IBM Entity Analytic Solutions

Perpetual Analytics

Data finds the data … Relevance finds the user

Jeff Jonas, Chief Scientist, IBM Entity Analytics
Blogging at www.JeffJonas.TypePad.com

My Background

- Founded Systems Research & Development (SRD) in 1983
- Moved headquarters to Las Vegas in early 90's
- Worked with the gaming industry to help them better understand who they were doing business with – resulting in the NORA (Non-Obvious Relationship Awareness) technology
- Funded by In-Q-Tel, the CIA's venture capital arm, in 2001
- Brought in a professional management team in 2002
- Acquired by IBM January 2005
- Now the Chief Scientist of the IBM Entity Analytics

If a .6% difference matters this much…

… no wonder traditional information systems lack so much intelligence!

Knowing What You Know and Having an Integrated Picture
is Fundamental to Real-time Understanding

=

Persistent Context
is Fundamental to Perpetual Analytics

Domains of Context

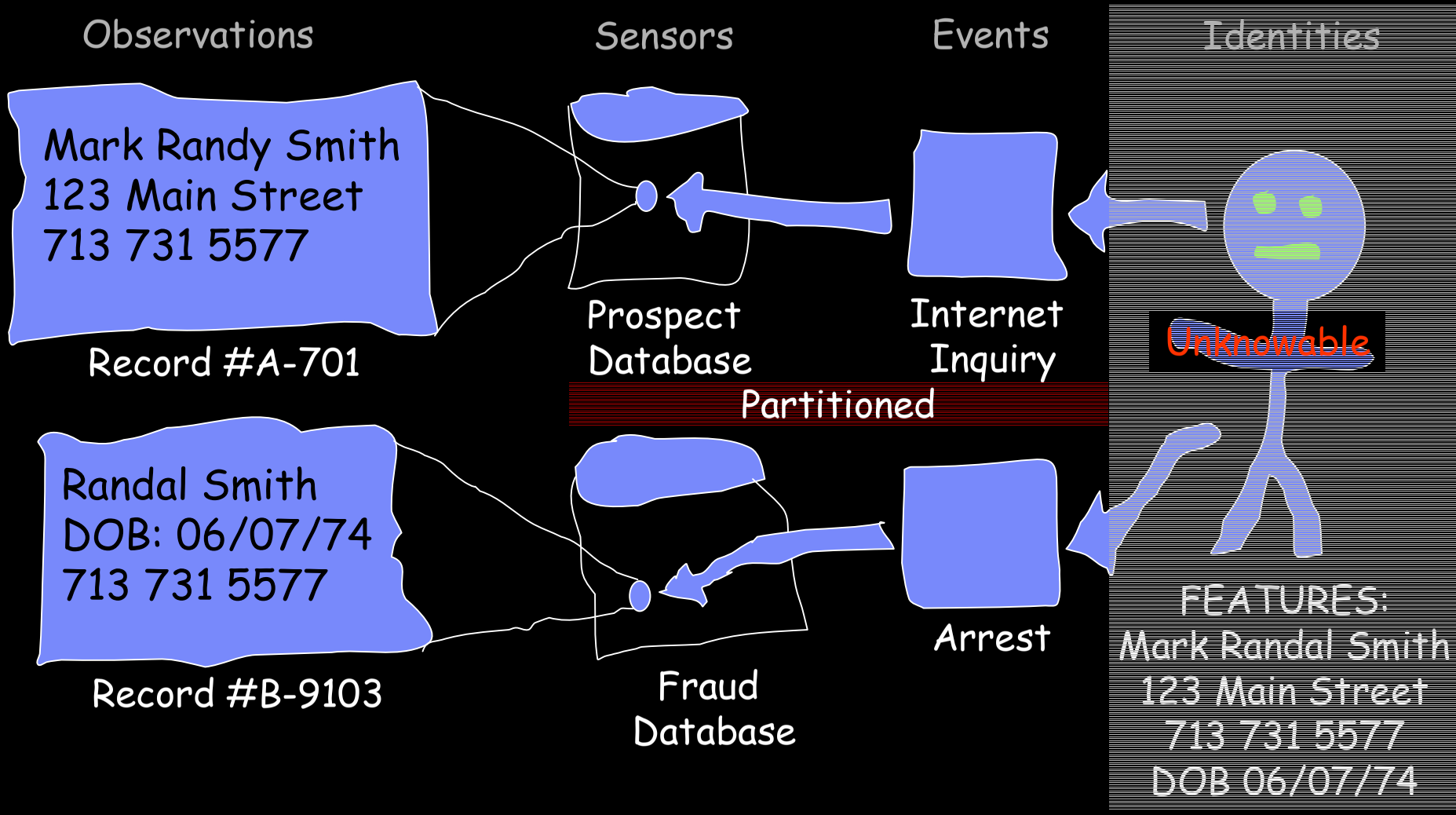
- Who, what, where, when
- Same, similar, related, dissimilar

Scope of this Presentation

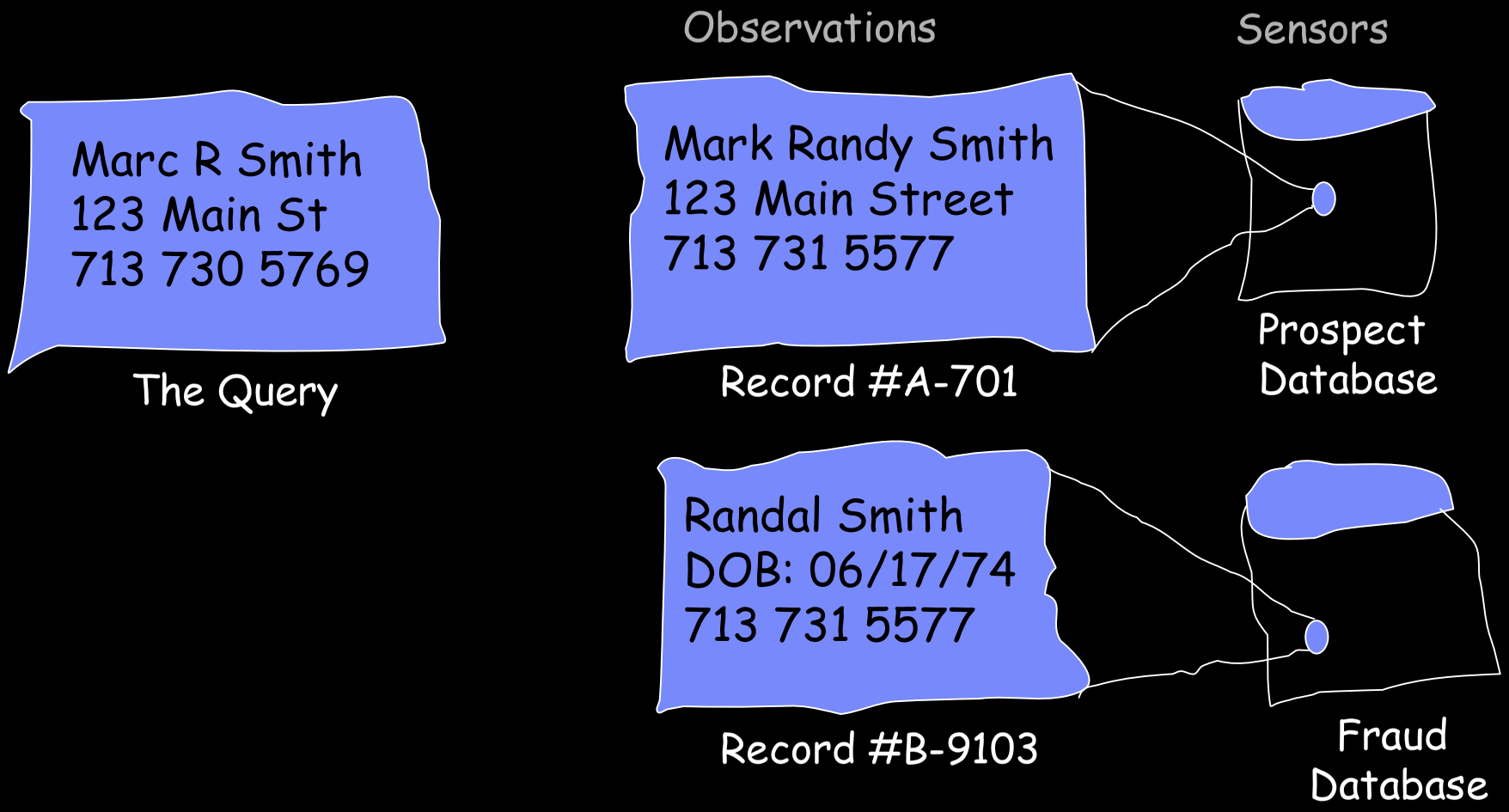
- Who = Identities (people and organizations)
- Same = Semantic Resolution

The Role of Identity Resolution Towards Perpetual Analytics.

