Big Data. Big Responsibility.

Jeff Jonas
Founder and CEO
jeff@senzing.com
A Surveillance Society is:

... inevitable

... irreversible

... irresistible!
Easy way to stay connected with family and friends
More convenient than owning a car
Easiest way to build & manage my professional network
Slippery Slope
Moved to Dropbox
Framed Items ... Digitized
Memorabilia ... Digitized
60 Ironman Medals Digitized
Terror Hits Pentagon, World Trade Center
Favorite Rock ... Digitized
Moving into the Cloud
Clothes Closet in the Cloud
As of August 23rd, 2018

The net weight of my personal possessions <200lbs
My only remaining keepsakes
What rights do I have re: my digital stuff?

Utterly unencumbered.
Digitization of mankind’s activity has barely started

Next: greater data diversity & distribution
Big Data. Big Responsibility.
Jeff Jonas

Data Scientist, Serial Entrepreneur

- 1983: Founded Systems Research & Development
- 1991: Relocated to Las Vegas

Non-Obvious Relationship Awareness
Jeff Jonas  
*Data Scientist, Serial Entrepreneur*

- 1983: Founded Systems Research & Development
- 1991: Relocated to Las Vegas
- 2001: Funded by In-Q-Tel
- 2005: SRD acquired by IBM
  - Chief Scientist of Context Computing
  - IBM Fellow
- 2016: Founded Senzing, an IBM spin-out based on my “G2” Entity Resolution technology
Presentation Outline

1. Introduction to entity resolution

2. Critical role entity resolution plays in privacy compliance
Entity Resolution (ER)

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

Robert Smith
123 Main Street
703.554.1214

Rob E. Smith
123 E Main St
+1 (703) 554-1214
Entity Resolution (ER)

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

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

Robert Smith Jr.
123 Main Street
703.554.1214

Rob E. Smith Sr.
123 E Main St
+1 (703) 554-1214
Entity Resolution Examples

- Duplicates in your personal address book
- Duplicates across your company’s prospect and customer database
- Previously terminated accounts now back masquerading as good accounts
- Watch-listed parties evading detection due to a slight name variation
Entity Resolution is fundamental to intelligence

... machine intelligence, human intelligence, or otherwise.
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.

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
New entity or belongs to the known 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

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

4. Patricia Smith
   smith@email.com

5.

6.

7.

8.

9.
When an email address is the same but the names are completely different, a new but related entity E2 is created.
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

Kim Smith
kim@email.com
Spouse: 3

Senzing
Entity E3 is created with a relationship to entity E1.

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

Kim Smith
smith@email.com
Spouse: Patricia Smith

Patricia Smith
smith@email.com

Disclosed Relationships
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
   kim@email.com
   Spouse: 3

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

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

Senzing
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.

<table>
<thead>
<tr>
<th></th>
<th>Name</th>
<th>Address</th>
<th>DOB</th>
<th>Phone</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Robert Smith</td>
<td>123 Main Street</td>
<td>12/11/1978</td>
<td>703.554.1214</td>
<td><a href="mailto:smith@email.com">smith@email.com</a></td>
</tr>
<tr>
<td>2</td>
<td>Bob Smith</td>
<td>1515 Adela Lane</td>
<td>11/12/1978</td>
<td>1.703.554.1214</td>
<td><a href="mailto:smith@email.com">smith@email.com</a></td>
</tr>
<tr>
<td>3</td>
<td>Rob Smith</td>
<td>123 E Main St</td>
<td>11/12/1978</td>
<td>(703) 554-1214</td>
<td><a href="mailto:bsmith@work.com">bsmith@work.com</a></td>
</tr>
<tr>
<td>4</td>
<td>Patricia Smith</td>
<td></td>
<td></td>
<td></td>
<td><a href="mailto:smith@email.com">smith@email.com</a></td>
</tr>
<tr>
<td>5</td>
<td>Kim Smith</td>
<td></td>
<td></td>
<td></td>
<td><a href="mailto:smith@email.com">smith@email.com</a></td>
</tr>
<tr>
<td>6</td>
<td>Bob R. Smith</td>
<td>123 E Main St</td>
<td>12/11/1978</td>
<td></td>
<td><a href="mailto:bsmith@work.com">bsmith@work.com</a></td>
</tr>
</tbody>
</table>

