REAL-TIME AI
FOR ENTITY RESOLUTION

Jeff Jonas
Founder and CEO
jeff@senzing.com
**Entity Resolution (ER)**

Recognizing when two observations relate to the same entity, despite having been described differently.

Conversely, recognizing when two observations do not relate to the same entity, despite having been described similarly.

And, the ability to remember relationships between entities.
NORA: Non-Obvious Relationship Awareness

- 15M+ customers
- 20k+ employees
- 18 watch lists

24 active players were known cheaters
23 players had relationships to prior arrests/incidents
192 employees had possible vendor relationships
7 employees were the vendor
NORA Grows Up

G2 Aspirations

- Real-time AI for entity resolution
  - No training, no tuning, no experts
  - Self-tuning; self-correcting

- Speed and scalability
  - Scale vertically and horizontally
  - 10’s billions of records
  - 1k’s of transactions a second

- New data sources, entity types, attributes on the fly

- Privacy by Design (PbD)
Early “G2” Success

Since 2012, G2 has helped us modernize U.S. voter registration while reducing unintended disclosure risk. The system requires less than one full-time person to manage over 350M records for 28 member states.

Shane Hamlin, Executive Director

Fake Identity Detection: 600k fake students detected; $300m saved

Maritime Domain Awareness: Malacca Straits monitoring for the Singapore Ministry of Defense
ENTITY RESOLUTION IN SLOW MOTION
Do these three entities become one, two or three resolved entities?

1. Robert Smith 123 Main Street DOB: 12/11/1978 smith@email.com 703.554.1214
2. Bob Smith 1515 Adela Lane DOB: 11/12/1978 1.703.554.1214 702.919.1600
3. Rob Smith 123 E Main St (703) 554-1214 ID: 00112233
4. 5. 6. 7. 8. 9.
Despite variations in name, address and phone, and the date of birth transposition, they become resolved entity E1.
New entity or belongs to the known entity E1?

<table>
<thead>
<tr>
<th></th>
<th>Name</th>
<th>Address</th>
<th>DOB</th>
<th>Email</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Robert Smith</td>
<td>123 Main Street</td>
<td>12/11/1978</td>
<td><a href="mailto:smith@email.com">smith@email.com</a></td>
<td>703.554.1214</td>
</tr>
<tr>
<td>2</td>
<td>Bob Smith</td>
<td>1515 Adela Lane</td>
<td>11/12/1978</td>
<td></td>
<td>1.703.554.1214</td>
</tr>
<tr>
<td>3</td>
<td>Rob Smith</td>
<td>123 E Main St</td>
<td></td>
<td></td>
<td>(703) 554-1214</td>
</tr>
<tr>
<td>4</td>
<td>Patricia Smith</td>
<td></td>
<td></td>
<td><a href="mailto:smith@email.com">smith@email.com</a></td>
<td>702.919.1600</td>
</tr>
</tbody>
</table>

E1

---

Number of entities: 4
When an email address is the same but the names are completely different, a new but related entity E2 is created.

Robert Smith
123 Main Street
DOB: 12/11/1978
smith@email.com
703.554.1214

Bob Smith
1515 Adela Lane
DOB: 11/12/1978
1.703.554.1214
702.919.1600

Rob Smith
123 E Main St
(703) 554-1214
ID: 00112233

Patricia Smith
smith@email.com

Discovered Relationships
How does this entity integrate with the existing entity graph?

Robert Smith
123 Main Street
DOB: 12/11/1978
smith@email.com
703.554.1214

Bob Smith
1515 Adela Lane
DOB: 11/12/1978
1.703.554.1214
702.919.1600

Rob Smith
123 E Main St
(703) 554-1214
ID: 00112233

Patricia Smith
smith@email.com
Entity E3 is created with a relationship to entity E1.

