Identity Matching Lab

Keith J. Miller, Principal Investigator

703-983-6920

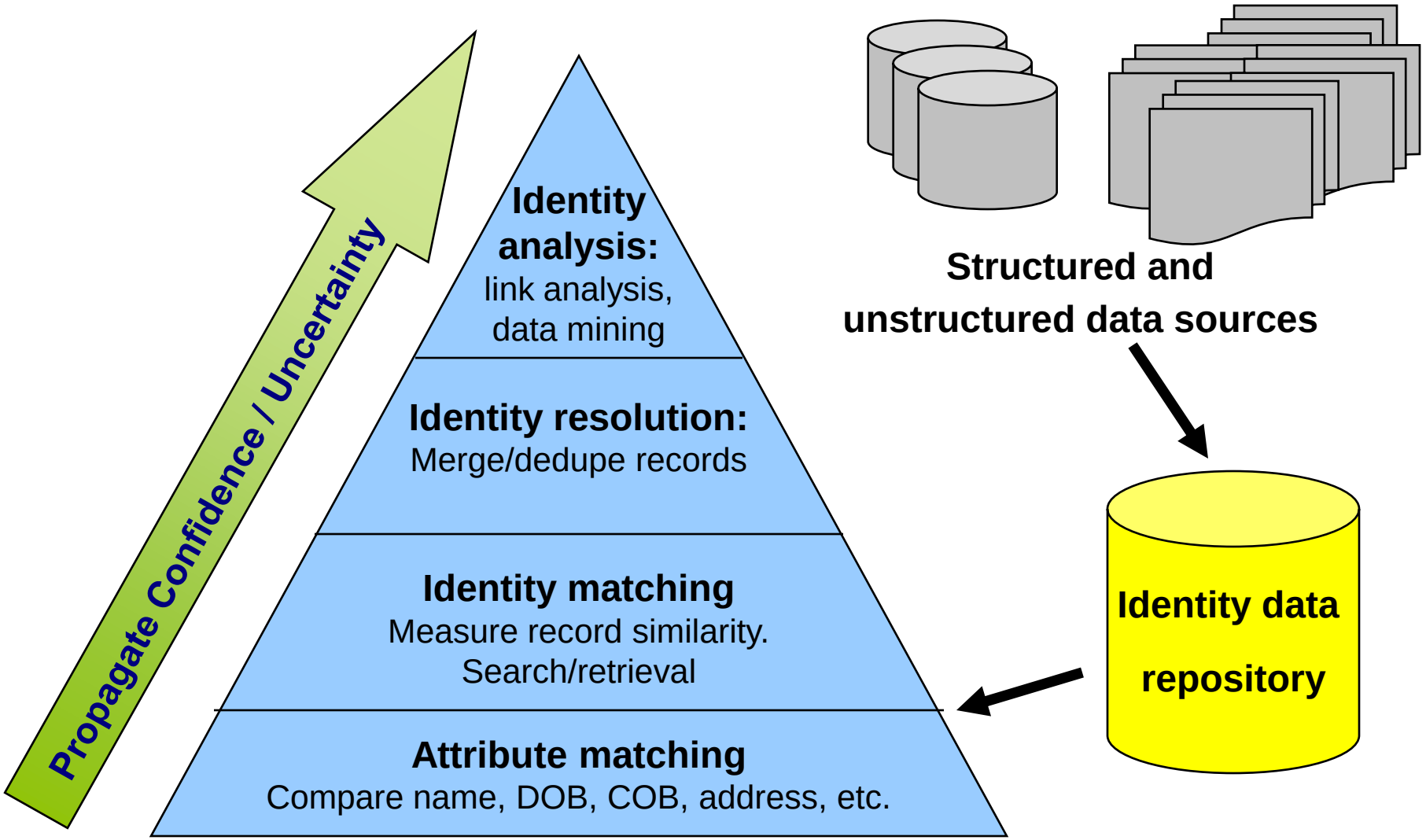
keith@mitre.org

Mark Arehart, Catherine Ball, Gail Hamilton, John Polk, Alan Rubenstein, Ken Samuel, Liz Schroeder, Genevieve Whiddon, Chris Wolf (E544, E549, G122, and F051)

