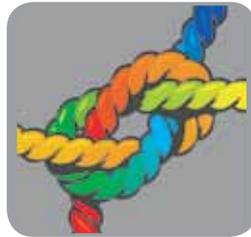


# White Paper

## Identity De-Duplication & Uniqueness Check Solution



**elise**



**WCC is a global leader in search and match technology with a market leading platform for identification that optimizes accuracy by combining biographic and contextual data with multi-modal biometric fusion. WCC's ELISE system offers unsurpassed capabilities to perform identifications using third party biometric algorithms, as well as built-in fuzzy logic algorithms and weighted criteria, very rapidly across very large data populations with extremely high uptime.**

# Identity De-Duplication & Uniqueness Check Solution

## Introduction

In the 21st century's global economy, identification documents have taken on an importance unimagined when they were first introduced. From driver's licenses to visas to passports and even to a company ID cards, these documents augment our traditional systems of personal trust to allow vast numbers of otherwise anonymous individuals to be known, trusted, and granted the privileges they are due: entrance to a corporate facility, airport, or military base, the right to drive a car, a bus, or airplane, the right to cross an international border for business and/or pleasure, the right to vote, receive healthcare, unemployment benefits or to receive asylum in a new country.

## Quality of Data

But with this added importance comes the added burden of ensuring that these documents accurately reflect the identity the person claiming them. Errors in data entry and bad actors can both reduce the usefulness of identity documents by undermining their credibility. Whether by innocent mistake or by fraudulent act, the issuance of multiple identity documents showing different identities for the same person introduces unnecessary risk, one that is often compounded when an initial document is used as a breeder document for other documents.

**Errors in data entry and bad actors can both reduce the usefulness of identity documents by undermining their credibility.**

If you are responsible for issuing identity documents – driver's license, visas, passports, corporate ID cards – you have undoubtedly built a database of identities. It may even be that a new enrollment to your document system does a lookup on a new enrollee to see if they are already in the database. Unfortunately, these are often simplistic searches that may miss data entry errors or bad actors. In fact, your database may already contain records that are actually duplicates without your knowledge.

## Cleaning Data and Preventing Duplicates

The WCC Identity De-Duplication and Uniqueness Check Solution can help establish and maintain the quality of your identity data by cleaning existing data and by preventing the addition of new duplicates.

With WCC's ELISE ID, errors in existing data, whether intentional or just accumulated mistakes, can be removed, and fraud or new data entry errors prevented. The de-duplication process aims at obtaining a uniform, complete and reliable database that consolidates as a single identity for the same user of multiple data sources as example passports, registrations, visas, naturalization records or other documents.

Along the same lines, identity information from multiple source databases with information on passports, visas, permits, criminal history etc... holding varying degrees of biometric data (fingerprints, photos) and biographic data can be consolidated into a single master identity record.

# Identity De-Duplication & Uniqueness Check Solution

ELISE ID uses a combination of fuzzy logic for typos, name changes, transpositions, and a variety of other common data entry errors with extensive name databases to cope with nicknames, name aliases and transliterations to find duplicate identities. These same capabilities can also easily spot fraudulent enrollments and help keep the database free of bad actors.

## Data Analytics

Many identity properties are somewhat correlated. For example, from a picture or a first name, often a gender can be deduced. Even last names sometimes may disclose a gender, like in Eastern Europe last names of females have an OVA-ending. Such analytics may suggest anomalies in the data.

Biometrics are checked for quality levels, like NFIQ for fingerprints, and ISO 19794-5 for face images. Images of insufficient quality are considered of lesser relevance in an identity check.

From experience we know that certain biometrics, like fingerprints, can be difficult to capture from young children, but also image quality deteriorates with age.

With enrollments of people above fifty we see a significant deterioration of fingerprint quality, suggesting possibly a different enrollment process for elderly people, like the use of a high(er) quality scanner.

In earlier projects we found that the proficiency of the operator, in case of supervised enrollment, does have a measurable effect on data quality and hence the relevance to the identification process.

This insight can be used to detect possible operator fraud.

The WCC Identity De-Duplication and Uniqueness Check solution also produces insight into the number of suspected duplicates per operator, age group, gender, location etc.

And finally the solution returns an exhaustive lists of all identities unusable for de-duplication, including the reasons why.