1. Robert Smith
   123 Main Street
   DOB: 12/11/1978
   smith@email.com
   703.554.1214

2. Bob Smith
   1515 Adela Lane
   DOB: 11/12/1978
   1.703.554.1214
   702.919.1600
   (703) 554-1214
   ID: 00112233

3. Kim Smith
   kim@email.com
   Spouse: 3

4. Patricia Smith
   smith@email.com

5. Disclosed Relationships

6. Senzing
### What becomes of this entity?

<table>
<thead>
<tr>
<th>Index</th>
<th>Name</th>
<th>Address</th>
<th>DOB</th>
<th>Email</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Robert Smith</td>
<td>123 Main Street</td>
<td>12/11/1978</td>
<td><a href="mailto:smith@email.com">smith@email.com</a></td>
<td>703.554.1214</td>
</tr>
<tr>
<td>2</td>
<td>Bob Smith</td>
<td>1515 Adela Lane</td>
<td>11/12/1978</td>
<td><a href="mailto:smith@email.com">smith@email.com</a></td>
<td>1.703.554.1214</td>
</tr>
<tr>
<td>3</td>
<td>Rob Smith</td>
<td>123 E Main St</td>
<td>12/11/1978</td>
<td><a href="mailto:smith@email.com">smith@email.com</a></td>
<td>702.919.1600</td>
</tr>
<tr>
<td>4</td>
<td>Patricia Smith</td>
<td></td>
<td></td>
<td><a href="mailto:smith@email.com">smith@email.com</a></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Kim Smith</td>
<td></td>
<td></td>
<td><a href="mailto:kim@email.com">kim@email.com</a></td>
<td>1.703.554.1214</td>
</tr>
<tr>
<td>6</td>
<td>Bob R. Smith</td>
<td></td>
<td></td>
<td><a href="mailto:bsmith@work.com">bsmith@work.com</a></td>
<td>702.919.1600</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Bob Smith**
  - AKA Bobby Jones
  - DOB: 12/11/1978
  - Email: bsmith@work.com

**Relationships:**
- E1: Parent/Child
- E2: Parent/Child
- E3: Parent/Child
Many people have the same name and date of birth, thus entity E4 is created as a possible match.
What becomes of this entity?

1. Robert Smith  
   123 Main Street  
   DOB: 12/11/1978  
   smith@email.com  
   703.554.1214

2. Bob Smith  
   1515 Adela Lane  
   DOB: 11/12/1978  
   1.703.554.1214  
   702.919.1600  
   (703) 554-1214  
   ID: 00112233

3. Rob Smith  
   123 E Main St  
   DOB: 11/12/1978  
   ID: 00112233

4. Patricia Smith  
   smith@email.com

5. Kim Smith  
   kim@email.com  
   Spouse:  
   3

6. Bob R. Smith  
   AKA Bobby Jones  
   DOB: 12/11/1978  
   bsmith@work.com

7. smith@email.com

8. 

9. 

E1  
E2  
E3  
E4  
E5
If E2 did not exist this would resolve to E1, and vice versa. E5 is created as it’s a possible match to both E1 and E2.
What becomes of this entity?

1. Robert Smith
   123 Main Street
   DOB: 12/11/1978
   smith@email.com
   703.554.1214

2. Bob Smith
   1515 Adela Lane
   DOB: 11/12/1978
   1.703.554.1214
   702.919.1600

3. Rob Smith
   123 E Main St
   (703) 554-1214
   ID: 00112233

4. Patricia Smith
   smith@email.com

5. Kim Smith
   kimm@email.com
   Spouse:

6. B. Smith
   1515 Adela Lane
   bsmith@work.com
   1.702.919.1600

7. smith@email.com

8. Bob R. Smith
   AKA Bobby Jones
   DOB: 12/11/1978
   bsmith@work.com

9. Patricia Smith
   smith@email.com

Senzing
This reveals that entities E1 and E4 are the same, causing the entity E4 to conjoin with entity E1.
What becomes of this entity?

1. Robert Smith
   123 Main Street
   DOB: 12/11/1978
   smith@email.com
   703.554.1214

2. Bob Smith
   1515 Adela Lane
   DOB: 11/12/1978
   1.703.554.1214
   702.919.1600

3. Rob Smith
   123 E Main St
   (703) 554-1214
   ID: 00112233

4. Patricia Smith
   smith@email.com

5. Kim Smith
   ksmith@email.com
   Spouse: 3

6. Bob R. Smith
   AKA Bobby Jones
   DOB: 12/11/1978
   bsmith@work.com

7. B. Smith
   1515 Adela Lane
   bsmith@work.com
   1.702.919.1600

8. Rob Smith Sr.
   DOB: 3/31/1954
   ID: 00112233
This is evidence that record 3 is not entity E1, causing the “Sr.” records to become the new entity E6.
What should happen to this entity?