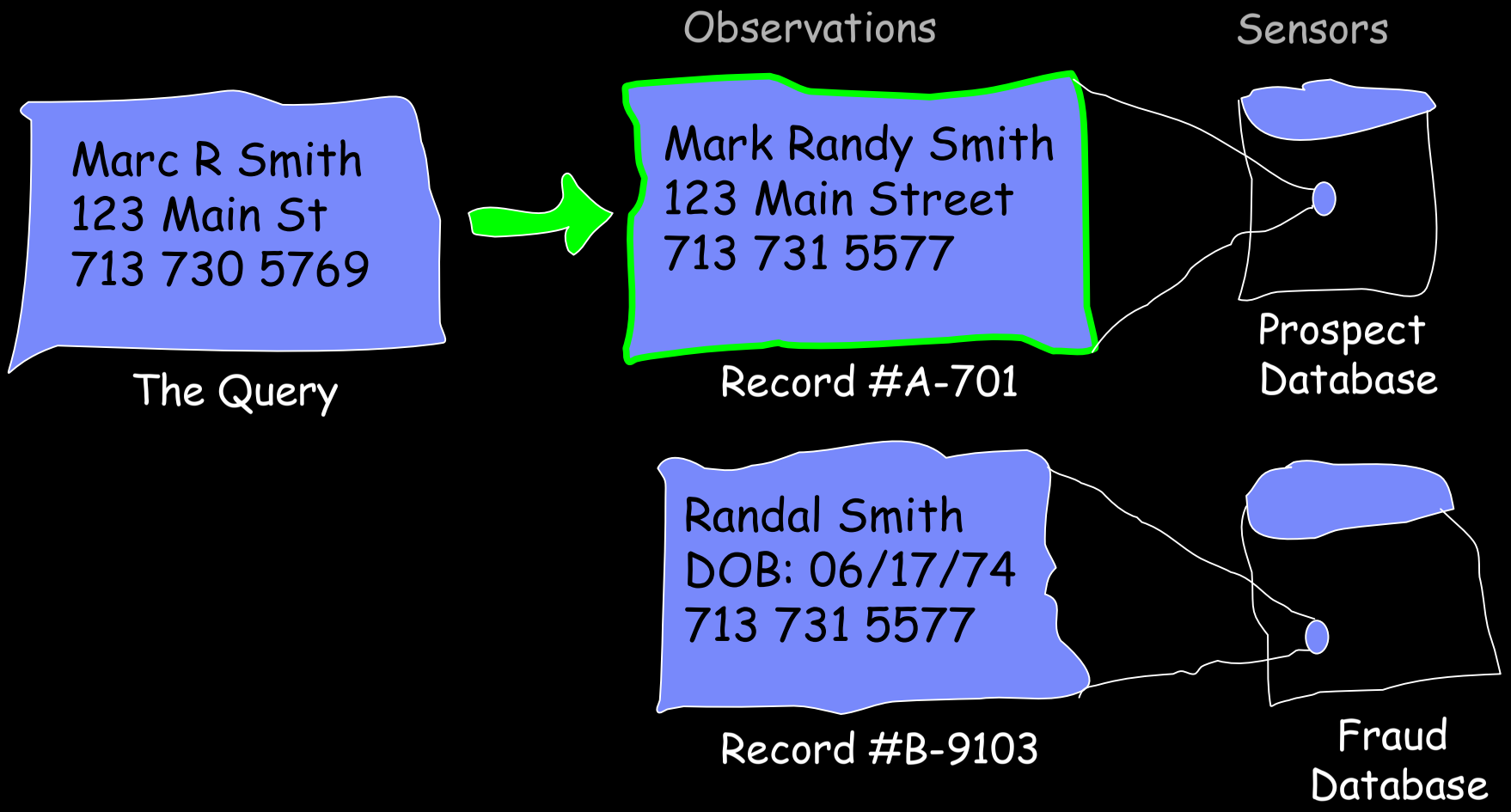
Identities, Events, Sensors and Observations



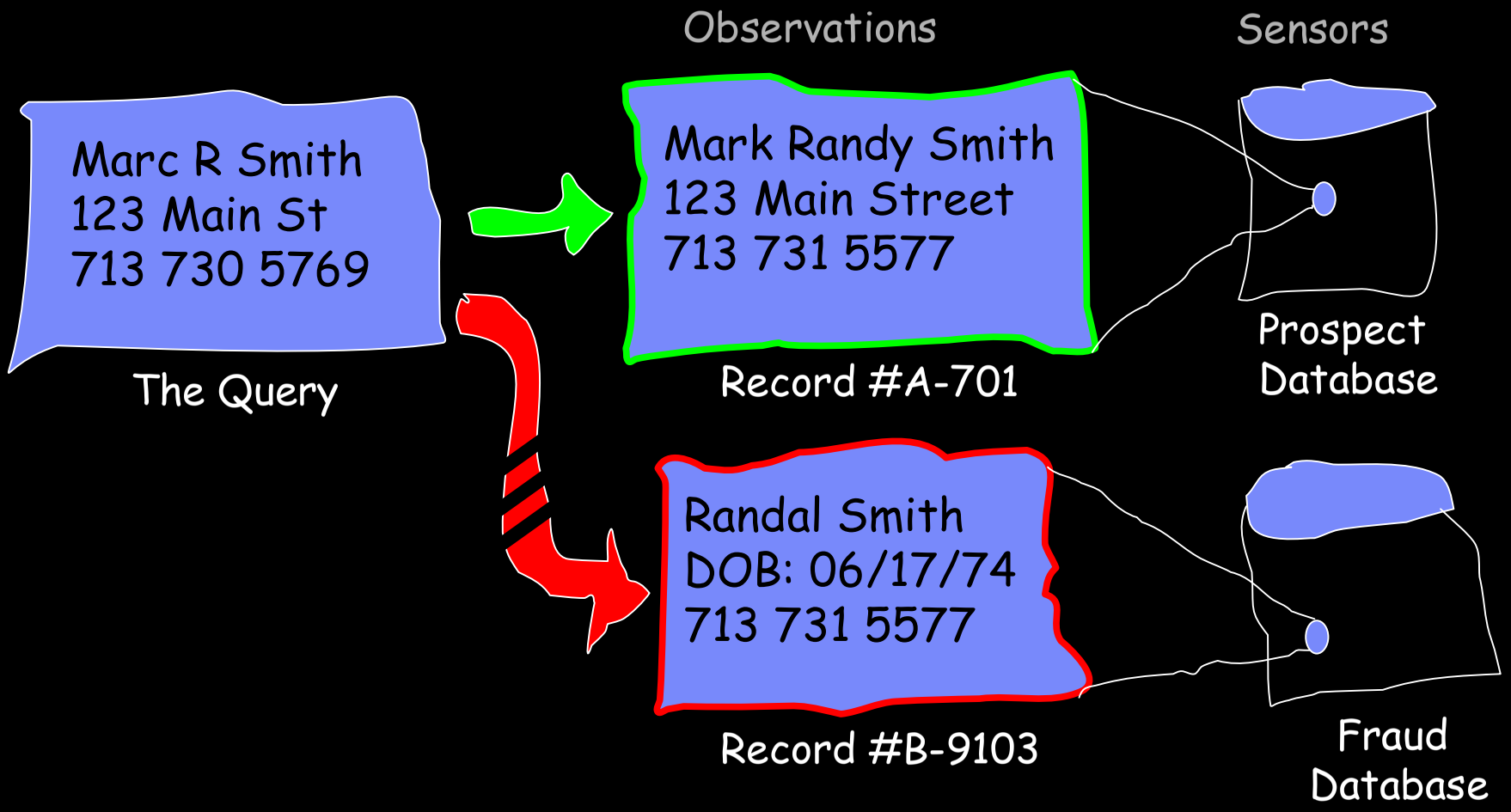
Consider the Query Against the Observables



