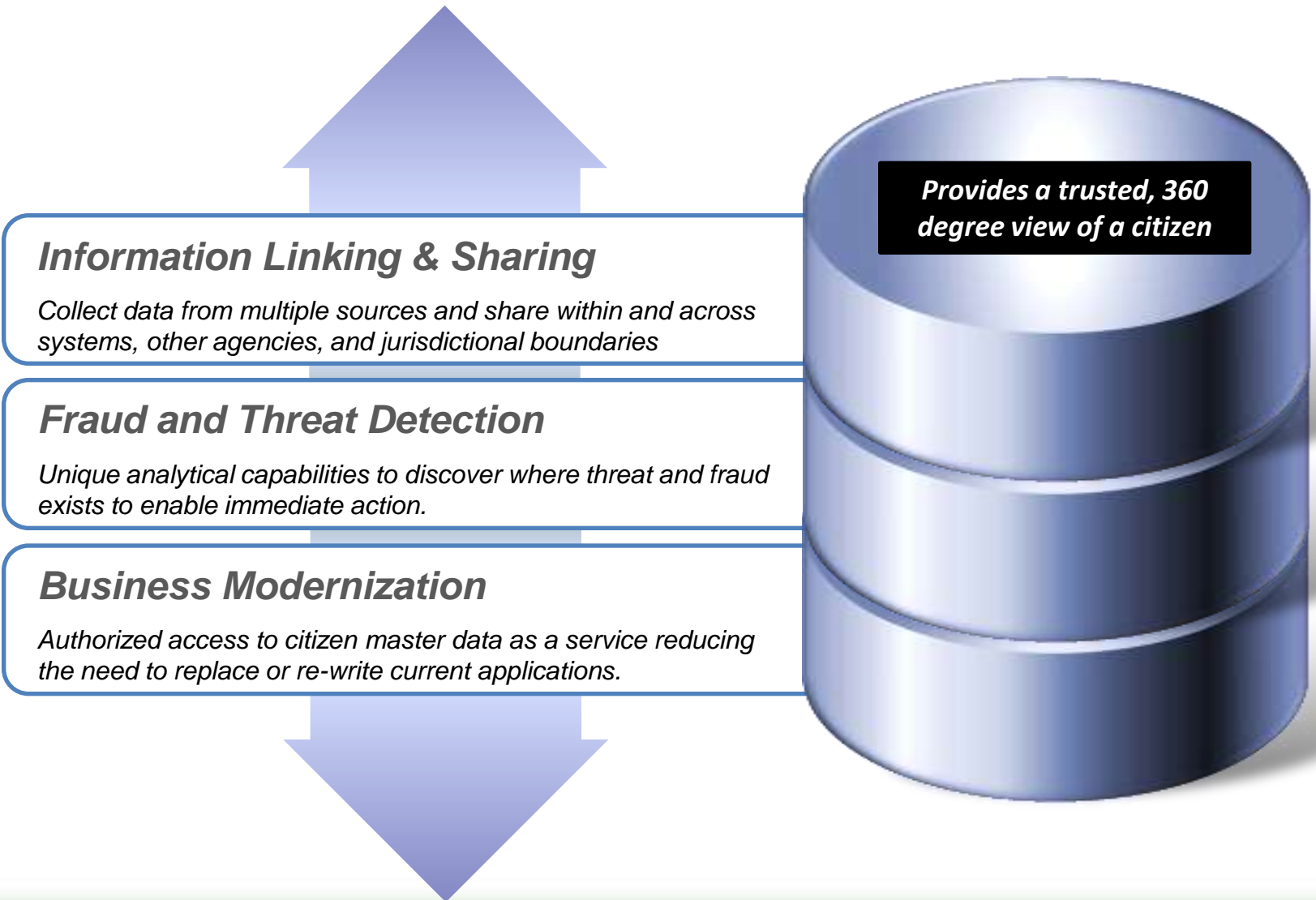


Make information sharing a strategic imperative

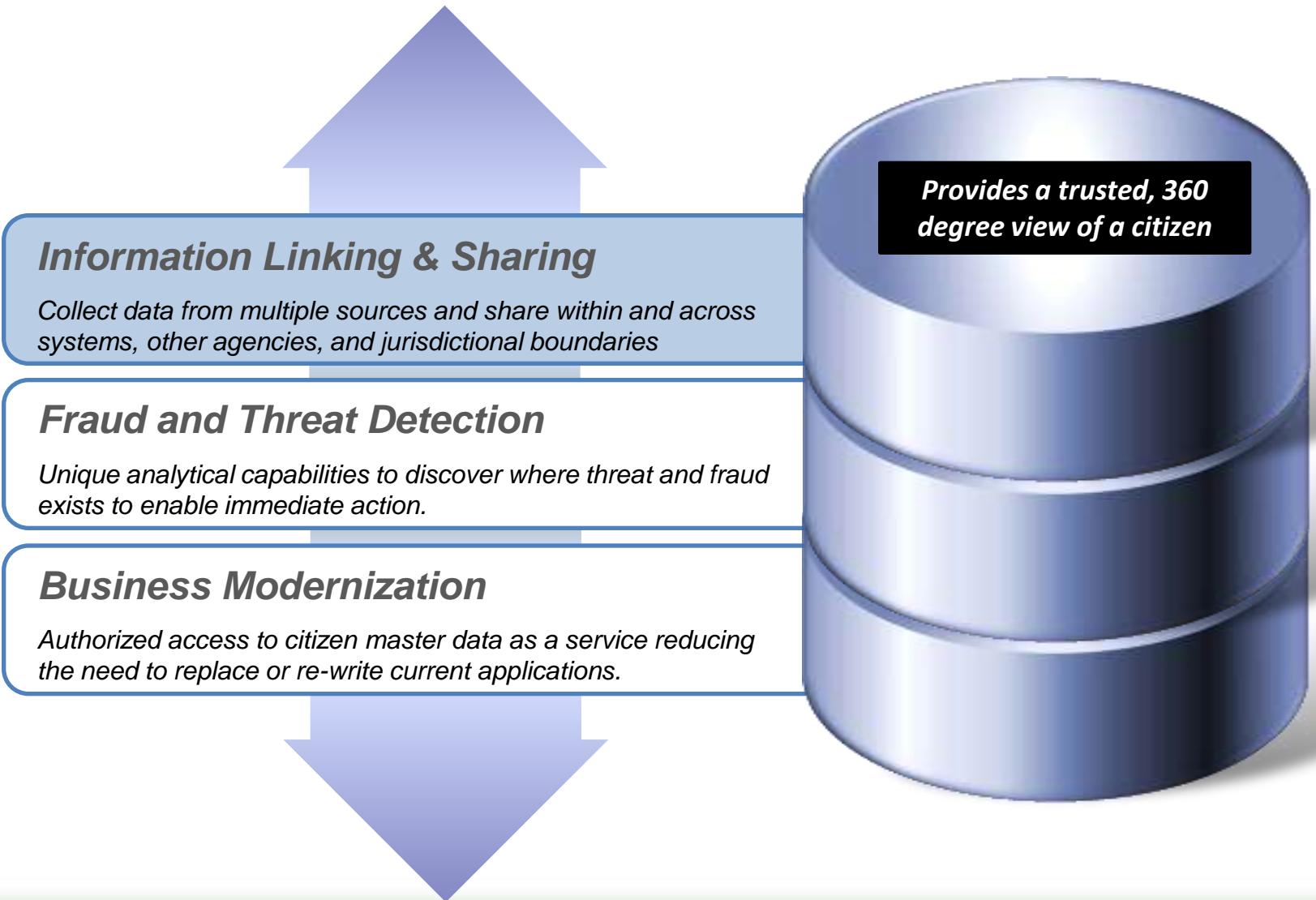
Adapt to an ever changing public safety landscape

Modernize and evolve to operate efficiently

Mastering Information Solutions



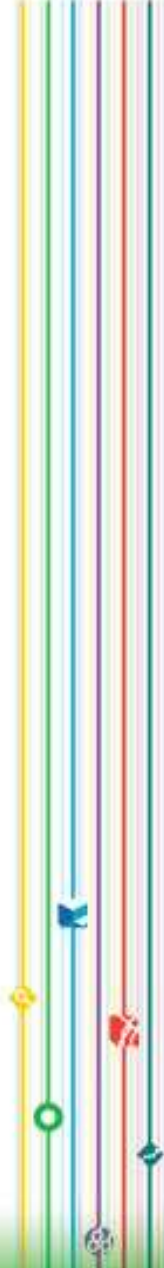
Mastering Information Solutions



Why Do We Need to Share Information?



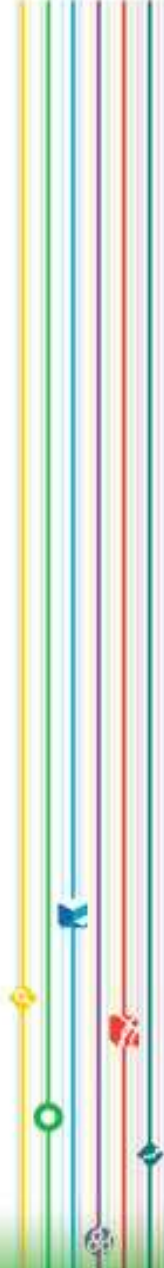
- Victoria Climbié, age 8
- Abused and murdered in 1999 by her guardians
- Individual agencies – police, social services, NHS, and local churches – had contact with her, and noted the signs of abuse.
- No coordination or sharing of information across agencies to recognize a pattern of continue abuse.
- Government investigation recommended a new information sharing system to enable coordination across agencies tasked with serving children.



Why Do We Need to Share Information?