Robert Smith  
123 Main Street  
DOB: 12/11/1978  
smith@email.com  
703.554.1214

Patricia Smith  
smith@email.com

Bob Smith  
1515 Adela Lane  
DOB: 11/12/1978  
1.703.554.1214  
702.919.1600

Kim Smith  
kim@email.com  
Spouse:  
3

Kim Smith  
kim@email.com  
Spouse:  
3

Bob Smith  
1515 Adela Lane  
DOB: 11/12/1978  
1.703.554.1214  
702.919.1600

Patricia Smith  
smith@email.com

Rob Smith  
123 E Main St  
(703) 554-1214  
ID: 00112233

Bob R. Smith  
AKA Bobby Jones  
DOB: 12/11/1978  
bsmith@work.com

B. Smith  
1515 Adela Lane  
bsmith@work.com  
1.702.919.1600

Rob Smith Sr.  
DOB: 3/31/1954  
ID: 00112233

Bob Jones  
123 Main Street  
702.919.1600  
bjones@email.com

Kim Smith  
kim@email.com  
Spouse:  
3

Rob Smith Sr.  
DOB: 3/31/1954  
ID: 00112233
This observation resolves to E1 because E1 contains a sufficient number of matching values.
What will happen to the entity graph if you delete (forget) record 8?

1. Robert Smith
   123 Main Street
   DOB: 12/11/1978
   smith@email.com
   703.554.1214

2. Bob Smith
   1515 Adela Lane
   DOB: 11/12/1978
   1.703.554.1214
   702.919.1600

3. Rob Smith
   123 E Main St
   (703) 554-1214
   ID: 00112233

6. Bob R. Smith
   AKA Bobby Jones
   DOB: 12/11/1978
   bsmith@work.com

8. B. Smith
   1515 Adela Lane
   bsmith@work.com
   1.702.919.1600

9. Rob Smith Sr.
   DOB: 3/31/1954
   ID: 00112233

10. Bob Jones
    123 Main Street
    702.919.1600
    bjones@email.com

The orange-colored box indicates that record 8 has been deleted (marked with an 'X').
As if record 8 never existed, entity E1 is re-evaluated – kicking out records 6 and 10 into entities E7 and E8

Robert Smith
123 Main Street
DOB: 12/11/1978
smith@email.com
703.554.1214

Bob Smith
1515 Adela Lane
DOB: 11/12/1978
1.703.554.1214
702.919.1600

Rob Smith
123 E Main St
(703) 554-1214
ID: 00112233

Bob R. Smith
AKA Bobby Jones
DOB: 12/11/1978
bsmith@work.com

B. Smith
1515 Adela Lane
bsmith@work.com
1.702.15.1600

Rob Smith Sr.
DOB: 3/31/1954
ID: 00112233

Self-correcting
(“Data Tethering”)
Common Use Cases

**Bad Guy Hunting**
Financial Fraud, Insider Threat, Watchlist Screening, Criminal Acts, Fake Identities

**Marketing 360**
Omni-channel Marketing, Next Best Action, List De-duplication

**Privacy Compliance**
Single Subject Search, Right to be Forgotten Monitoring, Central Disclosure and Consent Tracking

**Risk Analysis**
Credit Risk, Continuous Vetting, Brand Protection, Maritime Domain Awareness

**Public Safety**
Investigations, Humanitarian Assistance, School Safety

**Other**
Auto-labeling for Machine Learning
Meet Senzing

Reincarnated
2016 one-of-a-kind IBM spinout of G2 technology, team and intellectual property

Big Idea
Entity resolution made easy for programmers

Out of Stealth
June 2018

Product
You download Senzing entity resolution software (not SaaS)
THE FIRST REAL-TIME AI FOR ENTITY RESOLUTION

No email address required. No private data flows to Senzing.

FREE DOWNLOAD

DOWNLOAD NOW

We ship you software, you deploy onsite or on your cloud. No private data flows to Senzing.

System Recommendations
4 cores, 16GB RAM, 250GB flash storage (SSD or NVMe)

Minimum System Requirements
2 cores, 8GB RAM, 100GB storage

FREE DOWNLOAD

Windows 7/10 (64-bit) (1.3GB) macOS (1.3GB)

Up to 10k records – free email support!
APIs: C, Java, Python

ER Library
C++, SQL
Linux
Windows, macOS
White Paper: Uniquely Senzing

1. Purpose-built AI for Entity Resolution
2. Real-time Operations
3. Minimal Data Preparation
4. Built-In Privacy by Design (PbD)
5. Nonobvious Relationship Awareness
6. Speed and Scalability
jeff@senzing.com

@jeffjonas | @senzing
REAL-TIME AI
FOR ENTITY RESOLUTION