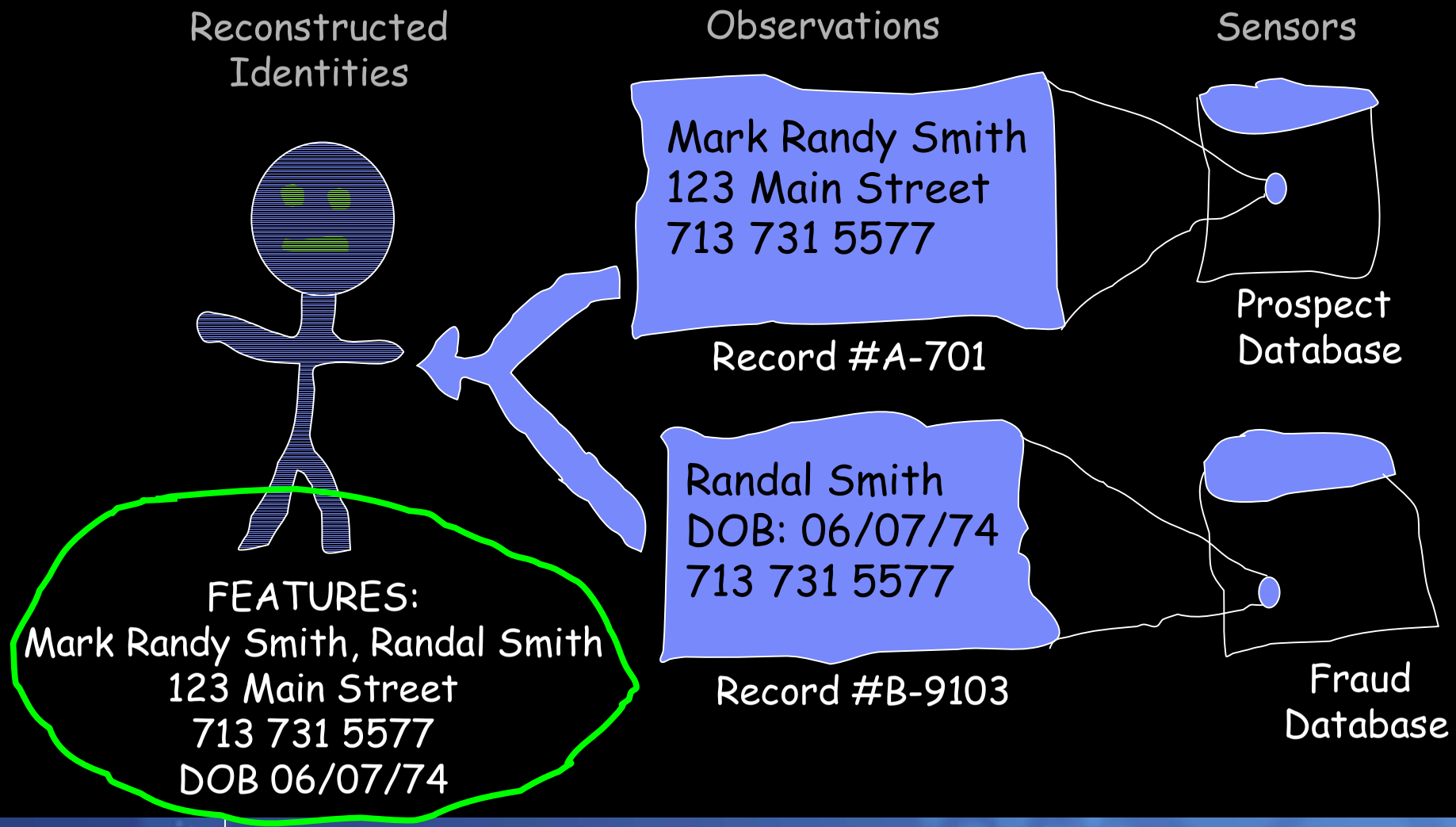
Discoverable



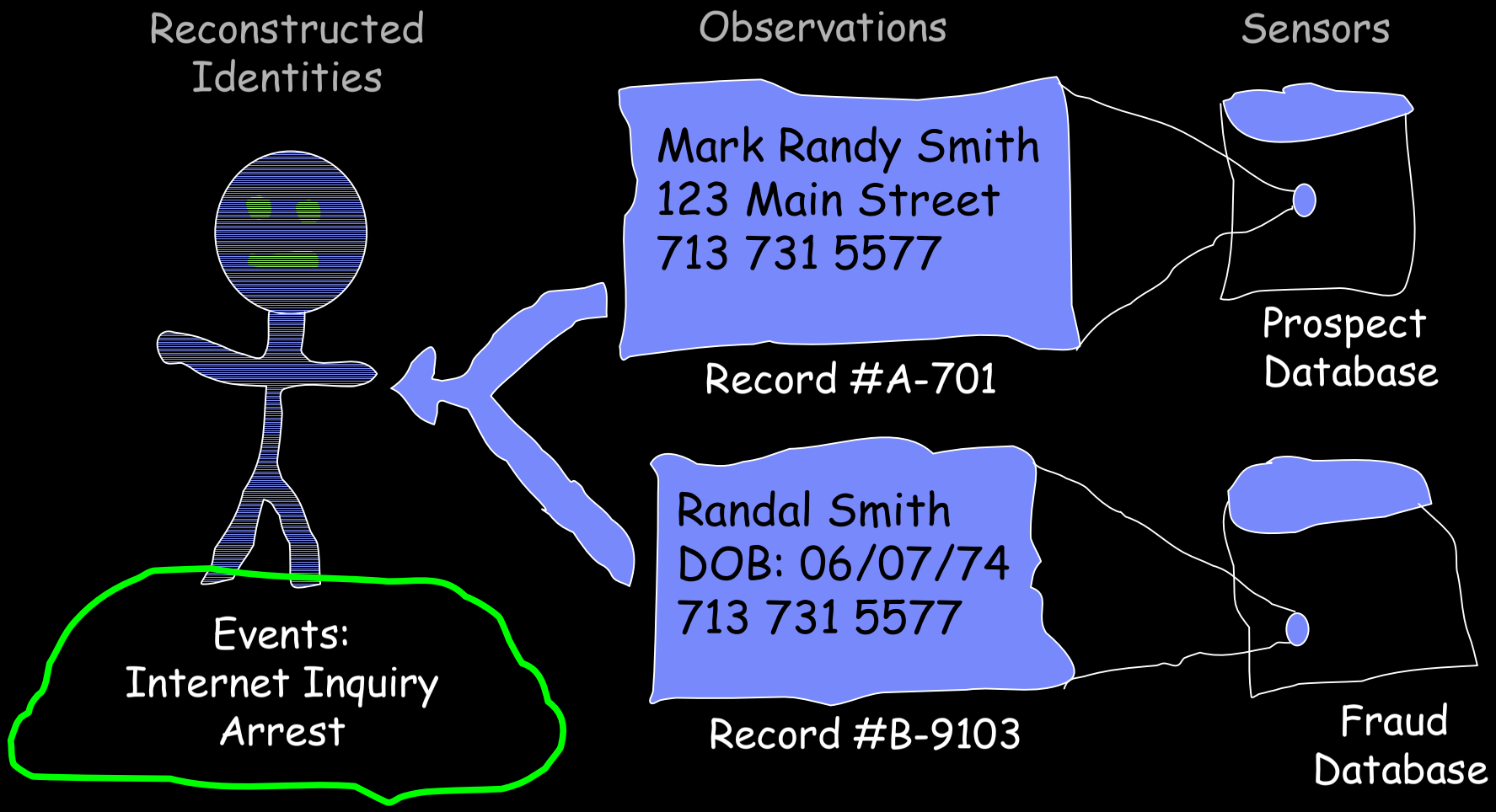
Other Observables … are Undiscoverable



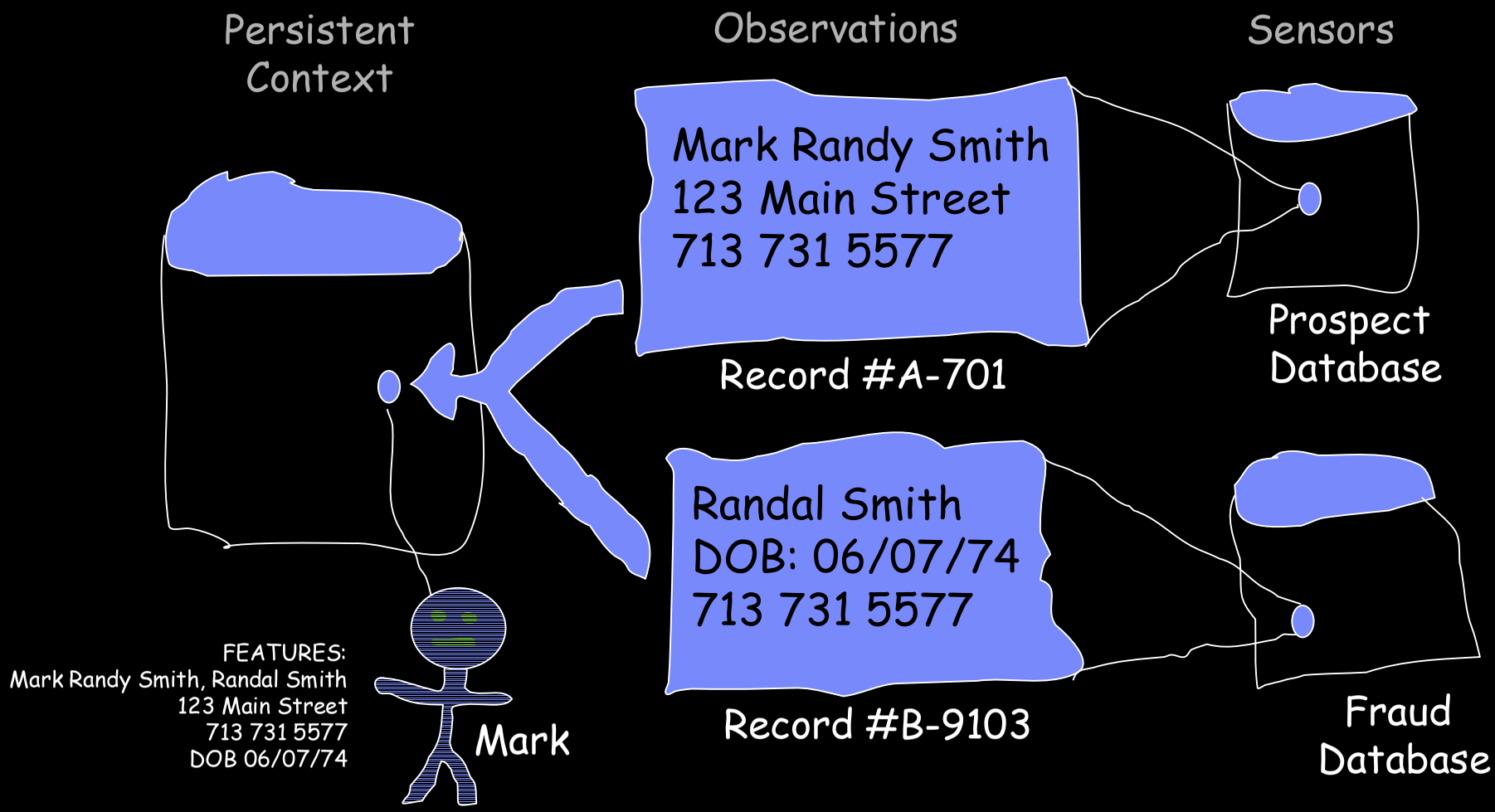
Identity Resolution … Reconstructs Context