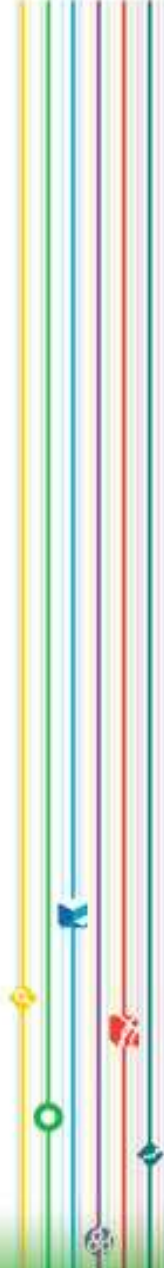


- Umar Farouk Abdulmutallab
- US Dept of State granted a US multiple-entry visit in 2008
- Met Anwar al-Awlaki at the Finsbury Park Mosque in London
- Trained with al-Qaeda in Yemen in 2009
- British intelligence tells US that “Umar Farouk” had pledged to support jihad
- Abdulmutallab’s father warned US embassy that his son had extreme religious views
- On Dec 25, 2009 Abdulmutallab attempts to blow up NW flight 253



Why Do We Need to Share Information?

- No single organization has all of the information
- Gaps in individual data sets
- Combine information to fill in the gaps
- Need to recognize commonality
- Entity resolution is the key to effective information sharing



Entity Resolution Examples

Are these records for the same person?

Attribute	Record 1	Record 2
Name	Rodolfo Gonzalez Lugo	Rodolfo Gonzalez Lugo
Birth date	03-01-1999	03-11-1999
Address		

Entity Resolution Examples

Are these records for the same person?

Attribute	Record 1	Record 2
Name	Felipe Valente Cholo	Felipe Rolando Valente Xolo
Birth date	27-06-2005	27-07-2005
Address	Villa de Cos, Veracruz	

Entity Resolution Examples

Are these records for the same person?

Attribute	Record 1	Record 2
Name	Sarahi Alberto Mendez	Sarai Mendez Alberto
Birth date	01-03-1987	07-03-1987
Address	13 Bravo, Agua Nueva, Veracruz	Agua Nueva, Veracruz

Example Information Sharing Requirements

Information is the lifeblood of policing. Managing the service's information assets effectively and making them available wherever they are needed, regardless of force boundaries, is crucial. Forces hold more than 65 million operational records in more than 350 separate local databases.

Most of these local databases don't communicate with each other, and consequently, information collected by one force is not available to colleagues in others. As a result, when it comes to exploiting their information assets, forces are hampered by artificial geographical and system boundaries, while criminals operate freely across borders, exploiting the lack of 'joining up'.

The program is radically improving the ability of the police service to share information across force boundaries, improving police effectiveness nationally, regionally and locally.

- *Law enforcement "information management" program*

Example Information Sharing Requirements

Lack of access to other agency data limits our ability to leverage whole of border information holdings, such as potential benefits in agencies having a complete picture of the border interventions in relation to a person, craft or consignment. This includes information held by non-border agencies such as the Police.

While individual agencies typically develop information systems against their own standards and requirements, there isn't agreement as to what standards should apply across the border. Information exchange is therefore complicated by the need to translate data, and to build and maintain interfaces to cope with different messaging technologies.

- *Border sector
“information management”
program*

Example Information Sharing Requirements

Currently police officers and staff use a number of key systems in support of their policing work across the process areas identified in the program scope. The systems and data are not well integrated so that users frequently have to log into multiple systems, re-key data and search across multiple systems for relevant data. The implication of this is that users do not have a complete picture / profile of the person leading to risks to public and officers.

There is a need to move away from the traditional method of event structured data to POLE (people, objects, locations, events) structured data. Currently data, across nearly all systems, is stored by 'event' and all associations, including people, are to that event. This makes it extremely difficult to identify a person across the multiple events.

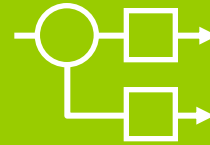
- *Law enforcement "information improvement" program*

Information Sharing Requirements



Dirty Data

Data from different shared sources varies in format, consistency, and quality.



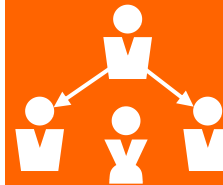
Ease of Integration

Multiple applications will want to consume the shared data. The data should be available via service-oriented architectures



Multi-cultural Data

Data from multiple sources comes in multiple languages and scripts



Multiple 'Versions of the Truth'

Different stakeholders may have different approaches to resolving identities and different perspectives on which data are accurate.



Sensitive Data

Prevent specific organizations and users from having access to or changing your data, down to the attribute level



Automatic Discovery

Users want to be notified when other source share information that relates to their work

Information Sharing Capabilities



Dirty Data

Data from different shared sources varies in format, consistency, and quality.