Possible Match ("Ambiguous")
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
   kim@email.com
   Spouse: 3

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

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

ID: 00112233

Patricia Smith
smith@email.com

Kim Smith
kim@email.com
Spouse: 3

B. Smith
1515 Adela Lane
bsmith@work.com
1.702.919.1600
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
   kim@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.
Data Drift vs. “Sequence Neutrality”

Percent of Error

Data Loading Over Time

- **Unstable**
  - Reloading required to correct data drift

- **Stable**
  - with Sequence Neutrality

Reload #1

Reload #2

new observations reversing earlier assertions

data drift
What should happen to 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
   kim@email.com
   Spouse: 3

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

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

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

9. Bob Jones
   123 Main Street
   702.919.1600
   bjones@email.com
This observation resolves to E1 because E1 contains a sufficient number of matching values.

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

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

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

7. “Entity-centric Learning”

8. Rob Smith Sr.
   DOB: 3/31/1954
   ID: 00112233
## “Channel Separation” – the Primary Deception Tradecraft

<table>
<thead>
<tr>
<th>Money Launderer</th>
<th>KYC Onboarding</th>
<th>Open Account</th>
<th>Employment Application</th>
<th>External Counterparty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob Jones</td>
<td>Bob Smith</td>
<td>Robert Smith</td>
<td>Bob R. Smith</td>
<td>Bobby Randal Smith</td>
</tr>
<tr>
<td>123 Main Street</td>
<td>1515 Adela Lane</td>
<td>123 Main Street</td>
<td>AKA Bobby Jones</td>
<td>1515 Adela Lane</td>
</tr>
<tr>
<td>702.919.1600</td>
<td>1703.554.1214</td>
<td>702.919.1600</td>
<td>703.554.1214</td>
<td>3232100123991123</td>
</tr>
<tr>
<td><a href="mailto:bjones@email.com">bjones@email.com</a></td>
<td>1703.554.1214</td>
<td>702.919.1600</td>
<td><a href="mailto:bsmith@work.com">bsmith@work.com</a></td>
<td></td>
</tr>
</tbody>
</table>
Detection Requires “Channel Consolidation”

Entity Resolution
Entity-centric Learning

Money Launderer
- Bob Jones
  - 123 Main Street
  - 702.919.1600
  - bjones@email.com

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

Open Account
- Robert Smith
  - 123 Main Street
  - DOB: 12/11/1978
  - 702.919.1600

Employment Application
- Bob R. Smith
  - AKA Bobby Jones
  - 1515 Adela Lane
  - DOB: 12/11/1978
  - bsmith@work.com

External Counterparty
- Bobby Randal Smith
  - 1515 Adela Lane
  - 3232100123991123
  - 702.919.1600
  - 703.554.1214
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

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

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

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

7. Bob Jones
   - 123 Main Street
   - 702.919.1600
   - bjonnes@email.com
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
1.702.919.1600

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

Self-correcting
("Data Tethering")
THE CRITICAL ROLE OF ENTITY RESOLUTION IN PRIVACY COMPLIANCE
GDPR and CCPA obligate organizations to provide data subjects with access to their personal data.
Compliance starts with one simple question:

“What do we know about the data subject?”
Privacy compliance starts with one simple question: “What do we know about the data subject?”

Unfortunately, this is an exceptionally hard question for organizations to answer.
Challenge #1: Lots of Places To Look

Will you remember to search every database you have?

CUSTOMERS
LOYALTY CLUB
MARKETING
CUSTOMER SERVICE
PAYROLL

OTHER COPIES
Reservoirs, Warehouses, Lakes, Data Marts, Email Systems, Spreadsheets, etc.