Feature and Event Reconstruction

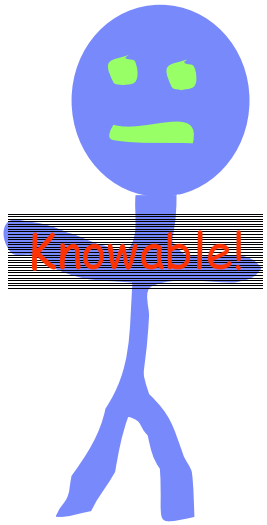


Constructed Context is Persisted in a Database



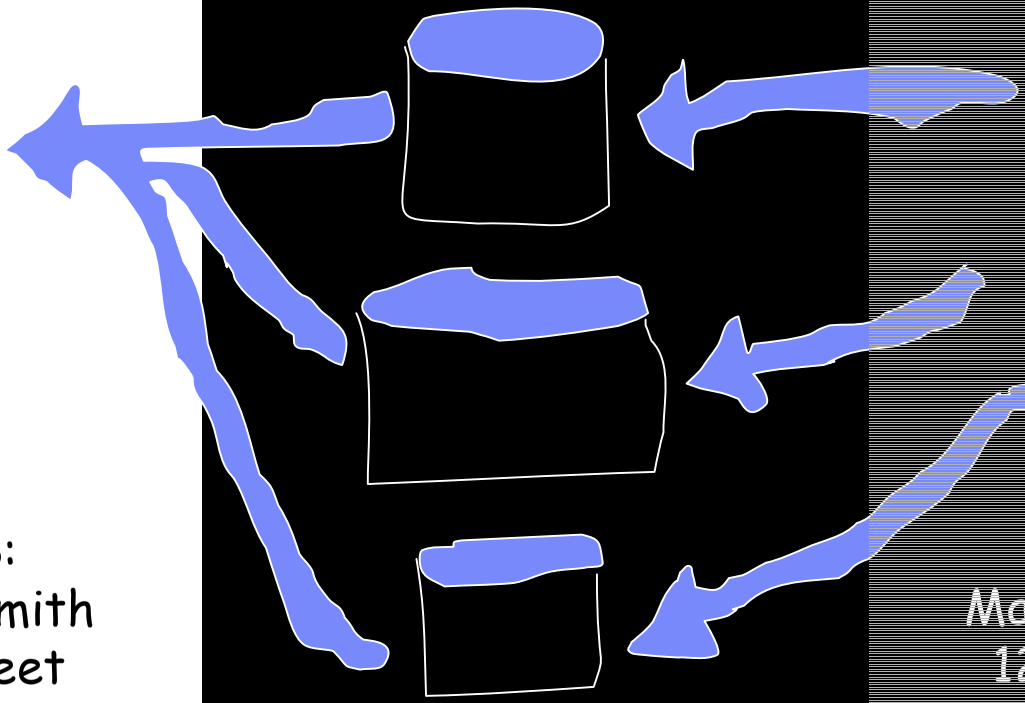
Multi-Sensor Fusion Re-Constructs the Unknowable

Reconstructed Identity



FEATURES:
Mark Randal Smith
123 Main Street
713 731 5577
DOB 06/07/74

Sensors

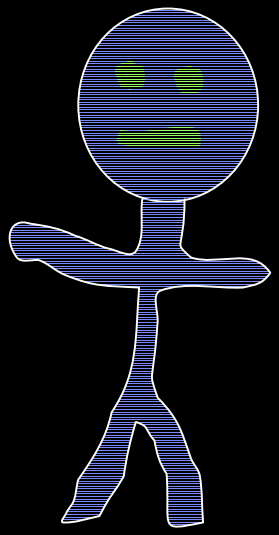


Identity

FEATURES:
Mark Randal Smith
123 Main Street
713 731 5577
DOB 06/07/74

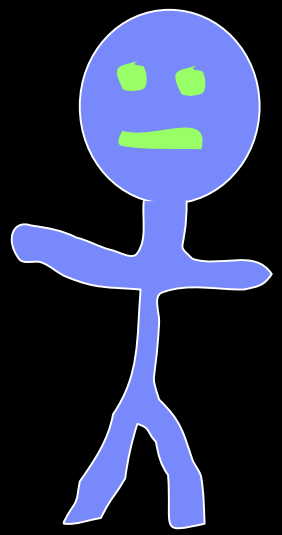
More Observations = Better Reconstruction

2 Observations

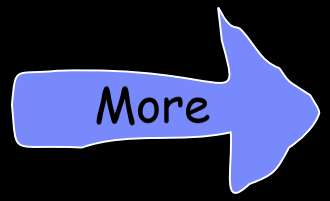


FEATURES:
Mark Randy Smith, Randal Smith
123 Main Street
713 731 5577
DOB 06/07/74

