

De-duplication for National Scale Unique ID Projects



Real World Challenges and Lessons
Learned Using Multi-Biometric and
Biographic Approach

Real World Deployment and Lessons Learned

❖ Scenario:

- Middle-East country national entitlement program
- Database: 1M Arabic profiles incl. 10 fingerprint & face
- Fingerprint stored on credential for biometric verification
- No fraud screen at enrollment
- Non-ISO images for both fingerprint and face
- Many records with missing information and biometrics

❖ Problem Statement:

- Lack of uniqueness check meant multiple records for same individual.
- Lack of trust in the credential created a ripe environment for fraud

❖ Solution - WCC'S Targeted De-duplication Services:

- ELISE ID – multimodal Smart Search & Match platform
- De-duplication Search optimization using Cascaded Fusion methodology.
- Customized Report Generation
- Analytics to drive process improvements

ELISE

~1 M Arabic profiles
~1 M Face images
~10 M Fingerprints

محمد عبدالرحيم



Challenge: Non-standard Biometric Images and Missing Information

❖ Uncontrolled enrollment environment

- Remote areas
- Untrained operators
- Capture equipment not suited for enrollment
 - Webcam face image
 - Single Fingerprint images

❖ Difficult demographic population to enroll

- Minimal technology interface
- High percentage of laborers and farmers (poor fingerprint quality)
- High percentage of elderly (worn fingerprints)
- Headdress and face veils created difficulty for face recognition

Challenge: Non-standard Biometric Images and Missing Information

De-duplication Solution with ELISE

WCC DE-DUPLICATION AND UNIQUENESS CHECK SOLUTION powered by elise

Analytics

Reports **Identities**

Biographic information ★★☆☆☆

use in match

Unique ID# 33796215

Full Name ميهاربيا درحما

Mother's Full Name

Date of Birth 07 / 16 / 1974

Marital Status

Gender

National ID

Government ID

Civil Affairs ID

Operator ID


Enrolment Date 10 / 11 / 2009

Fingerprint ★★☆☆☆

use in match






Face ★★☆☆☆

use in match




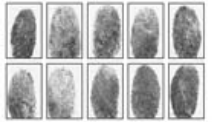
Clear all Search

Duplicate Reports **Duplicate Identities [5]**



Match %	Face	# Dups	Full Name	Mother's Name	ID	Gender	Type
100%		5	ميهاربيا درحما	يلوع نسح درحما	33796215	رلفذ	دعاقبتلا عىايه
98%		5	ميهاربيا درحما	يلوع نسح درحما	33796251	رلفذ	دعاقبتلا عىايه
95%		5	ميهاربيا درحما	يلوع نسح درحما	33796362	رلفذ	دعاقبتلا عىايه
90%		8					
80%		3					

Solution:
ELISE ID fused biometrics and Arabic text to increase accuracy and provide meaningful results

Operator ID 2148059140
Enrolment Date 11/10/2009
Unique ID 33796215
Full Name ميهاربيا درحما
Mother's Name يلوع نسح درحما
Date of Birth 16/07/1974

Operator ID 2148060679
Enrolment date 04/07/2009
Unique ID 33796251
Full Name عىايه بتجلا عىايه
Mother's Name يلوع نسح درحما
Date of Birth 16/07/1974

Use as Probe

« previous 1 2 3 4 5 6 ... 24 next »

Challenge: Limited Budget and Time Constraints

❖ Limited budgets created challenges for full solution stack

- Algorithms
- Large Scale Matching Platform
- De-duplication solution/service
- Server infrastructure
- Required hosted environment

❖ Political issues created timing urgency

- Fraud created strain on limited resources
- Fraud was source of funding for 'bad actors'
- Not providing entitlement benefits had potential of creating political instability

Challenge: Limited Budget and Time Constraints

The screenshot displays the ELISE ID Large-Scale Database Demo interface. The main search results table shows a 100.0% match for John Smiths, a male born on 11/18/1968, 6'0" tall, with blue eyes and brown hair, residing in Tampa, United States. The 'Cascading Match Settings' window is open, showing 'Cascaded Fingerprint Algorithms' with a filter for 'priv-ID Fast AFIS SDK v1.1' and a match using 'Neurotechnology VeriFinger SDK 6.0'. Below this is an 'ROC Curve - priv-ID' graph showing FRR vs. FAR, with a highlighted point at FAR: $7.5 \cdot 10^{-2}$ and FRR: $7.3 \cdot 10^{-2}$. The 'Solution' callout box contains the following text:

Solution

- Cascaded fusion minimize hardware footprint while maintaining accuracy and speed
- WCC's expertise provide de-duplication support and services for fast resolution of 1.5 million records