Fraud
Customer Engagement
Reporting

Search
Liz Reston
reston@home.com
Challenge #2: Data Variations

How long will it take to try all the different variations?

CUSTOMERS
Liz Reston
reston@home.com
123 East Court Rd

Elizabeth Reston
beth@work.com

LOYALTY CLUB
Liz Reston
beth@work.com
20 1111 2222

MARKETING
Beth Smith-Reston
123 East Court Rd
liz@old-email.com

CUSTOMER SERVICE
Beth Smith-Reston
20 1111 2222

PAYROLL
Elizabeth Smith
123 East Court Rd

Search
Liz Reston
reston@home.com
Challenge #3: Existing Systems Have Search Limitations

Do you need technical expertise for each search?

- CUSTOMERS
- LOYALTY CLUB
- MARKETING
- CUSTOMER SERVICE
- PAYROLL

Employee# :____:
Name:___________:
DOB :__/__/__:

Search
Liz Reston
reston@home.com

Elizabeth Smith
123 East Court Rd
reston@home.com
Challenge #4: Not All Things That Look Alike Are Alike

How often will you send information to the wrong person? (creating your own data breach)

**CUSTOMERS**
- Liz Reston
  - reston@home.com
  - 123 East Court Rd
  - Bob Reston
    - beth@work.com
    - reston@home.com

**LOYALTY CLUB**
- Liz Reston
  - beth@work.com
  - 20 1111 2222

**MARKETING**
- Beth Smith-Reston
  - beth@work.com
  - 20 1111 2222
  - liz@old-email.com

**CUSTOMER SERVICE**
- Beth Smith-Reston
  - 123 East Court Rd
  - liz@old-email.com
  - 20 1111 2222

**PAYROLL**
- Elizabeth Smith
  - 123 East Court Rd
  - reston@home.com

---

Search

Liz Reston
reston@home.com
Individual System Search

Challenge #1: Lots of Places To Look
Will you remember to search every database you have?

Challenge #2: Data Variations
How much time will it take to look at different variations?

Unreliable!
Individual System Search

Is like a library without an index ...

Thus having to look through every shelf in every aisle ...
The Missing Link

Data Access Request

Liz Reston
reston@home.com

Your Privacy Processes

The Missing Link:
Single Subject Search
The Missing Link

- Over 1,000 organizations surveyed
- Large companies, average:
  - 250 inquiries per month
  - 43 databases to search
  - 7 minutes per search
  - Requiring 7.5 full time people
Single Subject Search

- **CUSTOMERS**
- **LOYALTY CLUB**
- **MARKETING**
- **CUSTOMER SERVICE**
- **PAYROLL**

Data Access Request

Search
Liz Reston
reston@home.com
### Search Results

<table>
<thead>
<tr>
<th><strong>Search</strong></th>
<th><strong>Matches</strong></th>
<th><strong>Possible Matches</strong></th>
<th><strong>Possibly Related</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Liz Reston</td>
<td>Customers</td>
<td>#23456, #75688</td>
<td>Elizabeth Reston</td>
</tr>
<tr>
<td></td>
<td>Loyalty Club</td>
<td>#34-902456</td>
<td>Beth Ann Reston</td>
</tr>
<tr>
<td></td>
<td>Marketing</td>
<td>#1234</td>
<td>Liz Restn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Customers</td>
<td>#334311</td>
<td>Lizzy A. Reston</td>
</tr>
<tr>
<td></td>
<td>Loyalty Club</td>
<td>#34-908912</td>
<td>Bob Reston</td>
</tr>
</tbody>
</table>

**Like a Search Engine:**
These pointers direct you to the source information.
DEMO

- SINGLE SUBJECT SEARCH
Singapore Restaurants

Add Data Source
Choose a file from your file system to map and load.

