Identity Proofing Standards and Beyond

Kimberly Little Sutherland
LexisNexis Risk Solutions
Sr. Director, Identity Management Solution Strategy
Agenda

- Identity Proofing Primer
- Identity Proofing Standards
- Identity Proofing Challenges
- More on Identity Resolution
- Other Considerations
- Summary
Identity Proofing Primer
Identity Management is a process designed to correctly identify the right individual for the right purpose.

Identity Management risks can occur due to:

- Misidentification
- Impersonation or theft
- Fabrication or use of a synthetic identity, and
- Collusion
Identity Proofing Defined

• No consistent definition for identity proofing
• Goes by multiple names

Basic definition:
The process by which an organization collects and verifies information about a person for the purpose of account opening or issuing credentials to that person.

Other terms:
Identity Verification, Verification and Authentication
The top reasons for incorporating identity proofing processes:

– Regulatory compliance
– Enhancing fraud prevention
– Improving business efficiencies

Identity Proofing decisions impact nearly every user interaction, beginning at registration. These choices will not only impact a company’s technology decisions, but also user experience processes, and even privacy policies.
Identity Proofing Standards and Regulations

There is a patchwork landscape at the federal, state, and local level and for commercial industries for identity proofing and identity authentication.

**Government**
- **Federal**: NIST 800-63-2 (eAuthentication)

**Commercial**
- **Banking**: Section 326 of the USA PATRIOT Act (Customer Identification Program)
- **The Nevada Gaming Commission and State Gaming Control Board**: Regulation 5A (110 - Registration of Authorized Player)
NIST 800-63-2

NIST Special Publication 800-63-2

Electronic Authentication Guideline

William E. Burr
Donna F. Dodson
Elaine M. Newton
Ray A. Perlner
W. Timothy Polk
Sarbari Gupta
Emad A. Nabbous

COMPUTER SECURITY

NIST
National Institute of Standards and Technology
U.S. Department of Commerce
NIST 800-63 requirements include measures to ensure that:

• A person with the Applicant’s claimed attributes exists, and those attributes are sufficient to **uniquely identify a single person**;

• The Applicant whose token is registered **is** in fact the person who is entitled to the identity;

• It is difficult for the Claimant to later **repudiate** the registration and dispute an authentication using the Subscriber’s token.
In-Person Identity Proofing Levels of Assurance

ASSURANCE

LOW  MED-LOW  MED-HIGH  HIGH

RISK

HIGH  MED-HIGH  MED-LOW  LOW

Level of Assurance 1
Level of Assurance 2
Level of Assurance 3
Level of Assurance 4
Remote Identity Proofing Levels of Assurance

- Level of Assurance 1
- Level of Assurance 2
- Level of Assurance 3
- Level of Assurance 4
# The Basic Components of Identity Proofing

| RESOLVE | Resolve the identity  
Capture and disambiguate the data to identify one unique identity.  
“Who you are?”  
Bob Jones |
|---|---|
| VERIFY | Verify the identity  
Establish that the identity exists.  
“Does Bob Jones exist?” |
Identity Proofing at LOA2

**In Person**

**Verify Government ID**
Inspect photo-ID, compare picture to applicant, and record the ID number, name, address and date of birth
- Inspection and
- Recording

**Verify Contact Information and Ability to Receive Communication**
Confirmation of email, phone number or address

**Remote**

**Verify Personal Identity**
Confirm accuracy of application data to identify a unique individual
- Full name
- Address
- Date of birth (full or partial)
- Social Security Number*

**Verify Government ID Number**

**Verify Account Number**
- Verify Financial Account
- Verify Utility Account
- Verify Phone Number and Message Confirmation

Knowledge Based Authentication (optional)
Identity Proofing Challenges
Identity Proofing Challenges

Garbage in, garbage out
• Data hygiene and reducing duplicate records is critical to having effective identity proofing processes.

One way only
• Organizations should plan for alternate online workflows and contact channels, and try to limit manual reviews (more costly, not real-time).

Citizens of the world
• Organizations should ensure their KYC processes can be used for US citizens with established identities, those with limited societal footprints, and even individuals temporarily residing in or visiting the US.
Identity Proofing Challenges

In-Person vs. Remote
- Organizations must identify the risks for both in-person and remote proofing and establish some form of equivalence in both processes. Also account for hybrid processes.

Virtual Reality
- Organizations should avoid overreliance on self-asserted data in virtual/remote processes. Identify methods to strengthen and enhance the data collection process.

Universal Translator
- Without a common lexicon, equivalence of identity attributes, and consistent identity resolution processes, organizations need to seek clarity in their identity proofing processes.
More on Identity Resolution
IDPV Identity Attribute Framework

Below is a summary of our framework for selecting identity attributes:

1) EFFECTIVENESS: How EFFECTIVE is the attribute at distinguishing an identity?

