

UID Conference

Person Matching Overview

From Akbar to Zeb

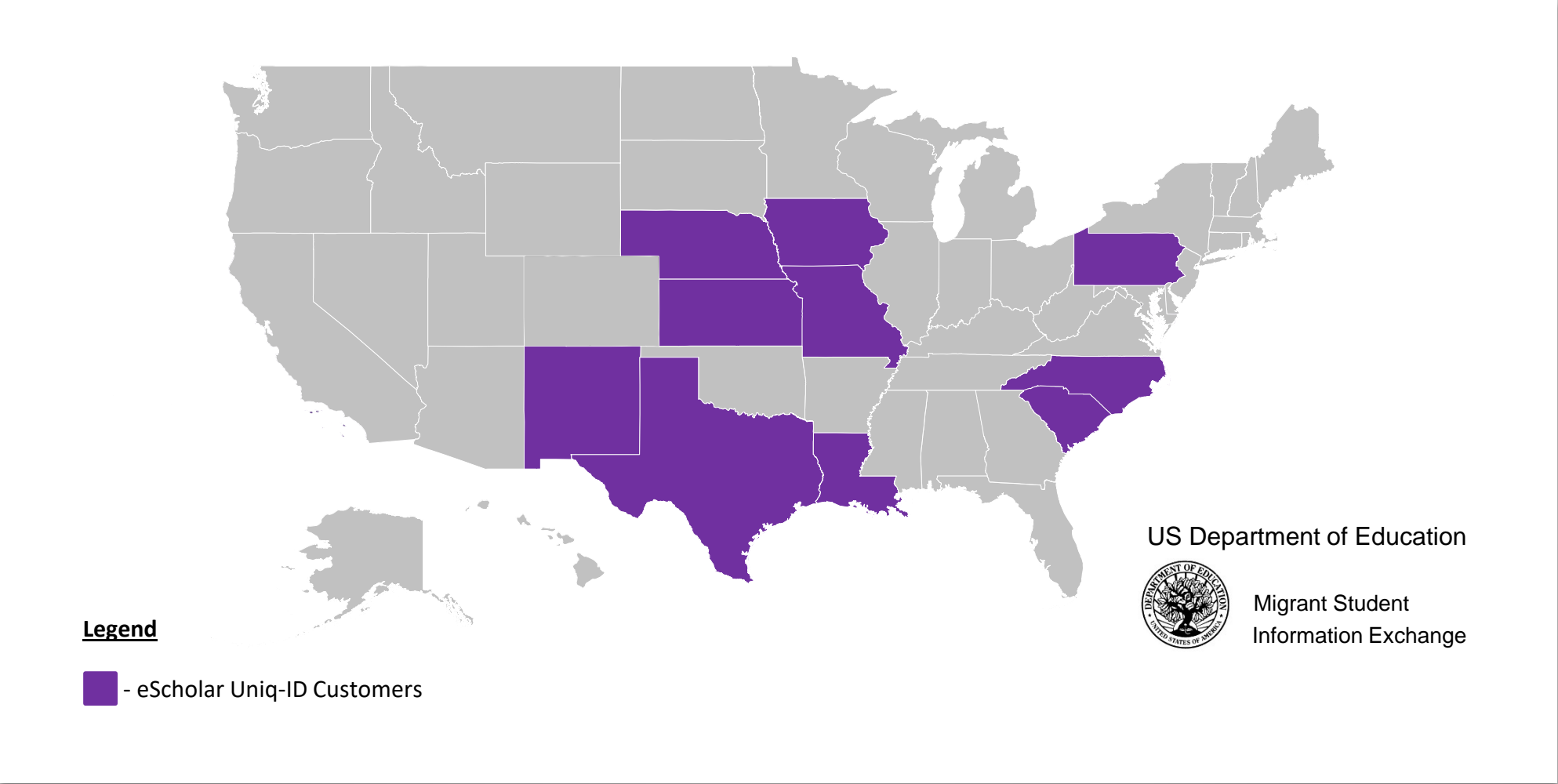
2018

Session Overview

- Learn how eScholar uses a blend of deterministic and probabilistic algorithms to match millions of students every year across 10 states and the Migrant Student Records Exchange Initiative (MISIX). Additionally, the team will introduce new ideas such as the Near Match Rule Engine which helps DirectMatch customers reduce the number of Near Matches they have to resolve manually.