Restaurant File 1
8,378 records
Singapore_File_1.csv
Nov 10, 2017, 7:47:06 AM

Restaurant File 2
21,788 records
Singapore_File_2.csv
Nov 10, 2017, 7:46:16 AM

Restaurant File 3
3,597 records
Singapore_File_3.csv
Nov 10, 2017, 7:49:04 AM
<table>
<thead>
<tr>
<th>Entity ID</th>
<th>ECL</th>
<th>Data Source</th>
<th>Name Data</th>
<th>Address Data</th>
<th>Phone Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>RESTAURANT FILE 1</td>
<td>Ocean Kingdom Seafood Restaurant</td>
<td>BUSINESS: 382 Joo Chiat Road, Singapore, Singapore</td>
<td>+65 6342 0382</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>RESTAURANT FILE 1</td>
<td>Queens Tandoor</td>
<td>BUSINESS: 11 Unity St, Singapore, 237955, Singapore</td>
<td>6562355803</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>RESTAURANT FILE 3</td>
<td>Queen's Tandoor</td>
<td>BUSINESS: Robertson Walk, 11 Unity St, #01-18/29, Singapore 237955</td>
<td>6562353005</td>
</tr>
<tr>
<td>20</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Pizza Hut</td>
<td>BUSINESS: 1 Jurong West Central 2, Singapore, 648886, Singapore</td>
<td>6567907146</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>RESTAURANT FILE 1</td>
<td>Pizza Hut</td>
<td>BUSINESS: 2 Jurong East Central 1, Singapore, 648866, Singapore</td>
<td>6562353535</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>RESTAURANT FILE 3</td>
<td>Pizza Hut</td>
<td>BUSINESS: 1 Jurong West Central 2, #01-33 Jurong Point Shopping Centre, Singapore</td>
<td>+65 6790 7146</td>
</tr>
<tr>
<td>26</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Mr Punch Public House</td>
<td>BUSINESS: 26 Seah Street, Singapore, 188382, Singapore</td>
<td>658398966</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>RESTAURANT FILE 3</td>
<td>Mr Punch Public House</td>
<td>BUSINESS: 26 Seah Street, Singapore 188382</td>
<td>+65 6339 6266</td>
</tr>
<tr>
<td>28</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Madame Patisserie</td>
<td>BUSINESS: 01-76, 76 Boat Quay, Singapore, 65, Singapore</td>
<td>6565367028</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>RESTAURANT FILE 3</td>
<td>Madame Patisserie</td>
<td>BUSINESS: 76 Boat Quay, #01-76, Singapore 049864</td>
<td>+65 65367028</td>
</tr>
<tr>
<td>37</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>La Cantina in Venezia</td>
<td>BUSINESS: Changi Village Hotel, 1 Netheravon Road, Singapore, 508502, Singapore</td>
<td>6565469190</td>
</tr>
<tr>
<td>37</td>
<td></td>
<td>RESTAURANT FILE 3</td>
<td>La Cantina in Venezia</td>
<td>BUSINESS: Changi Village Hotel, 1 Netheravon Road, #08/09-02, Singapore 508502</td>
<td>+65 6546 919</td>
</tr>
<tr>
<td>40</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Chef's Table by Chef Stephan Zoisl</td>
<td>BUSINESS: 61 Tras Street, Singapore, 079000</td>
<td>6662244188</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>RESTAURANT FILE 3</td>
<td>Chef's Table by Chef Stephan</td>
<td>BUSINESS: 61 Tras Street Singapore 079000</td>
<td>6224 4188</td>
</tr>
<tr>
<td>42</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Brij Bhoomi</td>
<td>BUSINESS: 1088 Serangoon Road, Singapore, 328188, Singapore</td>
<td>6563923482</td>
</tr>
<tr>
<td>42</td>
<td></td>
<td>RESTAURANT FILE 3</td>
<td>Brij Bhoomi(3 Branches)</td>
<td>BUSINESS: 1088 Serangoon Rd, Singapore 328188</td>
<td>63923482</td>
</tr>
<tr>
<td>43</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Jacks Place Restaurant</td>
<td>BUSINESS: 900 S Woodlands Dr, Singapore, 730900, Singapore</td>
<td>6563634803</td>
</tr>
</tbody>
</table>
## Singapore Restaurants