Self-learning algorithms **automatically determine the value of the data**

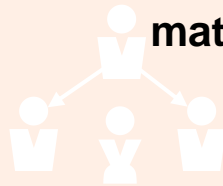
Matching criteria can vary by source

Fuzzy matching **algorithms that can match on partial information**

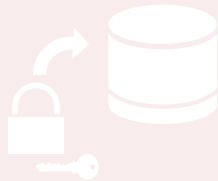


Multi-cultural Data

Data from multiple sources comes in multiple languages and scripts



Different stakeholders may have different approaches to resolving identities and different perspectives on which data are accurate.



Sensitive Data

Prevent specific organizations and users from having access to or changing your data, down to the attribute level

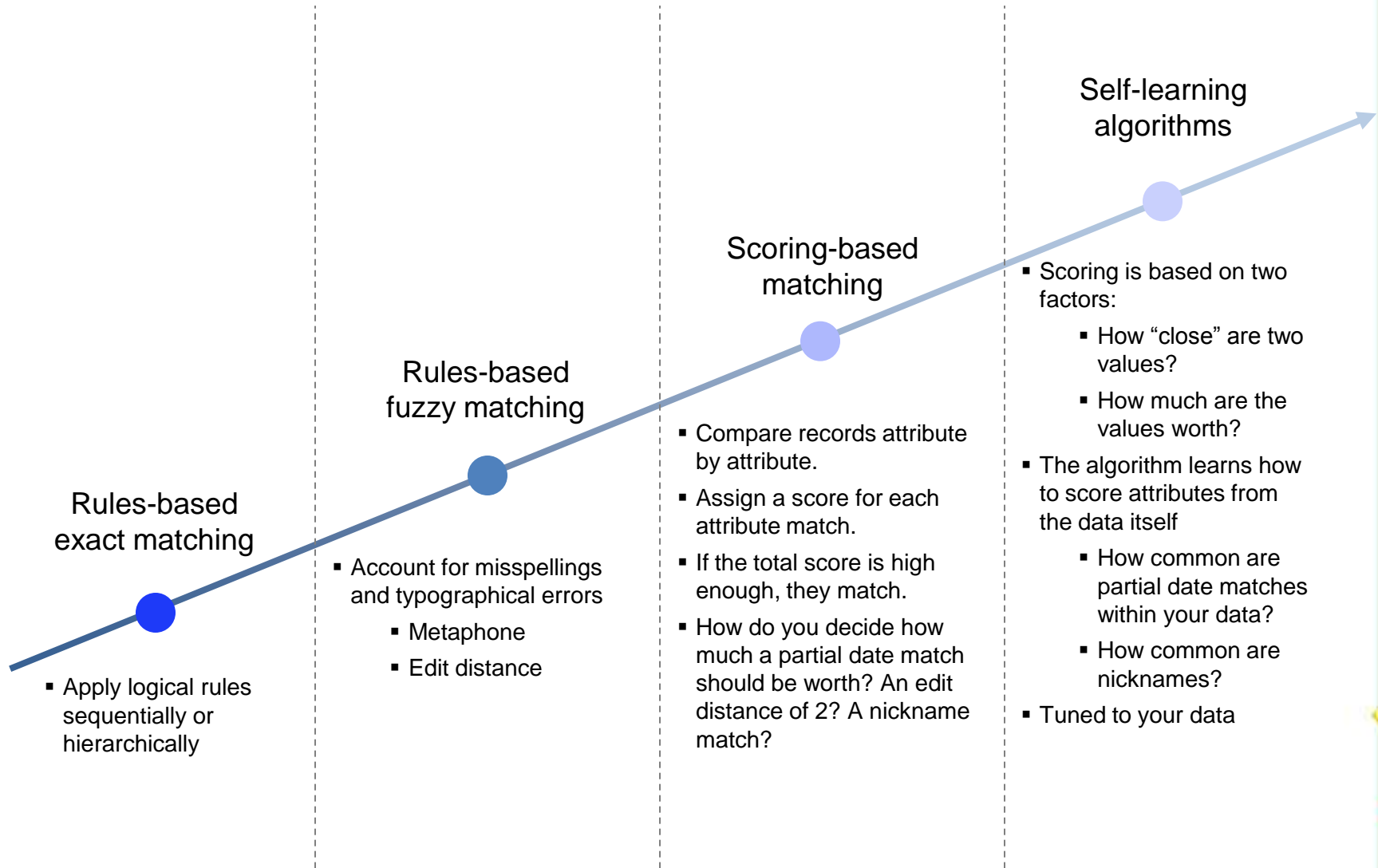


Automatic Discovery

Users want to be notified when other source share information that relates to their work



Self-Learning Algorithms



Information Sharing Capabilities



Dirty Data

Co-exist with existing systems & architectures. No need to create a centralized data store or modify source data



Multi-cultural Data

Data from multiple sources comes in multiple languages and scripts



Sensitive Data

Prevent specific organizations and users from having access to or changing your data, down to the attribute level



Match within languages and scripts (e.g., Arabic to Arabic)