Problem

- Maintain/search databases containing identity data
  - Sparse, noisy data from multiple sources
  - Matching multicultural name variants
  - Algorithms customized for multiple identity attributes (DOB, COB, etc.)
  - Combining evidence from multiple identity attributes (biographical and eventually biometric)
- Evaluation
  - Difficult to assess quality/accuracy of COTS/GOTS systems
  - No “standard” approaches or best practices
- Applications
  - Watchlisting/screening/credentialing
  - Identity management/identity resolution

Background



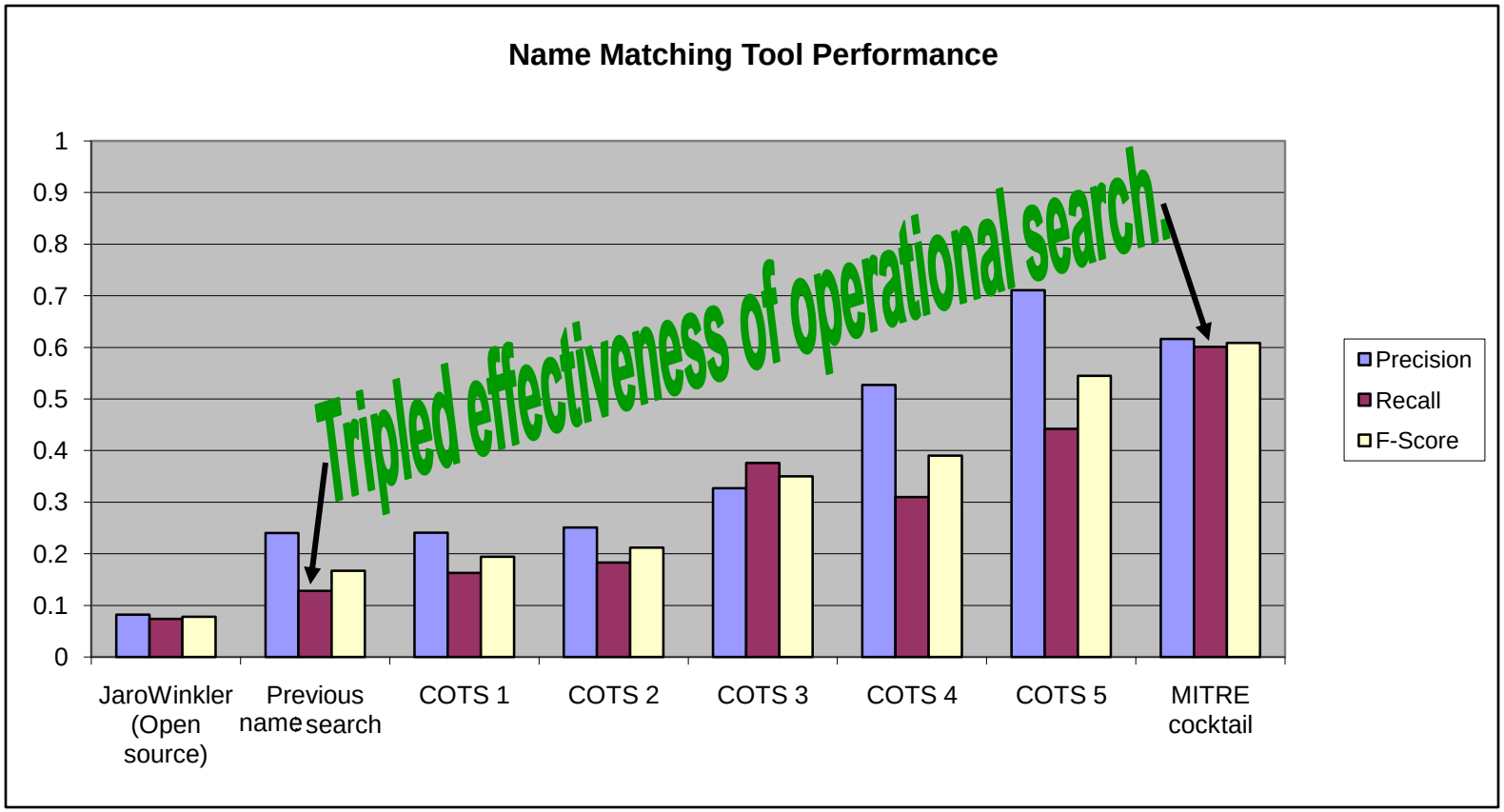
Objective

- Provide data, tools, and infrastructure for objective, user-focused measurement and evaluation of identity matching technologies
- Develop novel combinations of technologies that increase sponsors' ability to perform identity matching
- Develop technical solutions to maximize sponsors' effective use of these technologies