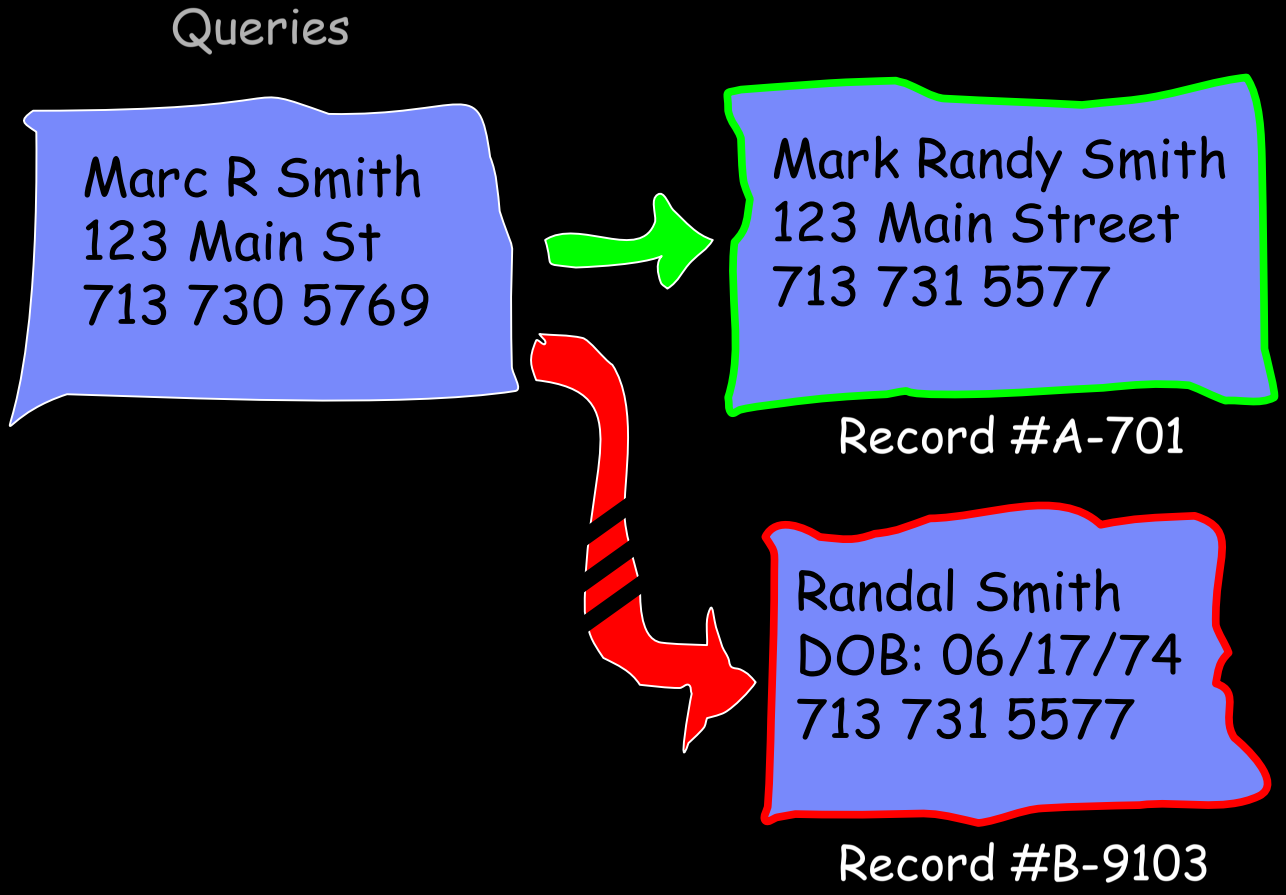
6 Observations



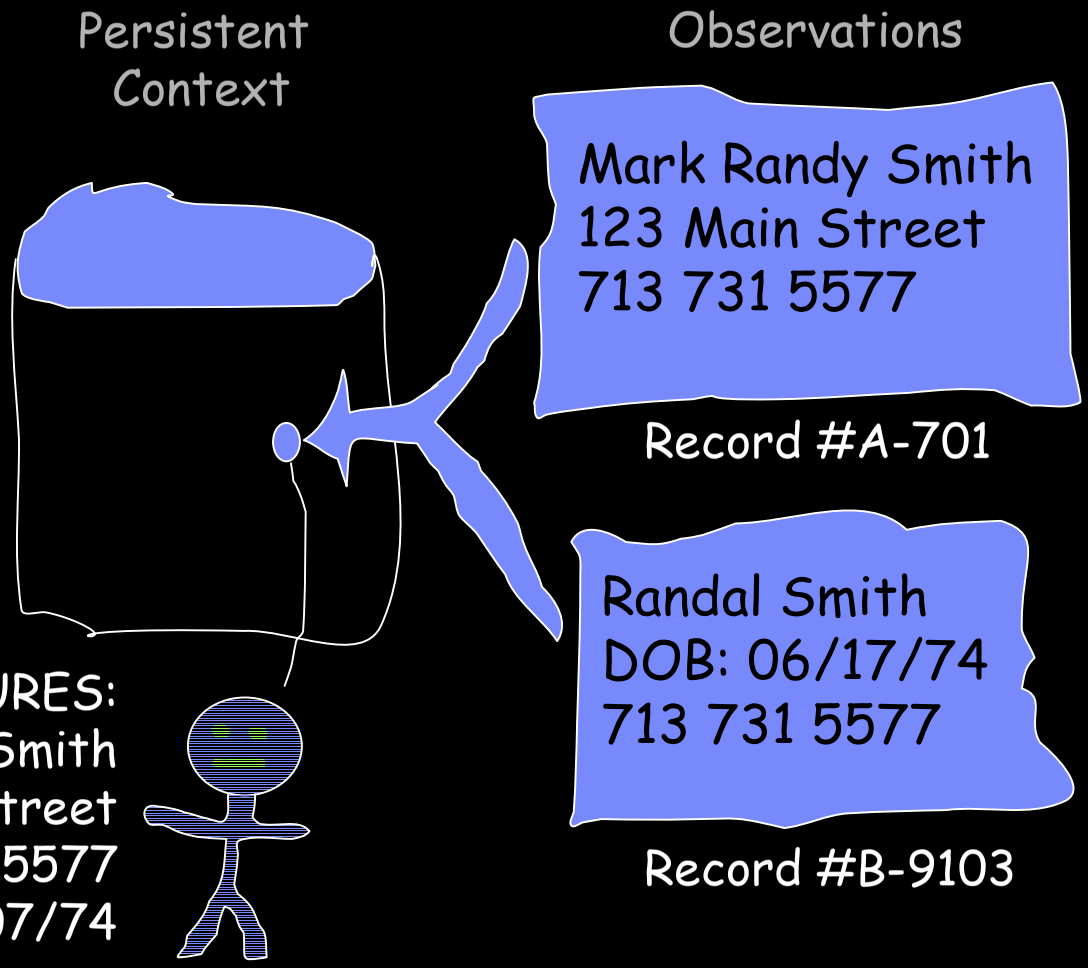
FEATURES:
Mark Randy Smith, Randal Smith, Randy Smith
123 Main Street, Flat 6 20 Lennox Gardens
713 731 5577, 796 064 03 04
DOB 06/07/74, Passport: 001003429002