### Matches: Restaurant File 1 to Restaurant File 3

<table>
<thead>
<tr>
<th>Entity ID</th>
<th>ECL</th>
<th>Data Source</th>
<th>Name Data</th>
<th>Address Data</th>
<th>Phone Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>236</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Oliva</td>
<td>BUSINESS: 34 Arab Street, Singapore, Singapore</td>
<td></td>
</tr>
<tr>
<td>243</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>The Drinking Market</td>
<td>BUSINESS: 226 East Coast Road, Singapore, 428923, Singapore</td>
<td>6564408939</td>
</tr>
<tr>
<td>245</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Shimbashi Soba</td>
<td>BUSINESS: 23 Serangoon Central, Singapore, 556083, Singapore</td>
<td>6566344786</td>
</tr>
<tr>
<td>249</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Leong Kee Bak Kut Teh</td>
<td>BUSINESS: 321 Bch Rd, Singapore, 199557, Singapore</td>
<td>6593801718</td>
</tr>
<tr>
<td>255</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Burgers vs Wings + Bar</td>
<td>BUSINESS: Orchard Central, 181 Orchard Road, Singapore, 238896</td>
<td>6566340423</td>
</tr>
<tr>
<td>256</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Artichoke</td>
<td>BUSINESS: 161 Middle Road (Sculpture Square), Singapore, 188978</td>
<td>6563366949</td>
</tr>
<tr>
<td>258</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>The Rice Table</td>
<td>BUSINESS: International Building 360 Orchard Rd 02-0910, Singapore, 238869, Singapore</td>
<td>6568353783</td>
</tr>
<tr>
<td>265</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Tai Dong Teochow Braised Duck Rice</td>
<td>BUSINESS: 49 Sims Place 01-04, Singapore, Singapore, 380049, Singapore, Singapore</td>
<td>6569476244</td>
</tr>
<tr>
<td>269</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Patisserie</td>
<td>BUSINESS: Sims Vista 49 Market And Food Centre, 49 Sims Place, 01-04, Singapore</td>
<td>6563387578</td>
</tr>
<tr>
<td>Entity ID</td>
<td>ECL</td>
<td>Data Source</td>
<td>Name Data</td>
<td>Address Data</td>
<td>Phone Data</td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>236</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Oliva</td>
<td>BUSINESS: 34 Arab Street, Singapore, Singapore</td>
<td>+65 6291 2297</td>
</tr>
<tr>
<td>243</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>The Drinking Market</td>
<td>BUSINESS: 226 East Coast Road, Singapore, Singapore</td>
<td>+65 6440 8939</td>
</tr>
<tr>
<td>245</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Shimbashi Soba</td>
<td>BUSINESS: 23 Serangoon Central, Singapore, Singapore</td>
<td>+65 6563 4786</td>
</tr>
<tr>
<td>249</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Leong Kee Bak Kut Teh</td>
<td>BUSINESS: 321 Beach Rd, Singapore, Singapore</td>
<td>+65 3801 718</td>
</tr>
<tr>
<td>255</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Burgers vs Wings + Bar</td>
<td>BUSINESS: Orchard Central, Singapore, Singapore</td>
<td>+65 6663 40423</td>
</tr>
<tr>
<td>256</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Artichoke</td>
<td>BUSINESS: 161 Middle Road (Sculpture Square), Singapore</td>
<td>+65 6451 4362</td>
</tr>
<tr>
<td>258</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>The Rice Table</td>
<td>BUSINESS: International Building 360 Orchard Rd 02-0910, Singapore, Singapore</td>
<td>+65 6835 3783</td>
</tr>
<tr>
<td>265</td>
<td>ECL</td>
<td>RESTAURANT FILE 1</td>
<td>Tai Dong Teochew Braised Duck Rice</td>
<td>BUSINESS: 49 Sims Place 01-04, Singapore, Singapore</td>
<td>+65 3800 49</td>
</tr>
</tbody>
</table>
Michelle Jonas | 3 records across 3 data sources

MICHELLE JONAS
WORK: 213 PACIFIC STREET SANTA MONICA... PRIMARY EMAIL: MS.MICHEL...
PRIMARY: 310-498-5493
MOBILE: +1 (310) 498-5493
HOME: 310-450-2216
ALT 1 EMAIL: MICHELLE@JEF...
EMAIL: MS.MICHELLEJONAS...