Full double-byte support

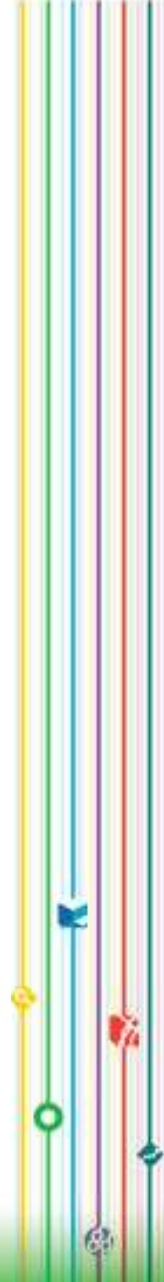


Match across scripts with transliteration and cross-script phonetics



Automatic Discovery

Users want to be notified when other source share information that relates to their work



Information Sharing Capabilities



Dirty Data

Co-exist with existing systems & architectures. No need to create a centralized data store or modify source data



Multi-cultural Data

Data from multiple sources comes in multiple languages and scripts



Sensitive Data

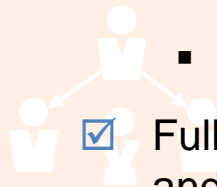
Prevent specific organizations and users from having access to or changing your data, down to the attribute level



Role-based access controls at multiple levels:

- Interaction
- Source

Users will want to consume the shared data. The data should be available via service-oriented architectures



Attribute

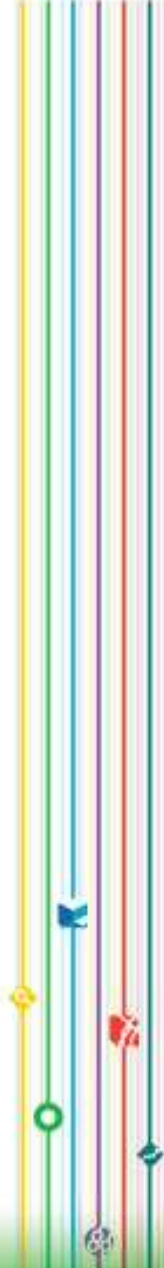
Full audit trail of reads, writes, and matches

Multiple 'Versions of the Truth' stakeholders may have different approaches to resolving discrepancies on which data are accurate.



Automatic Discovery

Users want to be notified when other source share information that relates to their work



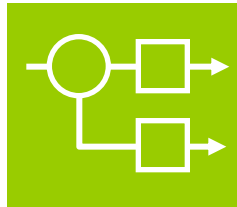
Information Sharing Capabilities

- ✓ Lightweight, embedded ETL for data ingest and extract or compatibility with Information Server
- ✓ SOA-based services for integration into consuming applications
- ✓ Option to deploy as a lightweight index without replicating all the source data



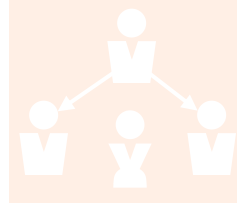
Sensitive Data

Prevent specific organizations and users from having access to or changing your data, down to the attribute level



Ease of Integration

Multiple applications will want to consume the shared data. The data should be available via service-oriented architectures



Multiple 'Versions of the Truth'

Different stakeholders may have different approaches to resolving identities and different perspectives on which data are accurate.



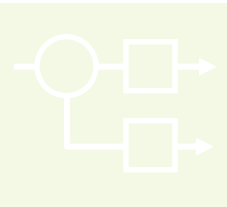
Automatic Discovery

Users want to be notified when other source share information that relates to their work



Information Sharing Capabilities

- ✓ Ability to apply multiple algorithms concurrently to the same underlying data
- ✓ Dynamic composite views give each stakeholder their own view of resolved entities



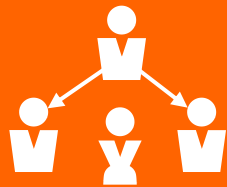
Ease of Integration

Multiple applications will want to consume the shared data. The data should be available via service-oriented architectures



Multi-cultural Data

Data from multiple sources comes in multiple languages and scripts



Multiple 'Versions of the Truth'

Different stakeholders may have different approaches to resolving identities and different perspectives on which data are accurate.



Sensitive Data

Prevent specific organizations and users from having access to or changing your data, down to the attribute level