Activities

- Evaluation
  - Evaluate COTS, GOTS, and open source name matching tools for TSC and Homeland Security stakeholders
  - Evaluate effects of data quality and data policy on identity matching performance
  - Develop framework for evaluating identity resolution tools
- Prototyping
  - Designed improved search “cocktail” for TSC
  - Methods for aggregating results from multiple algorithms
- Tools and resources
  - Web-based tools for collecting adjudications, viewing metrics
    - Web-based evaluation wizard for identity matching
  - Multicultural name-matching corpus for Federal Identity Matching Working Group

Highlight



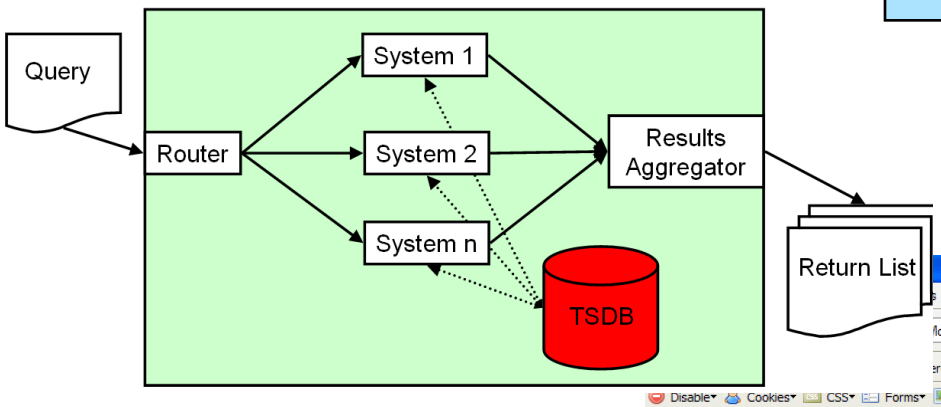
Original search at bottom of COTS performance