Challenge: Data Entry Errors

First name	Last name
John	Smiths
Smiths	John
	Smiths, John

Correct entry

Fields reversed

Fields concatenated
in one field

Solution

ELISE ID uses multiple algorithms to compensate for known error patterns

Challenge: Name Matching in Arabic Text

❖ Challenges in matching Arabic name

- Name variations in original script:

عبد الله - Abdallah

عبدالله – Abdella

احمد – Ahmad

- Name variations in romanized script”

‘Abdullah

Abd Allah

Abdalla

Solution

ELISE ID uses name databases to match variants of names – using native Arabic characters


Challenge: High Duplication Rates Explained in Actionable Reports

Fraud Report

Card ID: 9843111


Original Person	Duplication Score	Duplicate Person
Unique ID: 33796215	50%	Unique ID : 33796251
Full Name: ميهارب ا دمحم .	83%	Full Name: ميهارب ا دمحم ااا
Date of Birth: 1974-07-16	67%	Date of Birth: 1977-07-16
Mother's Name: يلع ن سح دمحم	100%	Mother's Name: يلع ن سح دمحم

Photo

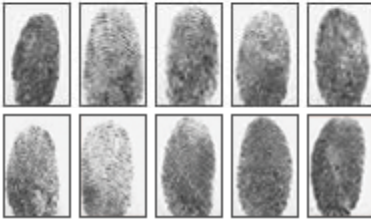


Photo

Duplication: 99.0%




Fingerprints



Fingerprints

Duplication 98%

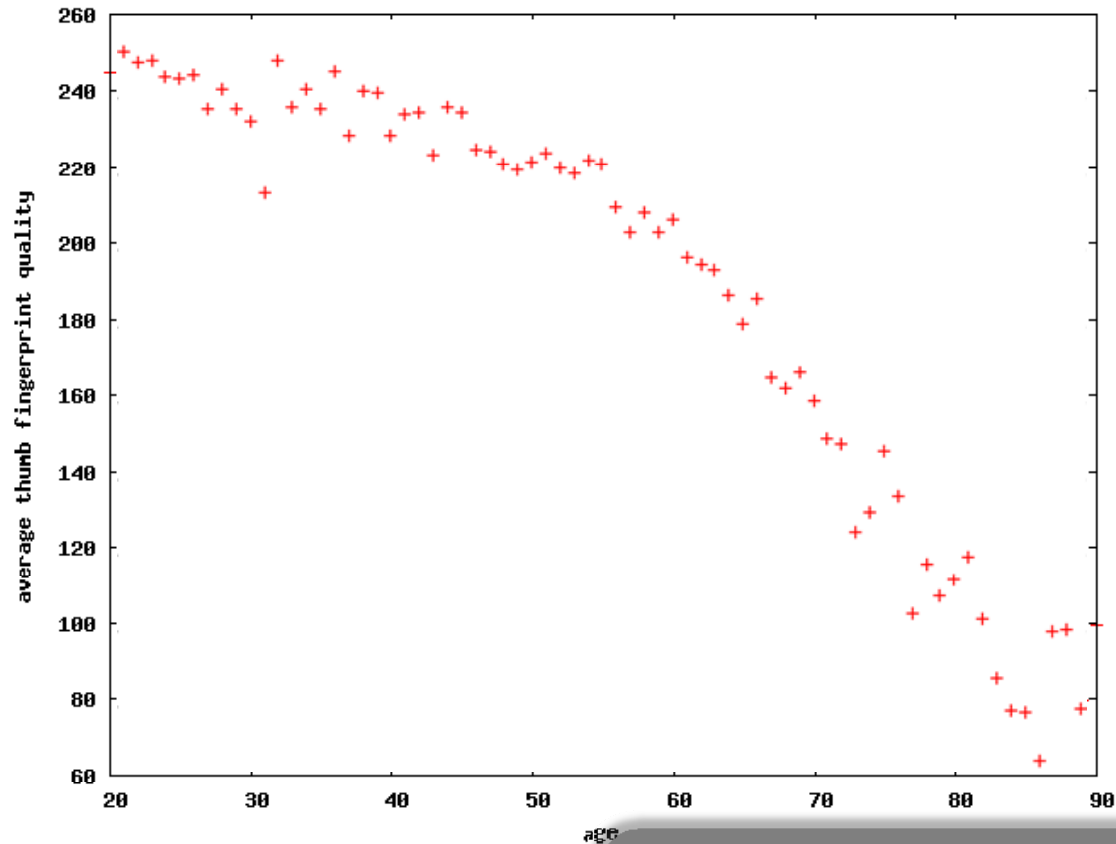


Solution
Use ELISE ID's ability to explain reasons for a match score to communicate to analysts in the field.