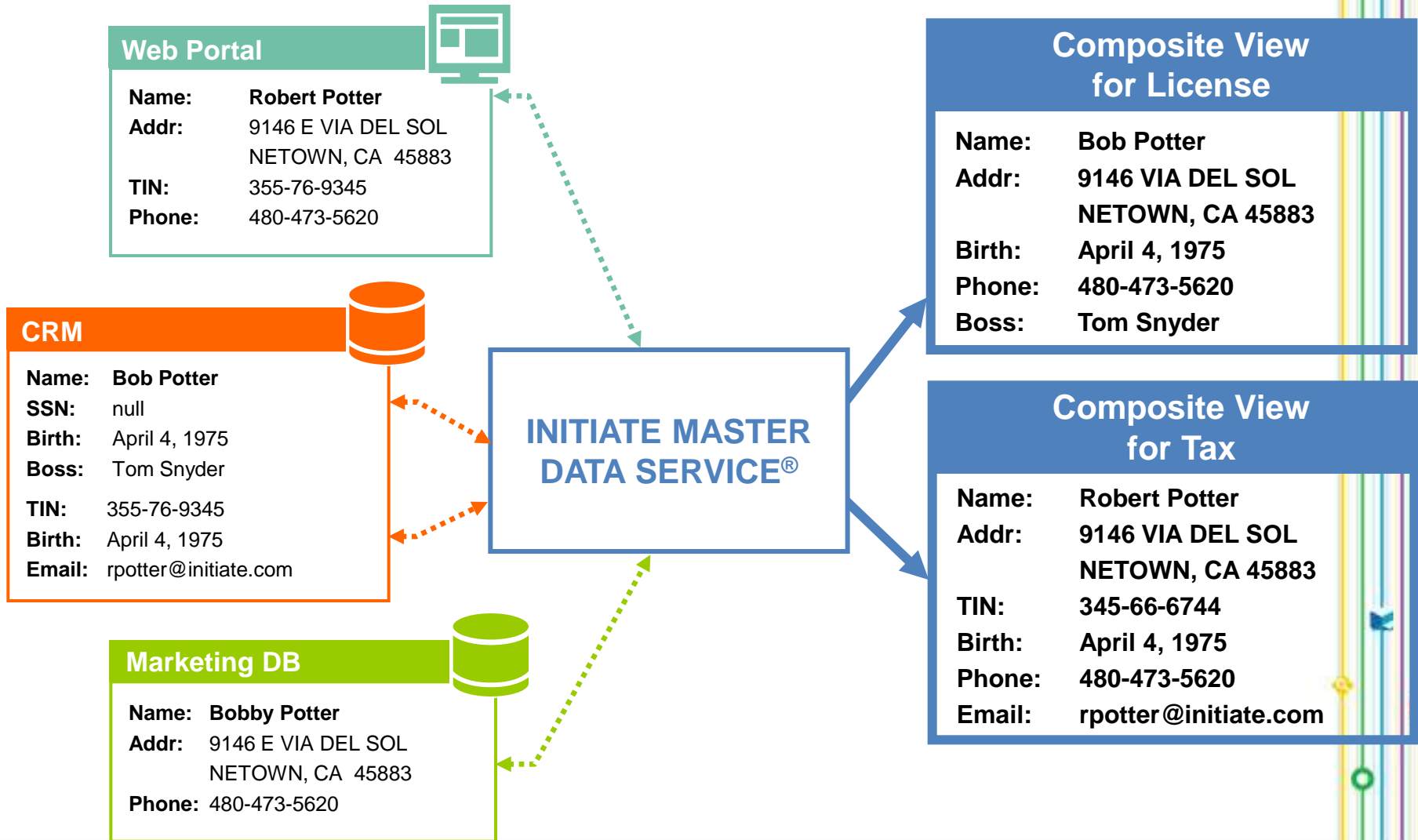


Automatic Discovery

Users want to be notified when other source share information that relates to their work



Composite Views



Information Sharing Capabilities

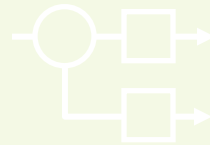
- ✓ Identify where other stakeholders have information of interest to me
- ✓ Allow users to save searches and subscribe to specific entities

- ✓ Notify users when changes occur within subscribed networks



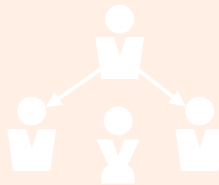
Sensitive Data

Prevent specific organizations and users from having access to or changing your data, down to the attribute level



Ease of Integration

Multiple applications will want to consume the shared data. The data should be available via service-oriented architectures



Multiple 'Versions of the Truth'

Different stakeholders may have different approaches to resolving identities and different perspectives on which data are accurate.



Automatic Discovery

Users want to be notified when other source share information that relates to their work

Case Study: N-Dex



FBI National Data Exchange Program

- 18,000 law enforcement agency participants
- Full Depth:
 - Local
 - State
 - Regional
 - National

Cross Boundary:

- Share data across disparate systems & political boundaries

- *National System for the Integration & Discovery of Criminal Justice Information*
 - Sharing of complete, accurate, timely, and useful information across jurisdictional boundaries and to provide new investigative tools that enhance the nation's ability to fight crime and terrorism
- *Leveraged Technology for Relation of Massive Amounts of Data into Useful Information*
 - Correlate data received from data suppliers and apply sophisticated business rules that may proactively notify specific users if certain relationships are discovered
 - Provide users with a single point of discovery to “connect the dots”
- *Case Sensitive Information - Multi-tier Data Access*
 - Three tier – Full, Pointer and Restricted
 - Controlled sharing of correlative data
- *Leverages Existing Standards, Systems & Networks*
 - Builds upon established trusted sharing networks
 - Provides for national capability