2) SENSITIVITY: Is the person SENSITIVE about the data attribute? Are they offended that we would ask such a question?

3) ACCESSIBILITY: Can the enroller verify the attribute? Does the enroller have ACCESS to resources that know?

4) PERMANENCE: Is this attribute stable over time?

5) NECESSITY: Is the attribute NECESSARY, something that I am required to collect in my business processes?
The total “core identities” in the LexisNexis analysis: 268,182,500
- Identities determined to be below the age of 18 have been excluded from the analysis
- Deceased identities will not be excluded from the analysis
- This is not intended to be a representation of the US population

US Population comparison
• The total US resident population, 18 years and older*: 234,564,071
  (population estimate from the 2010 US Census, US Census Bureau)

• The US annual deaths are estimated at 2.4M
  (based on Centers for Disease Control and Prevention’s National Vital Statistics System – mortality demographics: Work tables GMWK23F)
A total of 112 distributions were performed:

• Location-oriented - 29
• Time-oriented - 12
• Identifier-oriented – 12
• Advanced combinations – 60

* Name was used in all distributions
Identity Resolution Project: Distribution Details

Birth date-oriented identifying data (12):

- 1. Distribution of “Name” crossed by “DOB”
- 2. “Name” should include the following variants:
  - a. First Name, Middle Initial and Last Name
  - b. First Name and Last Name
  - c. First Initial and Last Name
  - d. First Name

- 3. “DOB” should include the following variants:
  - a. Full Date of Birth
  - b. Year of Birth
  - c. Month and Day
### Identity Resolution Project: Example of a Distribution

<table>
<thead>
<tr>
<th>count of Identities</th>
<th>count of DOB + fname + lname</th>
<th>% DOB + fname + lname of total core ADL</th>
</tr>
</thead>
<tbody>
<tr>
<td>cnt = 1</td>
<td>228,542,805</td>
<td>0.96291</td>
</tr>
<tr>
<td>cnt = 2</td>
<td>3,294,261</td>
<td>0.01388</td>
</tr>
<tr>
<td>cnt = 3</td>
<td>356,970</td>
<td>0.00150</td>
</tr>
<tr>
<td>cnt = 4</td>
<td>93,562</td>
<td>0.00039</td>
</tr>
<tr>
<td>cnt = 5</td>
<td>36,889</td>
<td>0.00016</td>
</tr>
<tr>
<td>cnt = 6</td>
<td>18,508</td>
<td>0.00008</td>
</tr>
<tr>
<td>cnt = 7</td>
<td>11,081</td>
<td>0.00005</td>
</tr>
<tr>
<td>cnt = 8</td>
<td>6,996</td>
<td>0.00003</td>
</tr>
<tr>
<td>cnt = 9</td>
<td>5,010</td>
<td>0.00002</td>
</tr>
<tr>
<td>cnt = 10</td>
<td>3,509</td>
<td>0.00001</td>
</tr>
<tr>
<td>cnt &gt; 10</td>
<td>14,278</td>
<td>0.00006</td>
</tr>
<tr>
<td>total DOB- oriented cnt</td>
<td>237,346,451</td>
<td>1.00000</td>
</tr>
<tr>
<td>total core ADL</td>
<td>268,154,531</td>
<td></td>
</tr>
</tbody>
</table>
The number of females born in 1970 with the name Kimberly

34,135
First Name, Last Name Combinations are Not So Unique

The number of females named Kimberly Little in the US: 1000+
First Name, Last Name with Location Provides More Focus

- **104** people named Kimberly Little in GA
- **3** people named Kimberly Little in Alpharetta, GA
  - All three have different years of birth (1962, 1963 and 1970)
  - One of us has a different zip code
Identity Resolution Project: Attribute Guidelines

Core Attributes must uniquely resolve at least 95% of identities:

– First Name, Last Name + SSN 9 (96.52%)
– First Name, Last Name + Full DOB (96.29%)
– First Name, Last Name + SSN4 + ....
Identity Resolution Project: Attributes Analyzed

**Name:**
- a. First Name, Middle Initial and Last Name
- b. First Name and Last Name
- c. First Initial and Last Name
- d. First Name

**Location:**
- a. Full Address
- b. Street Address
- c. City and State
- d. County and State
- e. House Number
- f. Secondary Number (Suite or Apt No.)
- g. Street Name
- h. Postal Code

**Time:**
- a. Full Date of Birth
- b. Year of Birth (YYYY)
- c. Month and Day of Birth (MMDD)

**Identifier:**
- a. Full Social Security Number (SSN9)
- b. First 5 digits of the Social Security Number (SSN5)
- c. Last 4 digits of the Social Security Number (SSN4)
### Table 1 – Selection of Core Biographical Attributes