FEEDBLITZ SUBSCRIBERS 460761732D6642E675E636..
EMAIL: ms.michelle.jonas@g...subscriberid: 57649308
listid: 55826

MAILCHIMP SUBSCRIBERS F3A46FB6EDEBA1351EC5...
EMAIL: ms.michelle.jonas@g...optin_ip: 47.151.157.225
confirm_ip: 47.151.157.225

OUTLOOK CONTACTS D2E8A5D43E9630E2C1FF...
Jonas Michelle
WORK: 213 Pacific Street Santa Monica CA 9... EMAIL: ALT 1: michelle@jeff...
PRIVACY BY DESIGN
IN ENTITY RESOLUTION SYSTEMS
Privacy by Design in Entity Resolution Systems

- Full Attribution ✓
- False Negative Favoring ✓
- Self-Correcting False Positives ✓
- Data Tethering
- Information Transfer Accounting
- Tamper Resistant Audit Log*
- Selective Field Hashing

* = Requires Blockchain integration

Privacy by Design
in the Age of Big Data

June 8, 2012

Ann Cavoukian, Ph.D.
Information & Privacy Commissioner
Ontario, Canada

Jeff Jonas
IBM Fellow
Chief Scientist, IBM Entity Analytics

TAMPER RESISTANT AUDIT LOG

USE CASE: REDUCING “SEARCH ABUSE”
Single Subject Search with Blockchain

Blockchain

Single Subject Search

Entity Resolution

Central Index

Search Results

Search
Sam Preston
sam@home.com
Tamper Resistant (Immutable) Audit Logs

Implementing a Trusted Information Sharing Environment
Using Immutable Audit Logs to Increase Security, Trust, and Accountability

A paper by The Markle Task Force on National Security in the Information Age

Markle Foundation

SELECTIVE FIELD HASHING
USE CASE: VOTER REGISTRATION MODERNIZATION
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 300M records for more than 24 states.

Shane Hamlin, Executive Director
Has the Voter Moved?

Maryland

Voters

Voted 2012
Robert E. Jones
6601 Ritchie Highway, MD
DOB: 12/14/75
SSN: 5598

Dept of Motor Vehicles

Virginia

Voters

Voted 2016
Bob Jones
382 Taylor Drive, VA
Born: 12/14/75
D/L: 5329

Dept of Motor Vehicles

Social Security

Deceased Persons

Bob Jones
382 Taylor Drive, VA
Born: 12/14/75
D/L: 5329 SSN: 5598

Insight: Maryland, your voter may have moved!
Has the Voter Moved?

Maryland

Voters

Dept of Motor Vehicles

Voted 2012
Robert E. Jones
6601 Ritchie Highway, MD
DOB: 00c9782a552a2d09b1b8
SSN: 7f2b6e48ea7d042bbee85e

Insight: Maryland, your voter may have moved!
IN CLOSING
3 Advantages of an Entity Resolved Index

1. Reduces the need to make bulk data transfers to new warehouses, lakes, etc.

2. Eliminates contamination introduced by federated search logs
Federated Search = Search Log Contamination

CUSTOMERS

LOYALTY CLUB

MARKETING

CUSTOMER SERVICE

PAYROLL

Search logs

Search logs

Search logs

Search logs

Search logs

Search

Liz Reston
reston@home.com
3 Advantages of an Entity Resolved Index

1. Reduces the need to make bulk data transfers to new warehouses, lakes, etc.
2. Eliminates contamination introduced by federated search logs
3. Entity resolved indexes improve search accuracy
“For the individual rights of the data subjects to be exercised and handled, **companies need to have an effective method for single-subject search** – to produce the records that respond to the request, while screening out all the other records.”
Big Data. Big Responsibility.