Challenge: Issuance process generated bad data

- ❖ “Dirty” data, missing data, inaccurate enrollments created several process issues at verification for Point of Distribution.
- ❖ High False Reject Rate at point of distribution create culture of by-passing biometric check.
- ❖ High False Reject Rate can create a volatile situation at Point of Distribution.

Challenge: Issuance process generated bad data



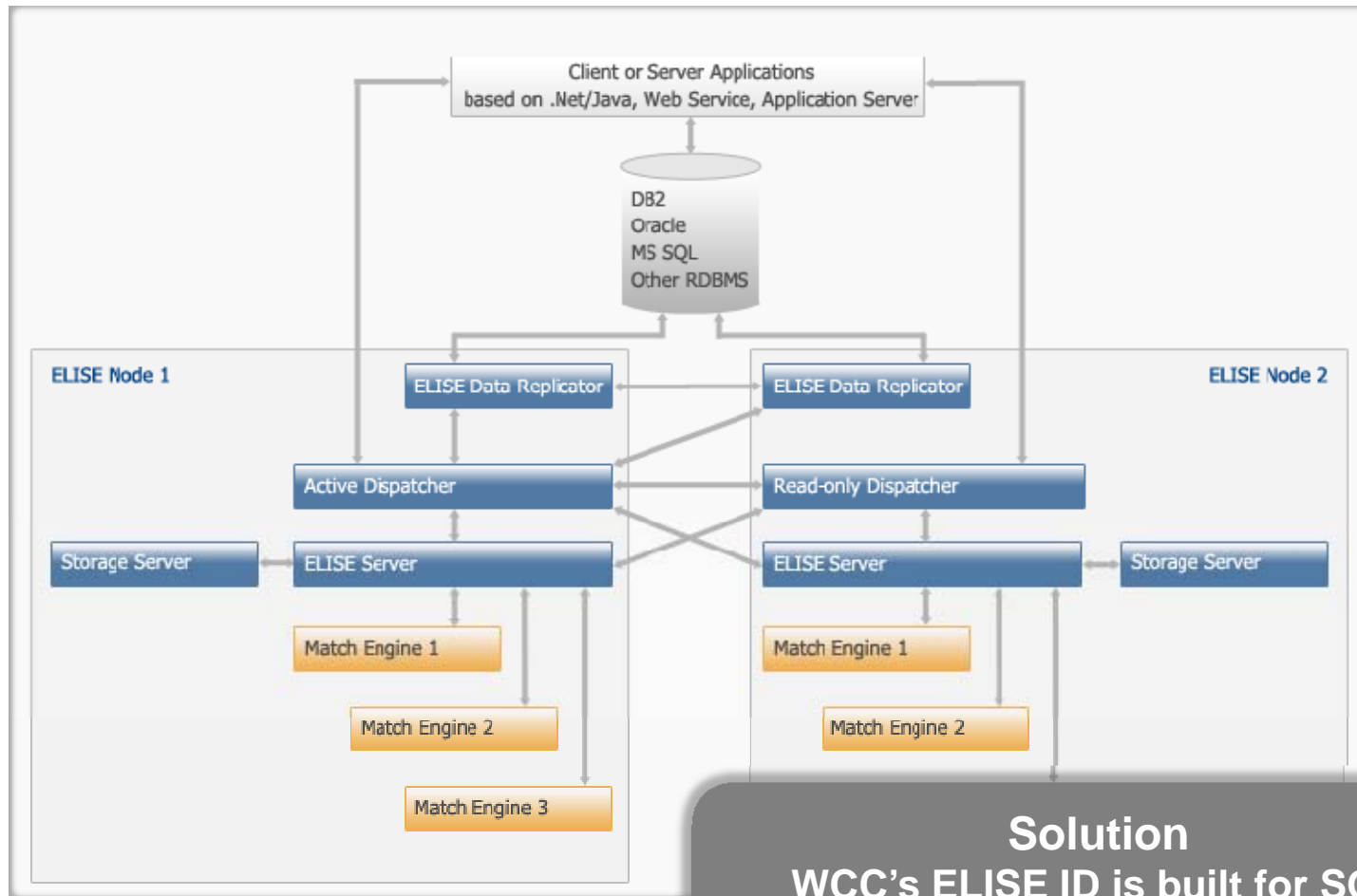
Solution

WCC's Analytics provides feedback to improve the enrollment and issuance process

Challenge: Entitlement Program to be Basis for Additional Programs

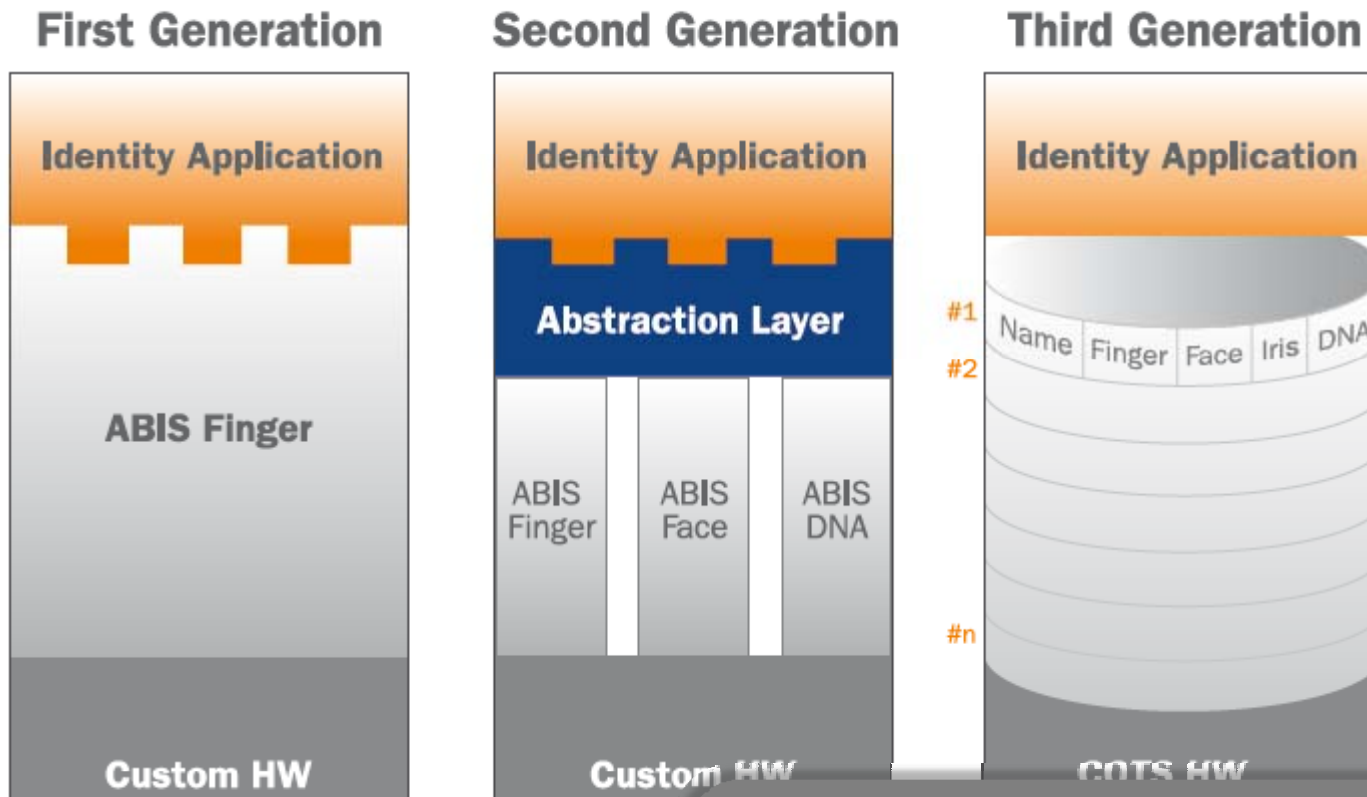
- ❖ Infrastructure used for entitlement program will be expanded to incorporate other national programs:
 - Driver's License Issuance
 - Motor Vehicle Registrations
 - Elections
 - Etc.
- ❖ System needs to scale and biometric technology needs to be future proof.

Challenge: Entitlement Program to be Basis for Additional Programs



Solution
WCC's ELISE ID is built for SOA environment and can scale horizontally and vertically

Challenge: Entitlement Program to be Basis for Additional Programs



Solution

Create interoperability at the algorithm level for easy integration of new biometric modalities and algorithm upgrades

Real World De-duplication - Lessons Learned

- ❖ All data is important; quality data is very important
- ❖ Fusion of biometrics with fuzzy search on biographic data improved de-duplication accuracy and performance
- ❖ Fusion methodologies optimize cost vs. speed vs. accuracy
- ❖ Analytics improve understanding and system performance
- ❖ System needs to be future proof to accommodate growth in size as well as technology improvements

Thank You

ELISE ID
The Platform for Multi-Modal Fusion

For more information

Tarvinder Sembhi
VP Sales, Identity Market

tsembhi@wcc-group.com

www.wcc-group.com