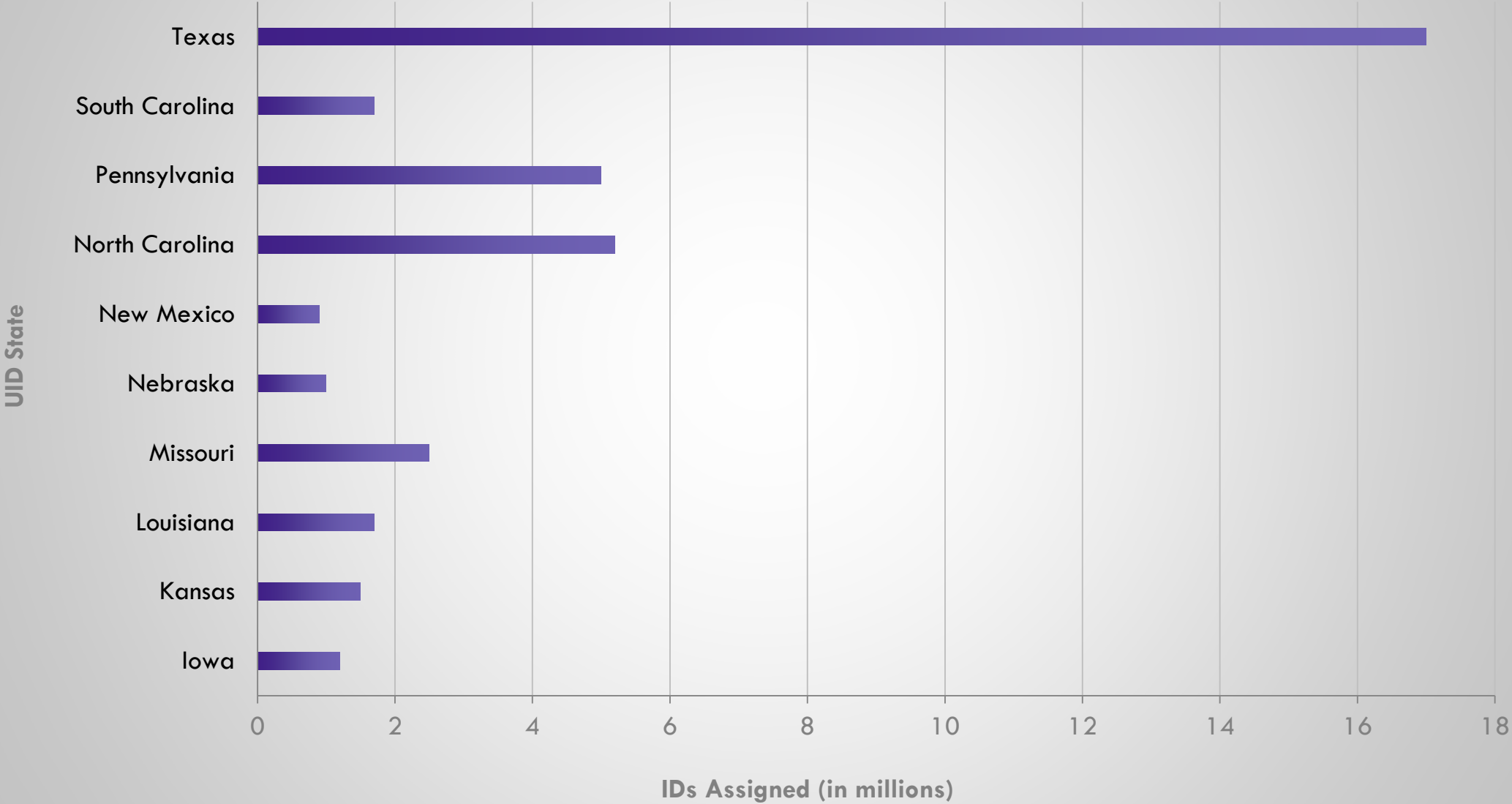
The Uniq-ID Community



42,000,000

ASSIGNED IDs!