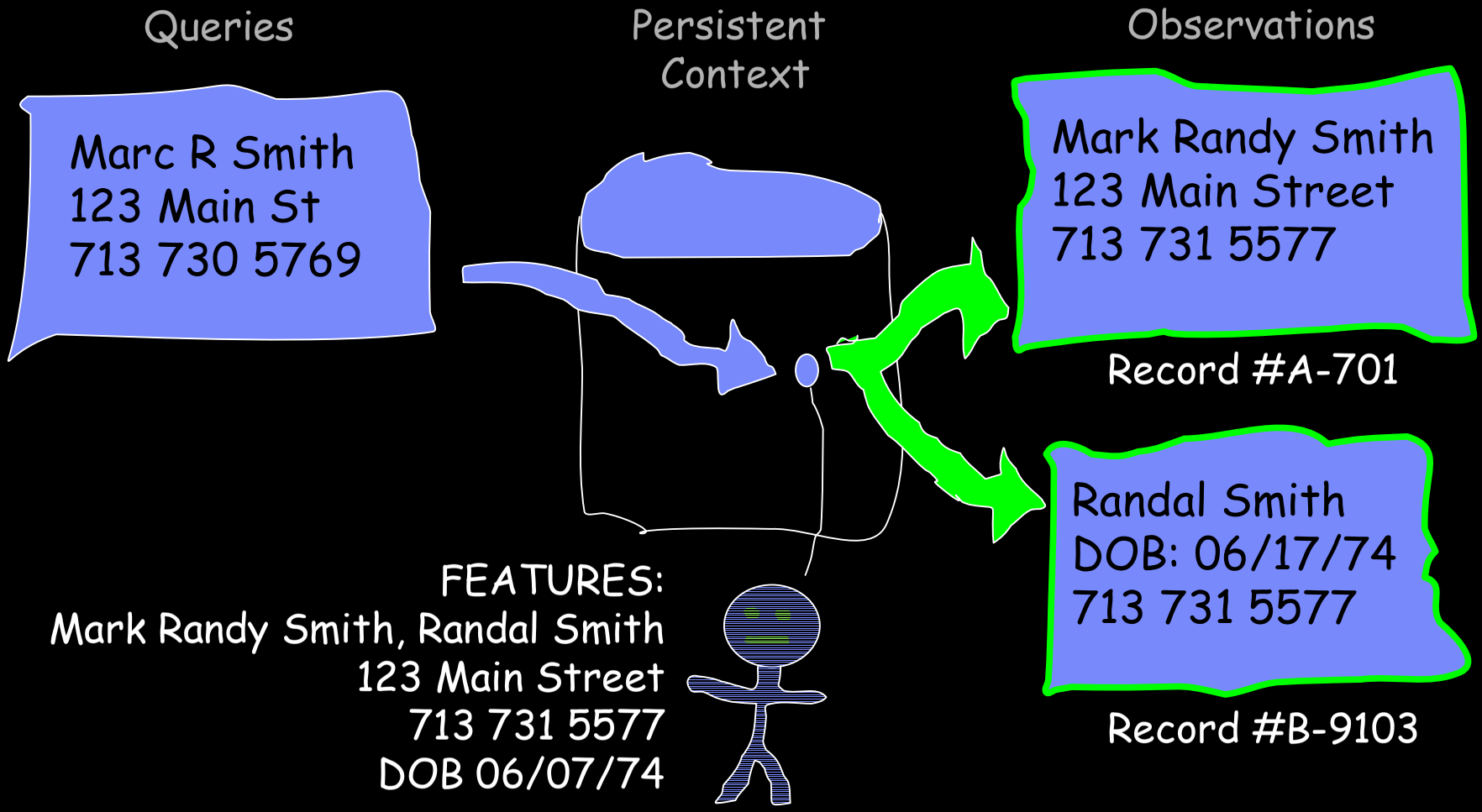
Now the Un-discoverable …



… After Context Reconstruction …



Enables a New Paradigm in Real-time Discovery



Additionally … All Data is First Treated as Query

Queries

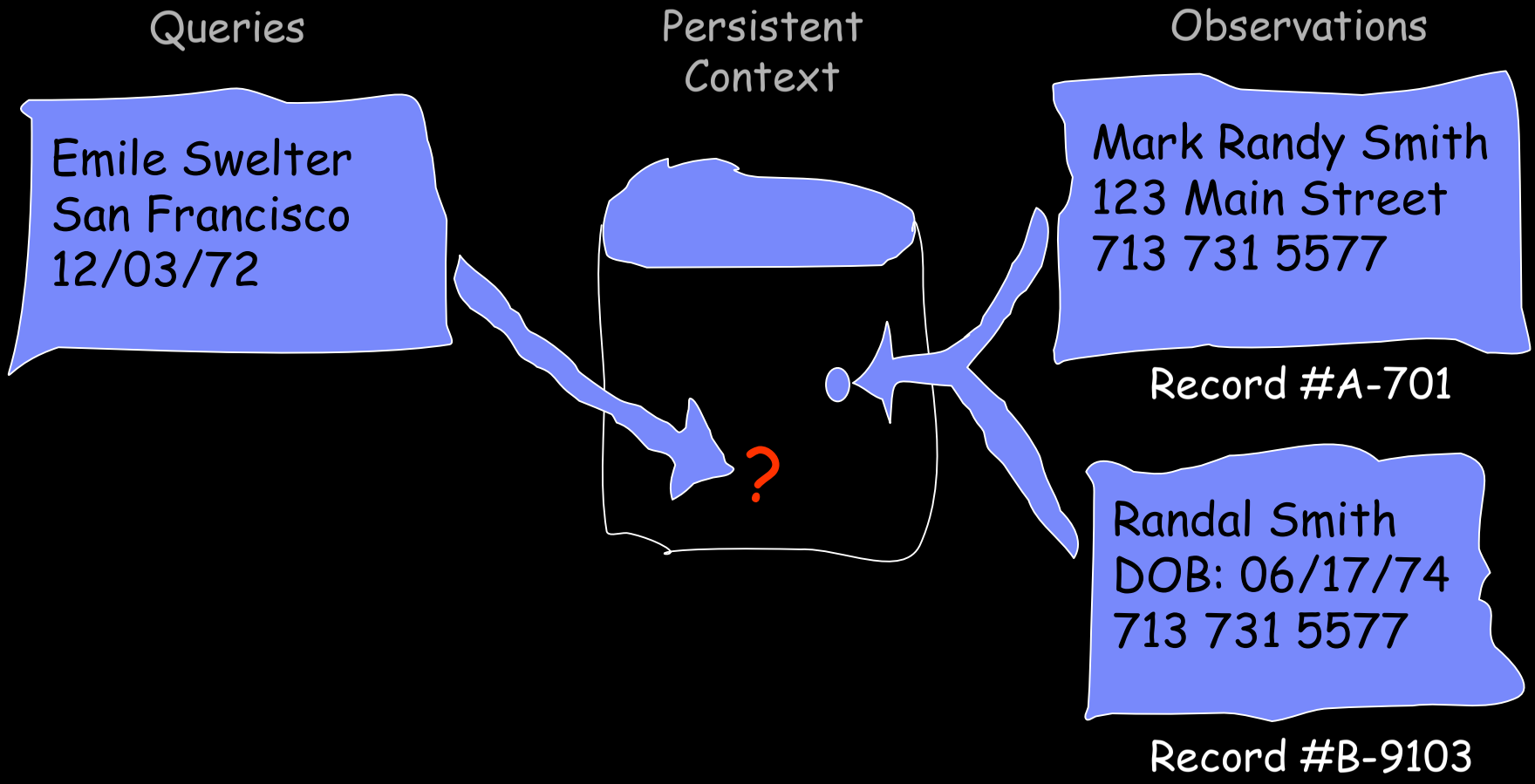
The query could be:

- A user with a question

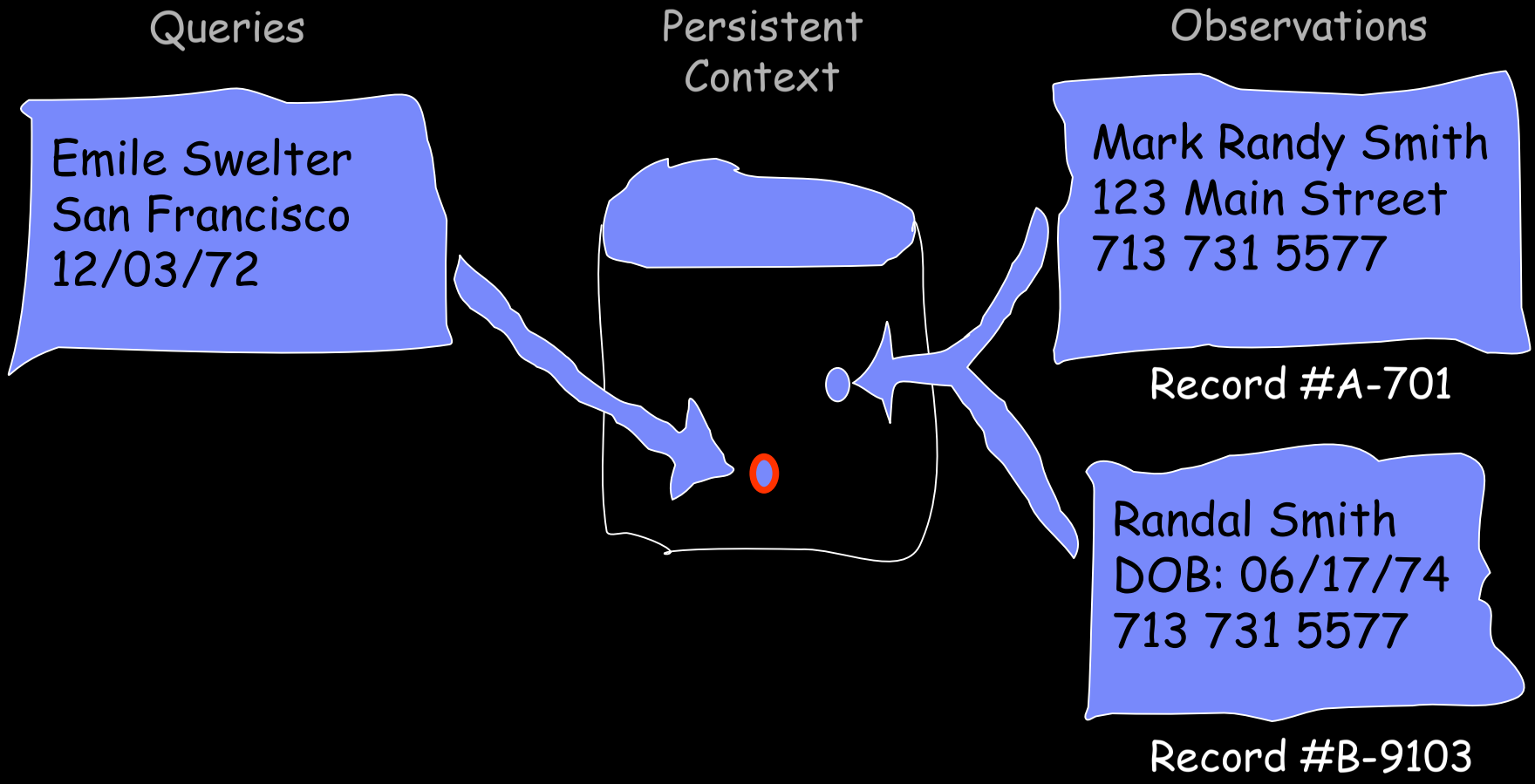
Or, also could be data:

- A new investigation
- A background check
- A new account
- An address change
- Deceased persons

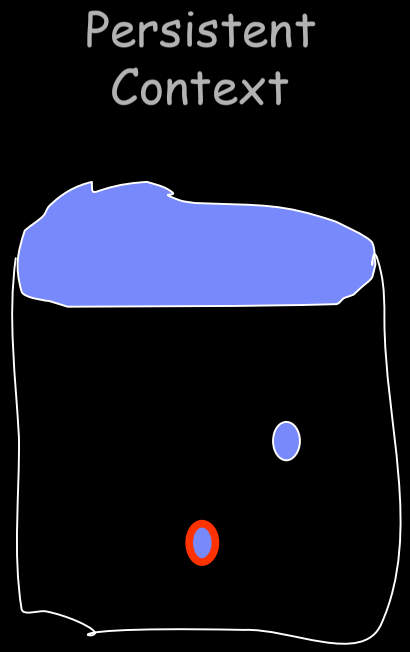
And … Any Query can be Treated as Data …