MITRE-developed composite search outperforms best COTS tool

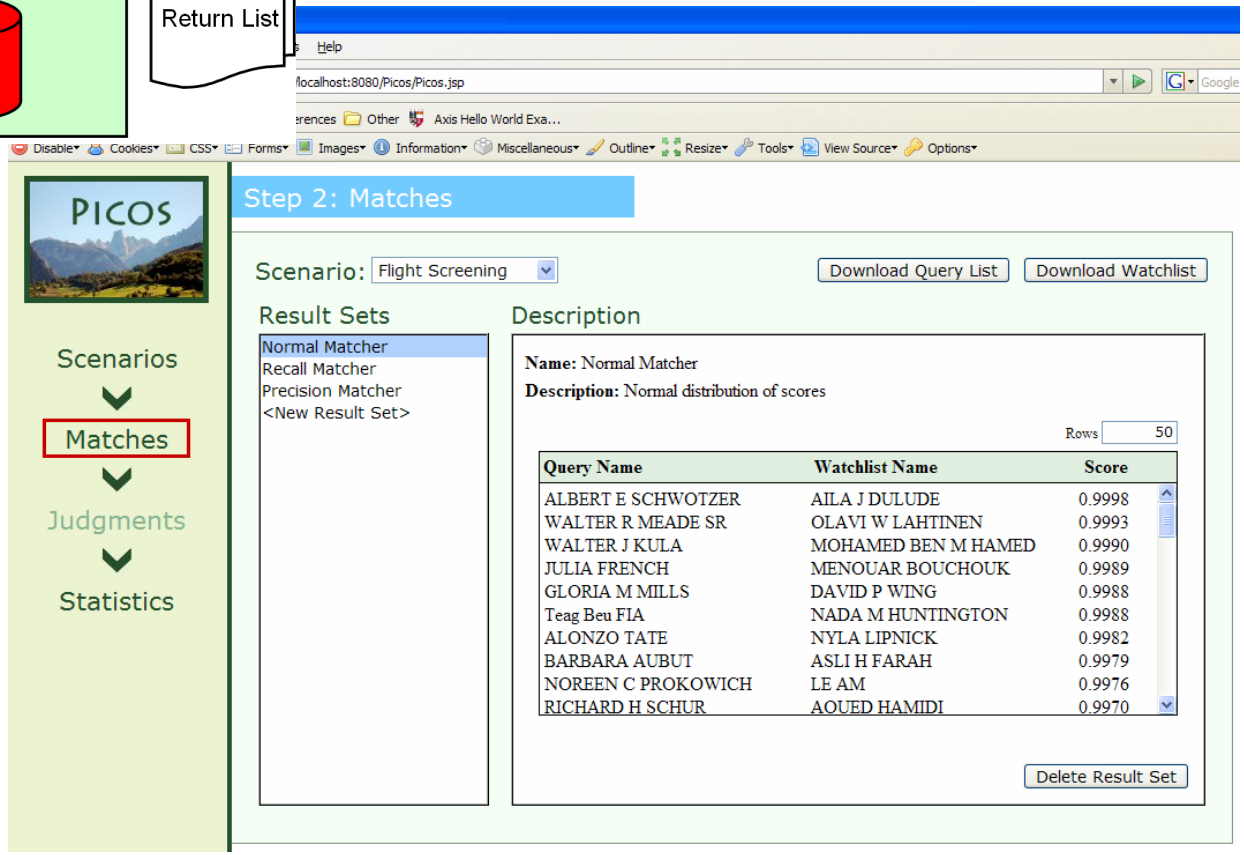
Demonstration

Recent briefs to various sponsors, coordination with several MITRE biometrics efforts

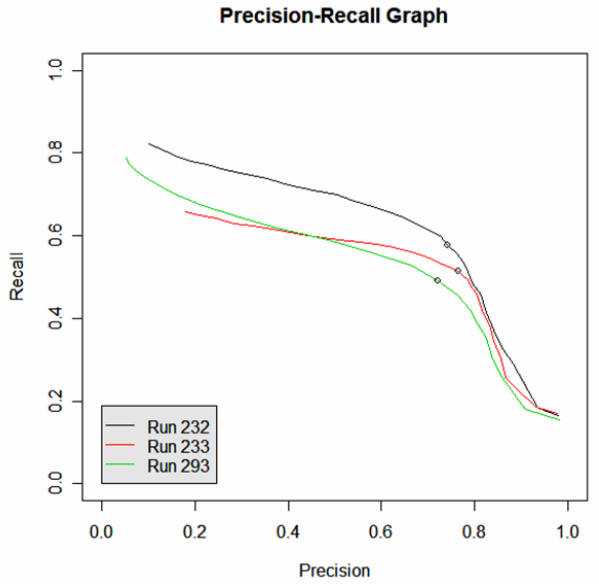
MITRE Innovation Exchange



Identity Matching Evaluation Wizard



System combination



MITRE

Impacts

- **Operational Identity Matching**
  - Designed “cocktail” that tripled the effectiveness of TSC’s identity matching
  - Developed operational prototype built on IML data model for another sponsor org.
- **Cross-MITRE / Cross-Government Collaboration**
  - IML replicated in Kevin DeFord’s iLab (G06A) to support sponsor work on entity resolution/management in collaboration with Eric Hughes (G062)
  - Discussions with MITRE staff working on anonymization/privacy concerns: Aberdeen (G063), Bakis and Garrison (G022), Shapiro (G022)
  - Currently participating in biometrics identity matching study team – MITRE POC: D. Lloyd (G122)
  - Supporting research on cross-script matching and evaluation (S. Condon, E544)
- **Support to TSC-sponsored Federal Identity Matching Working Group**
  - Defined guidelines, tools, and data set for evaluating name matching
  - Tools in beta at three USG sites; due to be released gov’t-wide summer 2008

Future Plans

- **Proposal for Infrastructure and methods for evaluating Identity Resolution tools**
  - **Matching on multiple identifying attributes**
  - **Address temporal nature of merging and splitting identities**
  - **Currently-funded MITRE Innovation Grant (IG) to begin investigation of appropriate evaluation methods**
- **Improving identity matching aggregation methods**
- **Researching approaches for minimizing the amount of manual "ground-truthing" required for evaluation**