IDs Assigned per State



Bye-Bye SSN



Matching Options

Deterministic



VS

Probabilistic

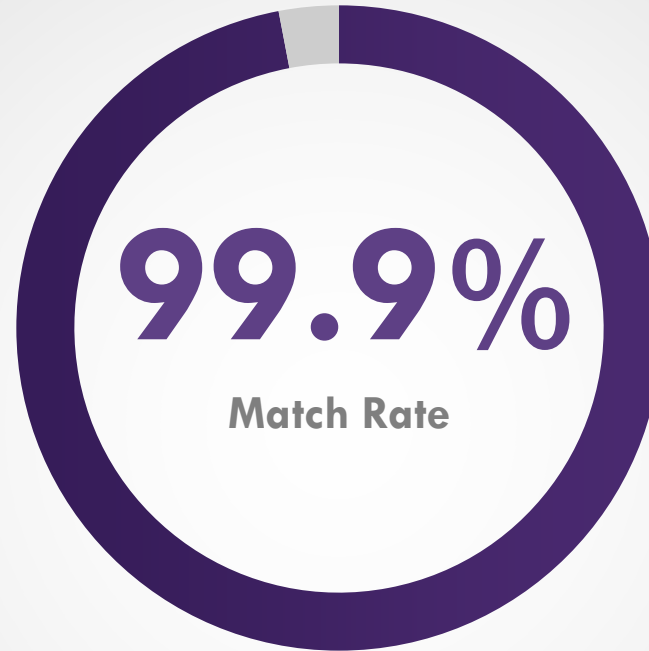


Two is better than one

Use both!



What is success?



Match Person Data with 99% accuracy

What is success?



Lightning Fast Matching



For probabilistic matching, you need to balance both

Your diverse student and staff population

Ethnicity

Nicknames

Frequency of Names

Multipart Names

Special Characters / Rules



Rules and Nicknames

Rules

support special characters: á, í, ü...

replace 0 (zero) with O

delete noise words: the, and

delete personal title: Mr, Ms, Dr...

delete prefix: (empty space)

join the following word: Mac

Nicknames

Jack = John

Chuck = Charles

Ted = Edward

Sally = Sarah

Dick = Richard

Tina = Christina

Multiple Data Sources



Source A

First Name
Last Name
Date of Birth



Source B

First Name
Last Name
Middle Name
Date of Birth
Gender



Source C

First Name
Last Name
Middle Name
Date of Birth
Gender
SSN
Alt. Last Name
Suffix

*Percentage is based on the availability of 8 matching data elements.

Flexible Configurations

Field Requirements



Field Weights



Matching Thresholds



Source Reliability



*The matching system is flexible enough to allow state administrators to configure matching properties for different datasets such as SNAP and Medicaid without needing to write new code each time fine tuning is needed.

Field Requirements and Weights

01

Optional or Required?

Matching should fail if a required field is not supplied. Fields should be set as individual fields or grouped fields for matching.

02

Weight

Each field should be scored, then multiplied by its weight, and the total score divided by the total weight.

03

Score Contribution

Each field should be designated as:

- 1) always contribute to the score,
- 2) only contribute to the score if above X,
- or
- 3) only contribute to the score if below X.

See it in Action

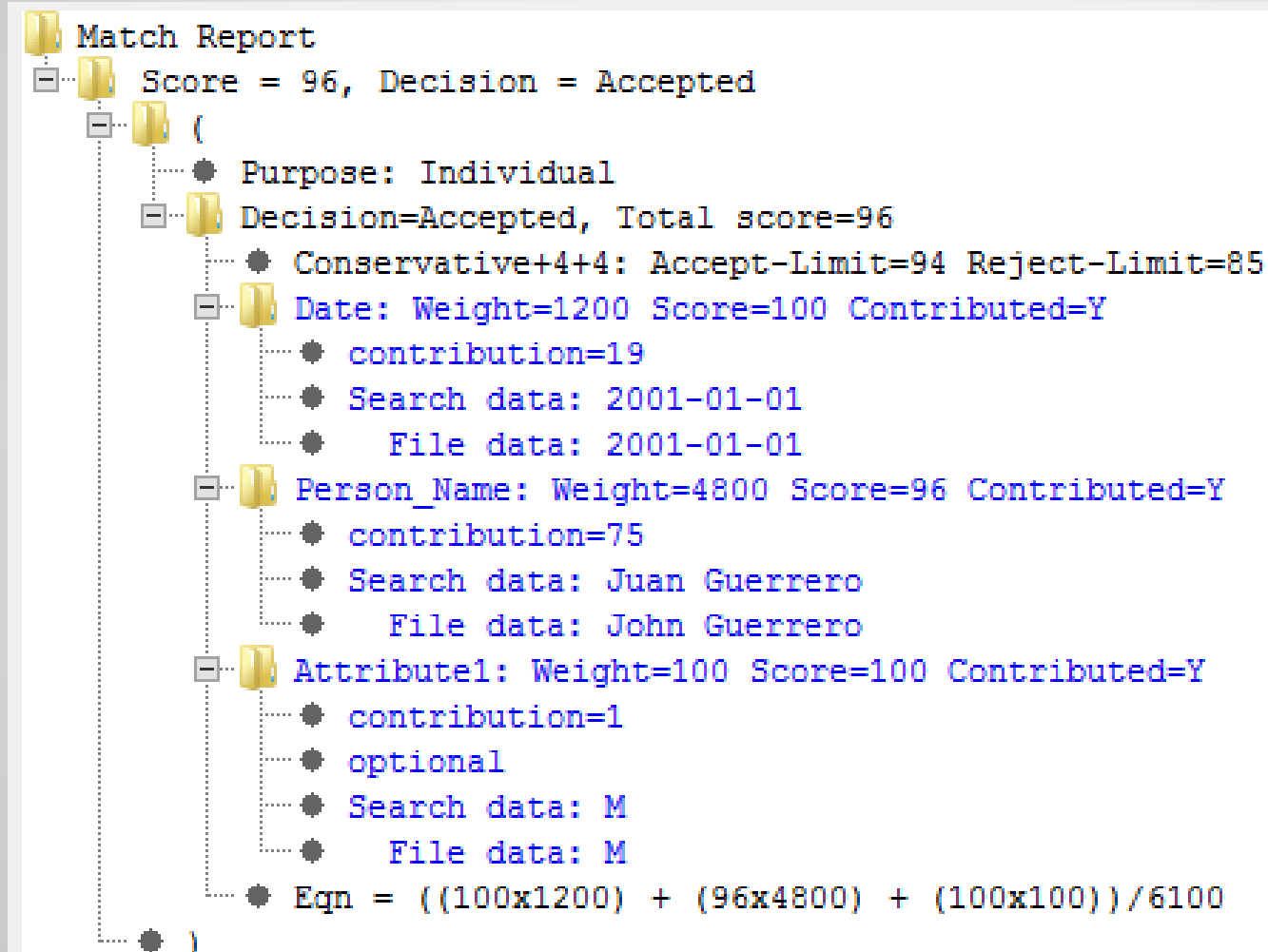
Search Data

```
*Person_Name*Juan Guerrero*DATE*2001-01-01*Attribute 1*M***
```