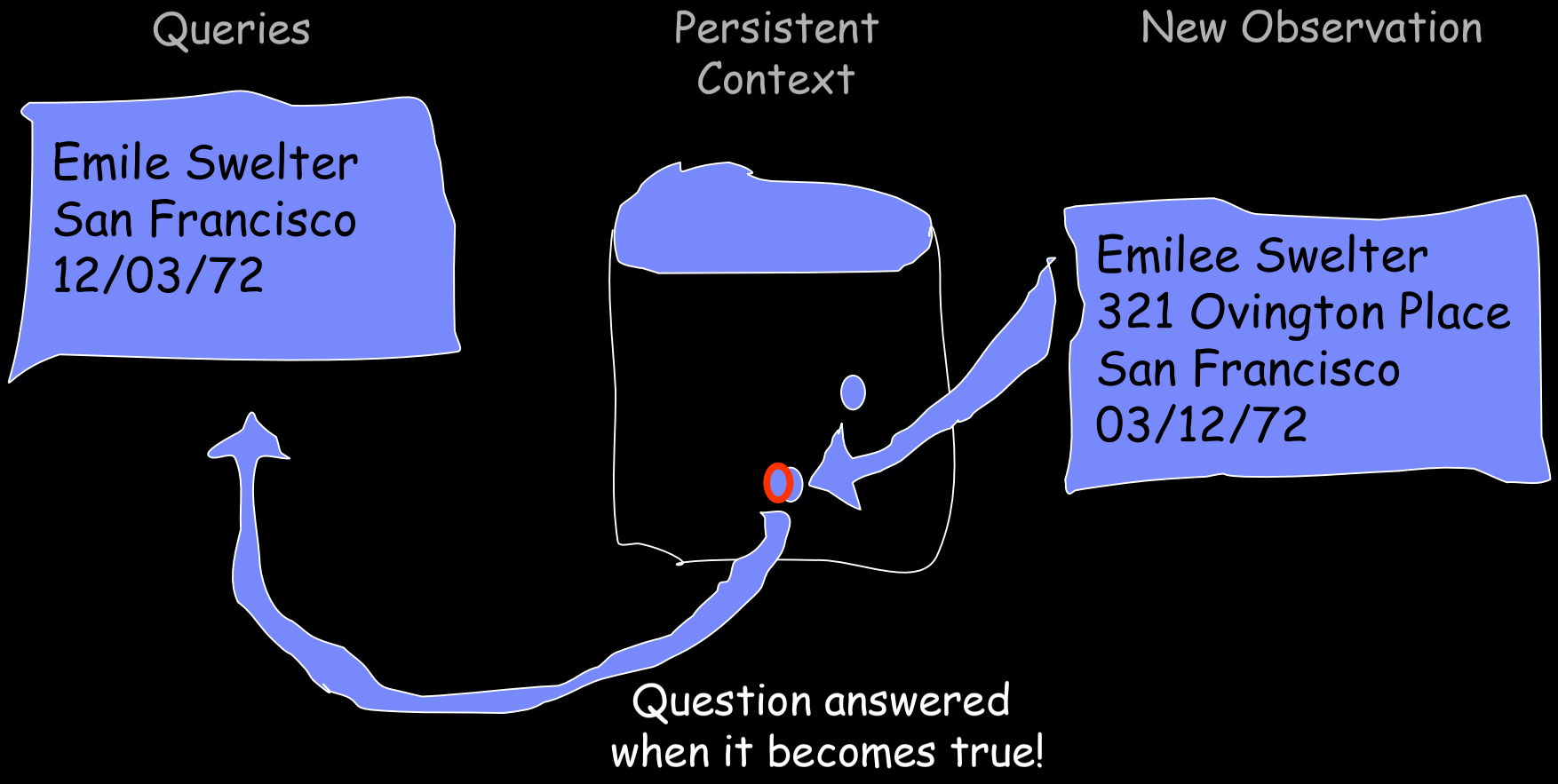
… In Which Case the Query can Stick (Persist)



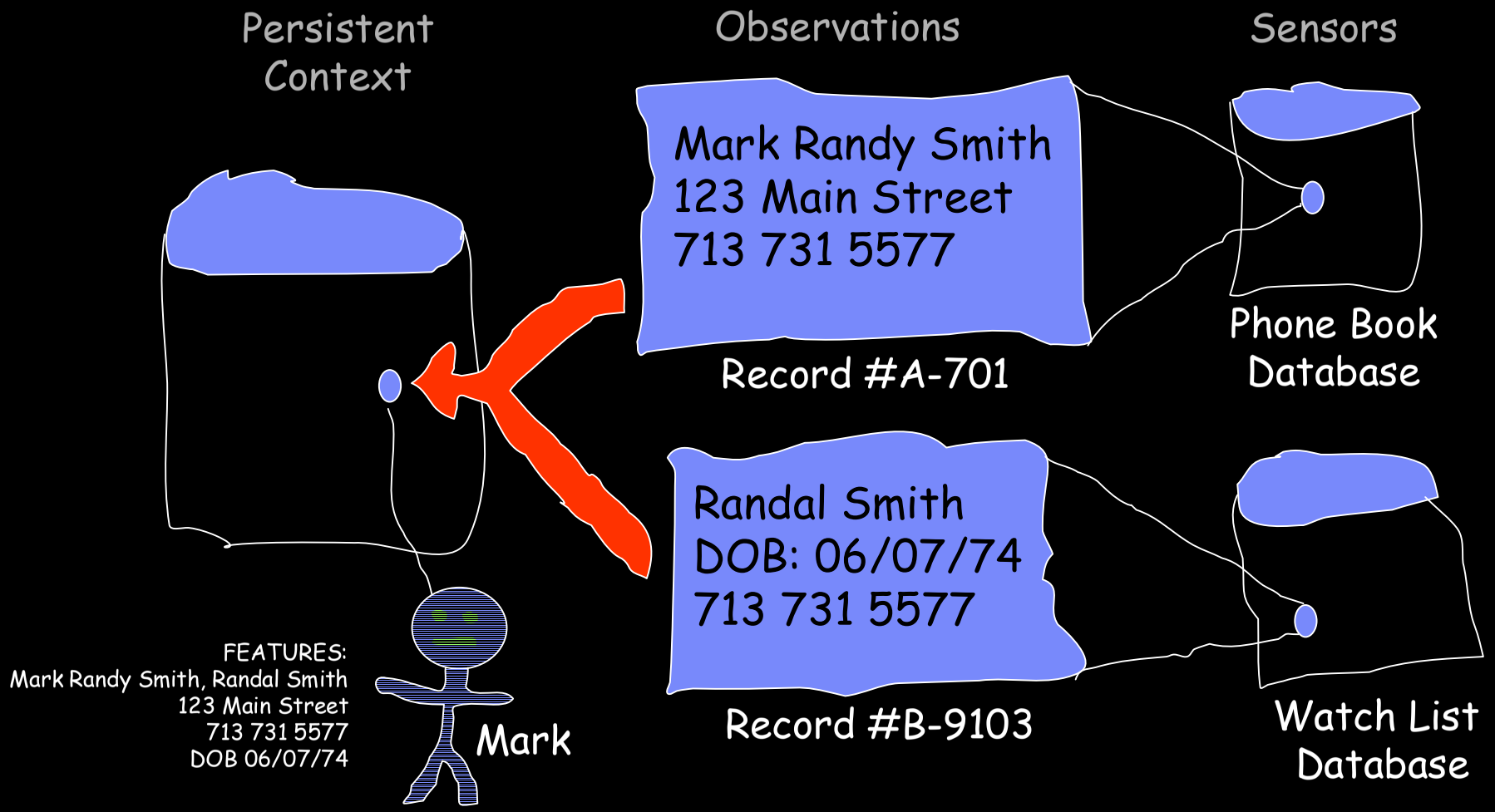
Notable, Stick in the Same Data Space



Now, New Observations Answer Persistent Queries



This is Identity Resolution



Semantically reconciled observations are necessary to understanding context.

Handling this in real-time, at scale and in a sustainable manner is the hard part!

Perpetual Analytics: The Game Changer

The "data finds the data" …

and relevance finds the consumer.

1st principal

If you do not process every arriving piece of data first like a query … then you will not know if you hold content that matters … until someone asks.