Case Study: Classified



Definition

- ▶ 100,000 inbound transactions per day requiring real-time entity resolution against:
 - Derogatory Data Sets
 - Watchlists
 - Open Source Data Sets
 - Internal Data Sets

Cross Boundary

- ▶ Share data across disparate systems and political boundaries

Defined

- Name matching and Relationship/Correlation analysis
- Secure information-sharing and real-time entity resolution for Intelligence Agencies
- Enables multiple agencies to collaborate in detection of real-time threats to national security by connecting the dots between people, places and events
 - Alerts on inbound data transactions from sister agencies containing common characteristics across derogatory, open source and internally owned data sets
 - Searches by similarities across various attributes including addresses, affiliations, associates, etc.—linking individuals, places, and things
 - Threat level assessments of individuals and addresses
 - Visualization and mapping features.

Information Security

- Access to information is strictly controlled by the agency who “owns” the info - agency decides what data to share, with whom and under what circumstances

Case Study: ContactPoint



- ▶ Every Child Matters, national program
- ▶ 400 million records
- ▶ 11 million children
- ▶ 125,000 initial users
- ▶ 1,000 data sources
- ▶ Four major national organizations:
 1. Department for Children, Schools and Families (DCSF)
 2. National Health Service (NHS)
 3. Office of National Statistics (ONS)
 4. Department of Works and Pensions (DWP)

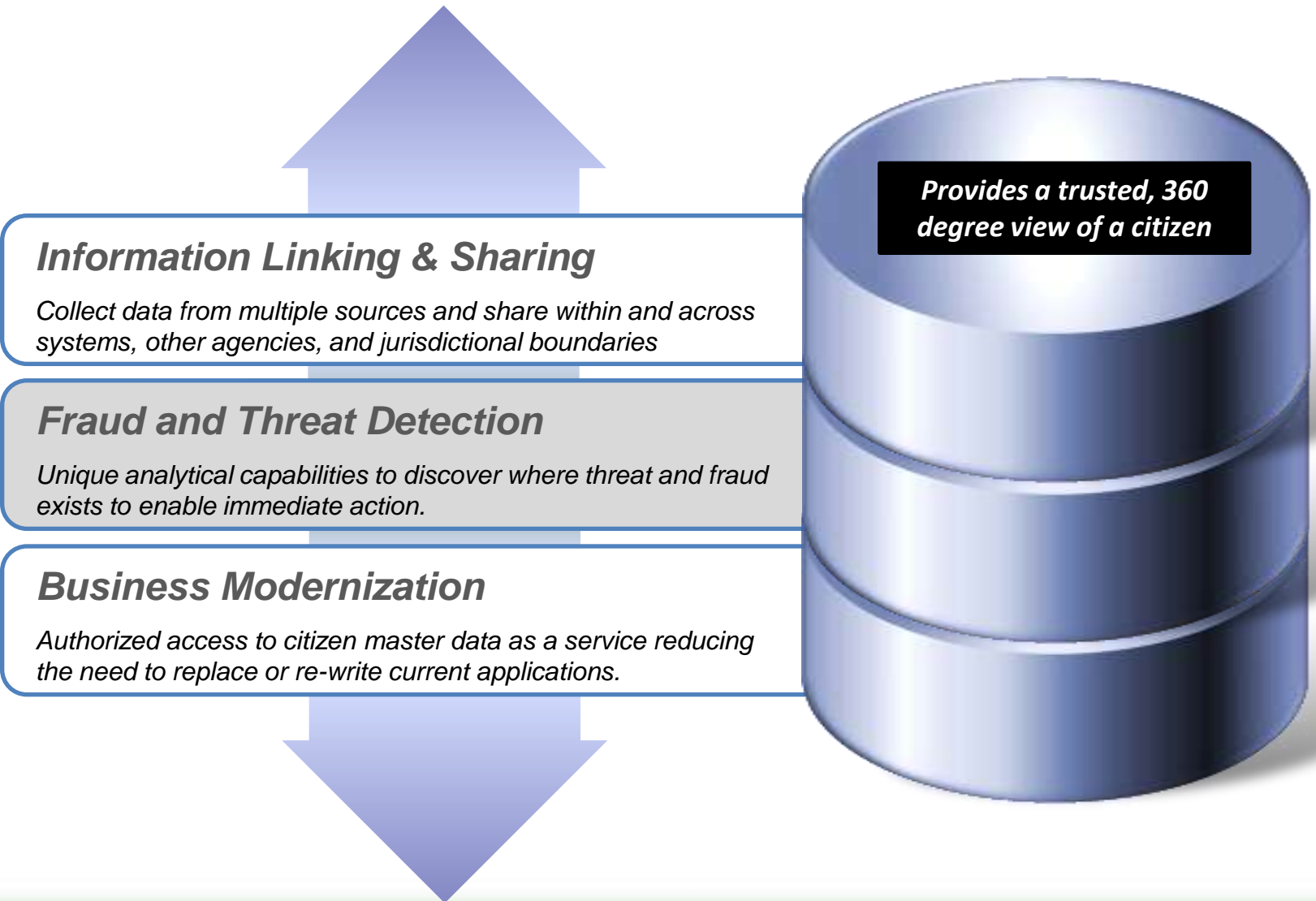


ContactPoint is a key element of the Every Child Matters initiative in the UK, transforming child services through prevention and early intervention