File Data

```
*Person_Name*John Guerrero*DATE*2001-01-01*Attribute 1*M***
```

See it in Action



Performance Considerations – Keys and Pre-Scoring

**John Michael
Smith**

S?H>>YVW
S?H>AZNQ
UU?\$>GVY
UU?\$AZNQ
Y?LMJ\$\$-
Y?NUA\$\$-
S?MEAZNQ
UU?/>G\$-

Jon M Smith

S*PS\$\$VW
S*VYIKPU
Y?KAR\$\$-
Y?PS\$\$KA

Yon Smith

S?H>AZNO
Y?LK\$\$\$\$-

Match Decision – Now What?

Matching Engine



User Interface

To achieve the highest possible match percentage, Uniq-ID incorporates human resolution. A workflow around near match decisions was built to allow users to resolve ambiguous matches through a User Interface

DirectMatch – Near Match Resolution Engine



Allows states the option to build rules to auto-resolve near matches

- ✓ Auto Cancel Near Matches
- ✓ Auto Resolve Near Match if a record has a score above a certain threshold, for example a score of 93
- ✓ All Rule Engine decisions are audited and available for future review

DM – Near Match Resolution Engine Example

STATE ID HOME

0000supr, CAYUGA ISD

Resolve Near Matches / Duplicates - Batch 653

STUDENT RECORD TO REVIEW AND SELECT

LAST NAME	FIRST NAME	MIDDLE NAME	ALT LAST NAME	SUFFIX	DATE OF BIRTH	GENDER	DISTRICT	SCHOOL	LOCAL ID
Alan	Charley				11/21/1999	FEMALE	178914	AAA	442261656

NEAR MATCHES / DUPLICATES FOUND

	LAST NAME	FIRST NAME	MIDDLE NAME	ALT LAST NAME	SUFFIX	DATE OF BIRTH	GENDER	DISTRICT	SCHOOL	LOCAL ID	MATCH PROBABILITY
<input type="radio"/>	Alan	Charlie				11/21/1999	FEMALE	250907	AAA	442261656	99 [NEAR MATCH]
<input type="radio"/>	Allen	Charley		Cabrera		10/11/1999	FEMALE	178914	AAA	371853966	88 [NEAR MATCH]

CANCEL RECORD

SELECT ANOTHER RECORD

CREATE NEW ID

ASSIGN SELECTED

Near Match Records – more than 1 record above 88

Lessons Learned



MORE IS NOT ALWAYS BETTER

Middle Name Initials



BEWARE OF OUTLIERS

Uncommon Matches



THE GOAL

Let the match process guide you to
the best matching logic

Success is more than just matching

Thank you!

Comments or questions? Visit us at www.escholar.com