<table>
<thead>
<tr>
<th>Purpose of Disclosure</th>
<th>Attribute Category</th>
<th>Ref No.</th>
<th>Attribute (Fields)</th>
<th>Equivalent Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D                                                                科</td>
<td>Set 1</td>
</tr>
<tr>
<td>Selection of Core</td>
<td>Name</td>
<td>CA1</td>
<td>Name (Legal First Name and Legal Last Name)</td>
<td>M</td>
</tr>
<tr>
<td>Biographical Attributes for: IAL 2; IAL 3; and IAL 4.</td>
<td>Location</td>
<td>CA2</td>
<td>Partial Current Address (Postal Code) or (City and State)</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>CA3</td>
<td>Place of Birth (City or County) and (State or Foreign Country)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>CA4</td>
<td>Partial Date of Birth (Month and Day) or (Year)</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>CA5</td>
<td>Full Date of Birth</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Identifier</td>
<td>CA6</td>
<td>Social Security Number – Last 4 Digits</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>Identifier</td>
<td>CA7</td>
<td>Social Security Number – Full 9 Digits</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Choice of Identity for IAL 1</td>
<td>CA8</td>
<td>Any Self Selected Personal Identifier</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The above table outlines the selection of core biographical attributes for different purposes, with specific attributes listed for name, location, time, and identifier purposes, along with their respective equivalent sets.
## Table 2 – Supplemental Biographical Attributes

<table>
<thead>
<tr>
<th>Purpose of Disclosure</th>
<th>Attribute Category</th>
<th>Ref No.</th>
<th>Attribute (Fields)</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
<th>Set 4</th>
<th>Set 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unique Identification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>SA1</td>
<td>Mother’s Name at Birth or Prior to First Marriage</td>
<td>O*</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA2</td>
<td>Middle Name or Initial</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>SA3</td>
<td>Place of Birth - Country</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA4</td>
<td>Place of Birth - State</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA5</td>
<td>Place of Birth - City</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA6</td>
<td>Current Address - State</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA7</td>
<td>Current Address - City</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA8</td>
<td>Current Address - Street</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA9</td>
<td>Previous City</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>SA10</td>
<td>Full Date of Birth</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Identifier</td>
<td>SA11</td>
<td>Social Security Number – All 9-digits</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA12</td>
<td>Social Security Number – First 5-digits</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA13</td>
<td>Social Security Number – Last 4-digits</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA14</td>
<td>Sex</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>Other Attributes Combined with the Name Attributes</td>
<td>Name Attributes</td>
<td>No. of Valid (Non Null) Identities</td>
<td>% Null Identities</td>
<td>% Resolved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>---------------------</td>
<td>-----------------------------------</td>
<td>-------------------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN9</td>
<td>SSN9 Only</td>
<td>266,933,071</td>
<td>0.46</td>
<td>82.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN9</td>
<td>+ fname + mname init + lname</td>
<td>226,014,252</td>
<td>15.71</td>
<td>97.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN9</td>
<td>+ fname + lname</td>
<td>259,421,904</td>
<td>3.26</td>
<td>96.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN9</td>
<td>+ fname init + lname</td>
<td>259,421,904</td>
<td>3.26</td>
<td>95.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN9</td>
<td>+ fname</td>
<td>259,802,288</td>
<td>3.11</td>
<td>95.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN5</td>
<td>+ fname + mname init + lname</td>
<td>226,014,252</td>
<td>15.71</td>
<td>96.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN5</td>
<td>+ fname + lname</td>
<td>259,421,904</td>
<td>3.26</td>
<td>93.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN5</td>
<td>+ fname init + lname</td>
<td>259,421,904</td>
<td>3.26</td>
<td>74.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN5</td>
<td>+ fname</td>
<td>259,802,288</td>
<td>3.11</td>
<td>18.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN4</td>
<td>+ fname + mname init + lname</td>
<td>226,014,252</td>
<td>15.71</td>
<td>96.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN4</td>
<td>+ fname + lname</td>
<td>259,421,904</td>
<td>3.26</td>
<td>93.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN4</td>
<td>+ fname init + lname</td>
<td>259,421,904</td>
<td>3.26</td>
<td>69.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN4</td>
<td>+ fname</td>
<td>259,802,288</td>
<td>3.11</td>
<td>13.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Below is a summary of our framework for selecting identity attributes:

1) EFFECTIVENESS: How EFFECTIVE is the attribute at distinguishing an identity?

2) SENSITIVITY: Is the person SENSITIVE about the data attribute? Are they offended that we would ask such a question?

3) ACCESSIBILITY: Can the enroller verify the attribute? Does the enroller have ACCESS to resources that know?

4) PERMANENCE: Is this attribute stable over time?

5) NECESSITY: Is the attribute NECESSARY, something that I am required to collect in my business processes?
Identity Resolution Project - Sensitivity

Q: Which of the following types of companies or organizations would you trust with your personal identifying information if you were to open and maintain an account with them?