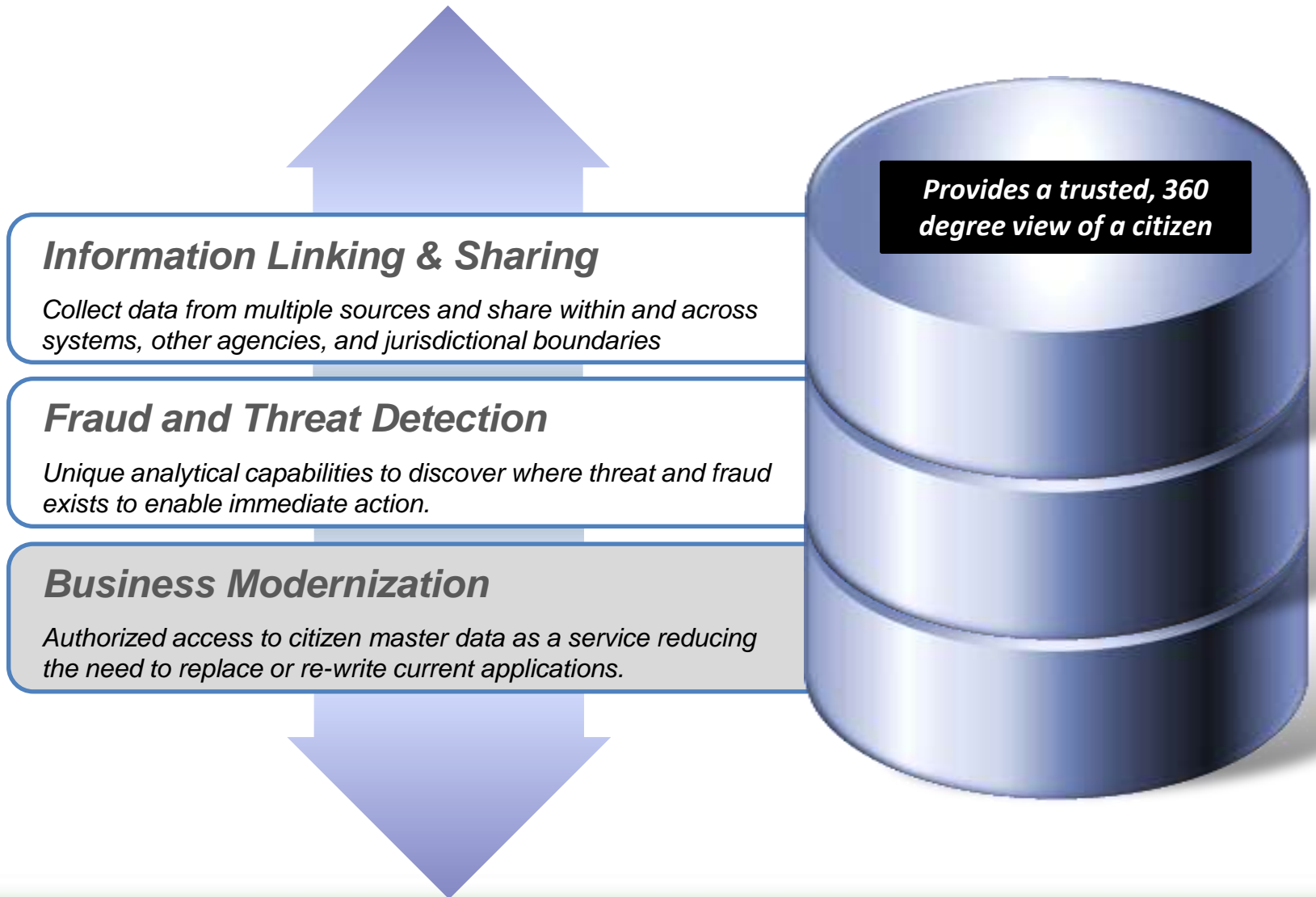
Project Goals:

- ▶ **Support effective delivery of children services**
index basic identifying information for every child in England from birth to age 18 – no case information
- ▶ **Provide secure access to authorized workers**
give access to workers that have been security-checked, trained and have necessary authentication
- ▶ **Leverage sharing of Master Index**
facilitate appropriate information sharing from four major national organizations to enable earlier identification and intervention and a more complete service delivery

Fraud & Threat Detection



Business Modernization



IBM

Software
Summit 2010 

Thank You