# Identity De-Duplication & Uniqueness Check Solution

## COTS Solution

The WCC Identity De-Duplication and Uniqueness Check Solution comes pre-packaged and ready to connect to your existing system, consisting of:

- ❖ De-Duplication application
- ❖ Uniqueness Check function
- ❖ Software license for ELISE ID Smart Search & Match Software
- ❖ Software connectors to your existing identity/credential database and enrollment system
- ❖ Support from WCC's consulting experts to ensure your system is configured correctly for optimal performance



The WCC De-Duplication Solution easily connects to an existing system or can be accessed as service via a secure network connection for easy integration.

# Identity De-Duplication & Uniqueness Check Solution

## Right Sizing

The WCC Identity De-Duplication and Uniqueness Check Solution comes in a single server configuration that is perfect for most corporate or smaller government needs. While somewhat dependent on the identity data you maintain, this system will handle a database with one million identity records and a daily enrollment load of fifty thousand uniqueness checks. When doing biographic check only, the system can handle 40 million identities with a daily load of 250,000 checks.

## Custom Sizing

Based on extensive experience that includes a variety of ELISE ID deployments, at US state-level (eg. DMV), at US national-level (eg. a national health care provider), and international nation-wide deployments (eg. a mid-east country-wide deployment), WCC can also offer custom configurations for handling very large databases with high usage.

For mission critical deployments, ELISE ID can be configured for scalability and resilience, offering extensive fail-over redundancy for maximum uptime and reliability.

## Identity Check as a Service

If you'd prefer to avoid management of infrastructures, you'll find the WCC Identity De-Duplication and Uniqueness Check Service a better alternative. By establishing secure communications between your master data and our data center, the service offers a great way to pay-as-you-go for de-duplicating legacy data or preventing duplicate enrollments – or both!

## More Information

For more information, white papers, and demos, please visit our website at [www.wcc-group.com](http://www.wcc-group.com) or send an e-mail to [info@wcc-group.com](mailto:info@wcc-group.com).

# Product Specification



---

**Speed**

- Sub-second response times for biographic matching.
- Biometric matching speed dependent on third party algorithm(s)

---

**Scalability**

Handles billions of records

Scales from running on a laptop to multiple, geographically dispersed datacenters, to cloud deployments

---

**Hardware and OS requirements**

- Intel or AMD based servers
- Microsoft Windows or Red Hat, SuSE, or CentOS Linux

---

**Biometric standards**

Supports BioAPI, CBEFF, PIV, ANSI INCITS 378 and 381, ISO 19794-4 and 19794-5. Other data exchange formats can be easily added through plug-in system

---

**SDK**

- Java or .NET
- SOA support
- XML support

---

**Connectors**

Data synchronization through ETL-tool with configurable connector for all relational database systems or file systems

---

**Client integration**

Supports clients applications on any architecture or OS

---

**Reliability**

- High availability, fully redundant architecture with no single point of failure.
- Proven reliability - many nation-wide, mission-critical deployments over the past ten years

---

**Architecture**

Massively parallel in-memory processing for high speed/high volume scenarios

---

**Customization**

- Fuzzy logic matching on text, and numeric data as well as setting weights for criteria importance
- Fully configurable fusion engine

---

**Algorithms**

- Proprietary text and numeric algorithms
- Incorporates third-party biometric algorithms, as well as matchers for other data types
- Score-level fusion algorithms
- Easy algorithm integration and replacement
- Geo-spatial analysis

---

**Text matching**

- Full set of query operators
- Graphic query interpretation
- Weighting by relevance groups
- Exact, phonetic, OCR, and typo matching
- Highlighting of matches

---

**Transparency**

Detail match shows how match score is calculated

---

**Management tools**

Tools for logging, monitoring, and data analysis

WCC Services B.V.  
Zonnebaan 19  
Utrecht ■ 3542 EA ■ The Netherlands  
Tel. +31 30 750 3200  
Fax +31 30 750 3299  
info@wcc-group.com  
[www.wcc-group.com](http://www.wcc-group.com)

Washington DC Office:  
Two Fountain Square  
Reston Town Center  
11921 Freedom Drive ■ Suite 550  
Reston ■ VA 20190  
Tel. +1 888 922 9224  
info@wcc-group.com