- A bank or financial institution: 60% Top 3 Box, 34% Neutral, 5% Bottom 3 Box
- A health care provider (such as, hospitals, clinics, healthcare insurance providers): 50% Top 3 Box, 43% Neutral, 7% Bottom 3 Box
- Government offices (such as IRS, human services, etc.): 45% Top 3 Box, 40% Neutral, 15% Bottom 3 Box
- Auto, life or home insurance providers: 38% Top 3 Box, 53% Neutral, 9% Bottom 3 Box
- A retail or consumer goods company to receive coupons or promotions: 17% Top 3 Box, 57% Neutral, 26% Bottom 3 Box

Consumer Willingness to Share PII: 
With Government Agencies When Filing a Tax Return

Q: You are filing your Federal tax return to receive a tax refund, please answer how willingly you will share the information in this scenario. Please rate on a scale of 1 to 10 where 1 means ‘Not At All Willing’ and 10 means ‘Totally Willing’. TOP 3 BOX SHOWN

Consumer Willingness to Share PII: When Opening a New Financial Account

Q: You are opening a bank checking or savings account (either at a branch or online over the internet). For each of these personal identifying information, please answer how willingly you will share the information in this scenario. Please rate on a scale of 1 to 10 where 1 means ‘Not At All Willing’ and 10 means ‘Totally Willing’. TOP 3 BOX SHOWN

- Age: 81%
- Full Name (First and Last Name): 79%
- Sex at birth: 79%
- Partial Date of Birth (Month and Day only): 78%
- Partial Address (Street Address, City and State or Zip only): 78%
- Full Address (Street Address, City, State and Zip): 77%
- Full Date of Birth (Month, Day and Year): 75%
- Full Name (First, Middle and Last Name): 74%
- Email Address: 73%
- Telephone Number: 73%
- Partial Social Security Number (First 5 Digits or Last 4): 66%
- Place of Birth (City or County and State or Country): 64%
- Mother’s maiden name or name at birth: 56%
- Full Social Security Number (9 Digits): 54%
- Screen name or user ID: 52%

December 2013, n = 606.
Base: Consumers assigned to Financial Services Group.
Consumer  Willingness to Share PII: With Government Agencies When Registering To Receive Program Details or Benefits

Q: You are registering to receive general program or benefit details, please answer how willingly you will share the information in this scenario. Please rate on a scale of 1 to 10 where 1 means ‘Not At All Willing’ and 10 means ‘Totally Willing’. TOP 3 BOX SHOWN

Q: You are returning an online purchase for a refund. For each of these personal identifying information, please answer how willingly you will share the information in this scenario. Please rate on a scale of 1 to 10 where 1 means ‘Not At All Willing’ and 10 means ‘Totally Willing’. TOP 3 BOX SHOWN

December 2013, n = 605.
Base: Consumers assigned to Retail Group.
Below is a summary of our framework for selecting identity attributes:

1) **EFFECTIVENESS**: How EFFECTIVE is the attribute at distinguishing an identity?

2) **SENSITIVITY**: Is the person SENSITIVE about the data attribute? Are they offended that we would ask such a question?

3) **ACCESSIBILITY**: Can the enroller verify the attribute? Does the enroller have ACCESS to resources that know?

4) **PERMANENCE**: Is this attribute stable over time?

5) **NECESSITY**: Is the attribute NECESSARY, something that I am required to collect in my business processes?
Future Research: Establishing Equivalence of Identity Attributes

**HIGH TRUST**
- Biometrics
- Expert/electronic ID document authentication
- Verified bank account

**MODERATE TRUST**
- Dynamic knowledge based authentication
- Electronic copy of passport
- Electronic copy of drivers license
- Mobile number
- Social login

**LOW TRUST**
- Email address

*Source: MiiCard (adapted)*
Summary

1. Despite the fact that identity proofing is a foundational activity for user registration/new account opening activities and credential issuance, there are limited standards to guide organizations in this process.

2. Resolution to one unique individual should be clearly understood in the proofing process. Extra attention should be given to identity resolution when multiple databases and/or vendors are used in the proofing process.

3. Identity proofing should align with:
   - The assessed risk based on the type of user activities to be permitted and whether the process will performed be in-person, remote or a hybrid.
   - The interaction channel being leveraged (via face-to-face, kiosk, website, a mobile browser or mobile app).

4. Future research is needed for identity proofing in terms of identity resolution, establishing equivalence of identity attributes, new methods for identity proofing and strengthening the proofing process.
Questions?
Thank You

Kimberly Little Sutherland,
Sr. Director, Identity Management Solution Strategy
LexisNexis Risk Solutions
idmanagement@lexisnexis.com
www.lexisnexis.com/risk