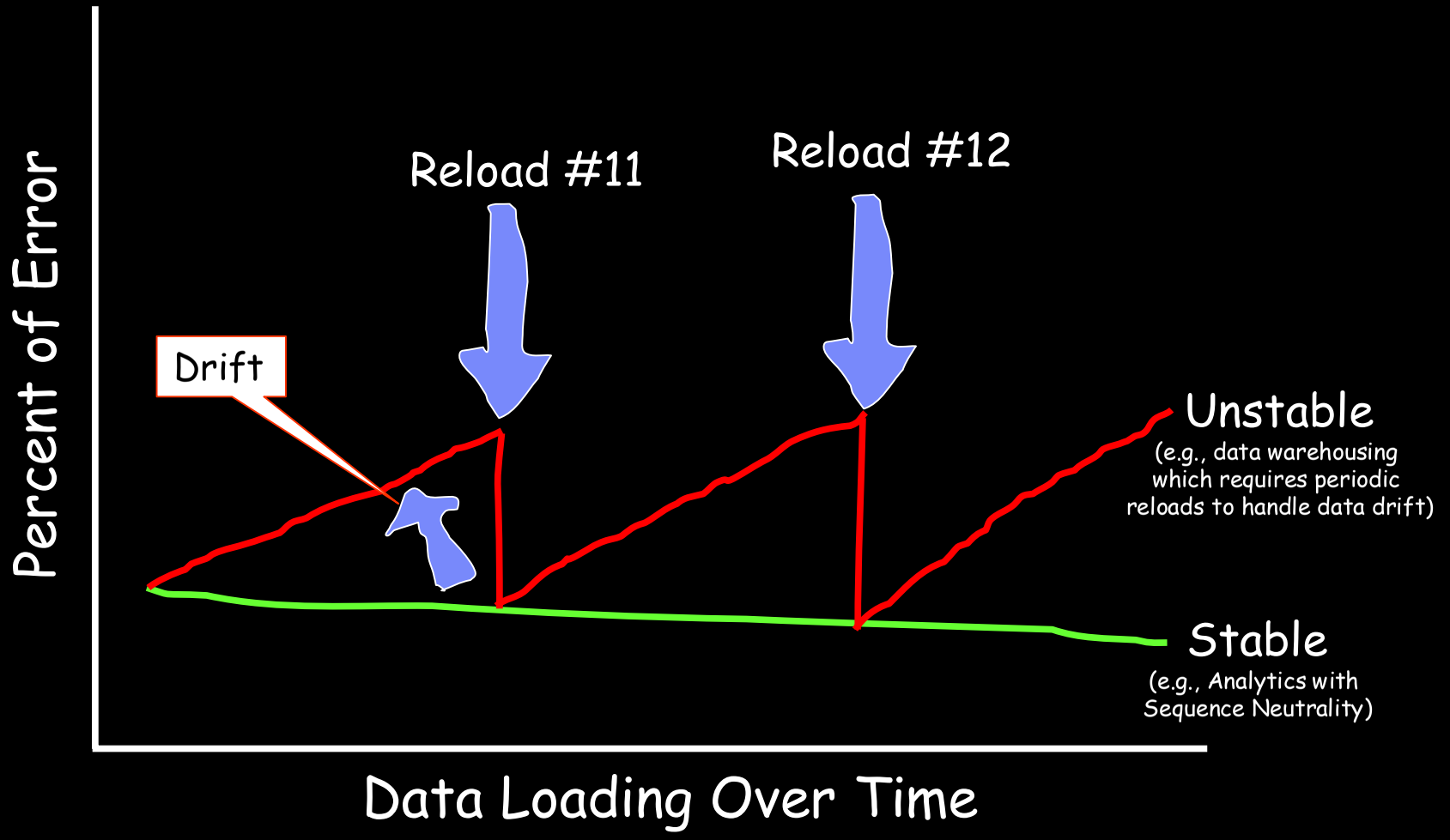
2nd principal

If you do not assemble and persist context on data streams … computational costs for after-the-fact assembly are unbearable.

Perpetual Analytics Requires …

- Persistent Context
  - Received data is reconciled to historical holdings and persisted (versus context-on-the-fly)
- Tethered to Source Systems
  - Processing of adds, changes and deletes from source systems
- Full Attribution
  - Every row must retain its pedigree (no data survivorship processing)
- Data and Query Equality
  - Processing new observations as a queries
  - Persisting queries as data (as selected and with expirations)
- Sequence Neutrality
  - New data corrects previous outcomes improving accuracy over time
  - The database end-state is the same despite the arrival order or timing of the data

Sequence Neutrality is Critical for Context Stability



Semantic Reconciliation: 23 Years of Practical Experience

- Deterministic with real-time and self-correcting probabilistic thresholding is essential to deliver the highest possible accuracy and scalability
- Sustainability requires:
  - No dependence on training data sets (initial or otherwise)
  - No merge and purge (data survivorship) processing
  - No in-memory full data set persistence
- Federated (i.e., context-on-the-fly) matching/linkage architectures cannot scale
- The greatest degree of discovery and intelligence comes from analytics on data streams (versus batch processing)
- The most complexity is caused by attempting Sequence Neutrality at scale

Currently Available Technology

- Context
  - Same identities (Identity Resolution)
  - Related identities (Relationship Resolution)
- Streaming context of
  - People
  - Organizations
  - Certain discernable objects (e.g., boats, planes, etc.)
- Scalability
  - 3B rows, 600M resolved entities, >2,000 contextualized observations per second
- Sustainability
  - Significant sequence neutrality processing at ingestion

Latest Advance: Anonymous Resolution

A new technique that allows "*n*" data holders to share anonymized identity-based observation data …

whereby context is assembled and persisted while the data remains in a cryptographic form …

resulting in a more secure and privacy-enhancing way to deliver multi-party, large-scale perpetual analytic systems

Analytics in the Anonymized Data Space – The Future!

"If information can be shared in an anonymized form whereby a materially similar result can be achieved …

why would an organization share information any other way?"

Blogging About All Of This At:

www.JeffJonas.TypePad.com

Information Management
Privacy
National Security

and Triathlons

The Next Big Leap

Integrated observations
which are … contextually reconciled
in … real-time
with … sequence neutral learning
where … data and hypotheses coexist
creating … persistent awareness
toward new levels of … active intelligence

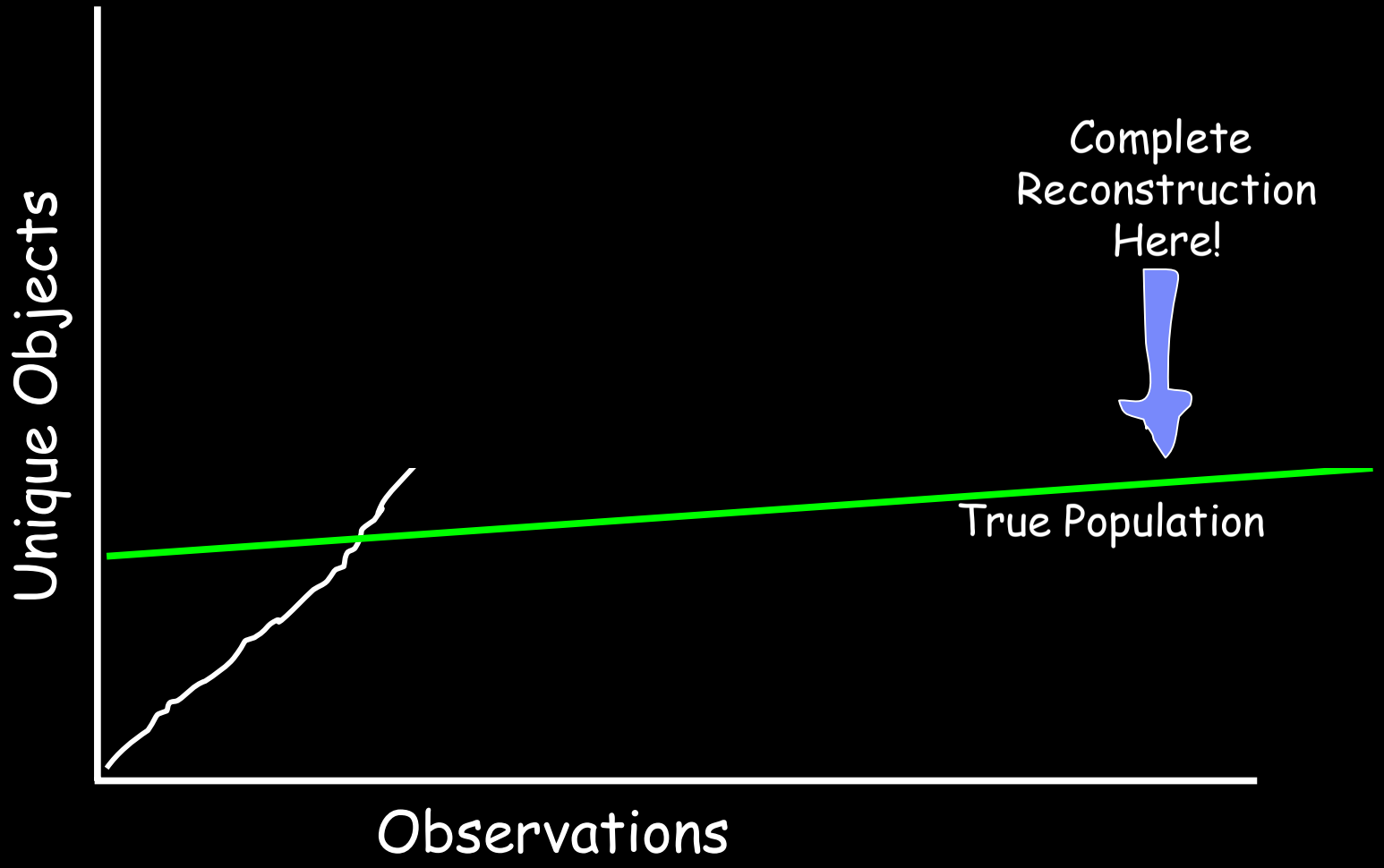
IBM Entity Analytic Solutions

Perpetual Analytics

Data finds the data … Relevance finds the user

Jeff Jonas, Chief Scientist, IBM Entity Analytics
Blogging at www.JeffJonas.TypePad.com

Saturated Observations = Complete Reconstruction



In Contrast: Human Contextualization

- Conscious – streaming contextualization
- Dreaming – deep re-contextualization
- Relevance of this notion:
  - We have a long way to go towards intelligence on streams before we must resort to off-line processing