FEDERATED SEARCH OPTIONS FOR DATA ACCESS & SECURITY

IDENTITY MANAGEMENT IN REMOTE, DISTRIBUTED & DISPARATE DATABASES

WHITE PAPER

Introduction

In an ideal world, to search is to find. Information is freely accessible, even in remote, disparate, or distributed databases. Search results are relevant; legacy and new data are merged without conflict; additional external systems complement existing systems. Big data is no big deal. And of course, data security is guaranteed.
In reality, well... things are different.

When organizations share identity data, they do so selectively and within strict limits. They want their data protected, their privacy prioritized, and their boundaries respected. But full access to data is crucial in person identification: to protect citizens, prevent fraud, and simply to minimize costs.

This paper explores solutions for effective and secure searching in remote, distributed, and disparate systems.

What are the challenges?

Data access and data security are complex, especially when it comes to person identification. Different organizations have different architectures and database setups which they query through different search methodologies based on different policies and legislations. Each architecture or system has its own rules and requirements.

Challenges arise in four domains:

1. Security

Data security is the topmost concern when sharing and transferring data. In person identification, this often translates to privacy as well. Federated systems must also guard data ownership.

2. Scale

We live in the era of big data. The world produces data at staggering speed, volume, and complexity levels. Searching in huge databases is often expensive, time-consuming, and error-prone.

3. Data quality

Individual databases widely differ in design, format, fields, and data quality. They interpret search terms and results differently. They can contain both new and legacy data, both accidental and fraudulent errors.

4. Search result quality

The final challenge is making sure that search results are transparent, reliable, and relevant.

What is the objective?

To end users, it does not matter whether their query is sent to a single database or to many. All they want is optimal results. For the organizations that use remote, disparate, and distributed databases, there is an additional concern: security. Efficiency and ease of use must never compromise security. In short, the objective is to reconcile the need for data access with the need for data security.

FEDERATED SEARCH OPTIONS
FOR DATA ACCESS & SECURITY

What are the solutions?

Federated search solutions must address challenges in all four of the aforementioned domains.

1. Security

To protect data security, you need control over who is granted access to data and how search results are presented to different users. Role-based and rule-based access help ensure that end users only receive the information that they are allowed to process.

WCC's ELISE provides this capability: its match behavior can be customized to enforce access levels. In addition, users can specify which fields to search, the weight of those fields, and the algorithms used to use, which provides greater control of the match results compared to classical search. ELISE can return search results containing only the match scores and unique identifiers (called keys) of the records. This effectively separates the security protocols from the search or the data, and allows merging of the different result sets while protecting the privacy of the records. If more details about a record from the result set are needed, ELISE can use the unique identifiers and separate access control protocols.

Encryption is a good additional security measure, which ELISE provides by default. Encrypted communication channels and configured firewalls protect both the data storage and the communication layer. Back-end software like ELISE can also be encapsulated by system or application security, which adds yet another layer of security.

2. Scale

In mission-critical and high-volume deployments, software must run efficiently in multiple data centers. This ensures that even if one crashes, the software continues to run in the others. Data spread across different data centers must be synchronized and connected no matter how high the data volume. Finally, growing data volumes, increased loads, and higher match speeds must be easy to accommodate. WCC's ELISE is high-availability software with excellent scalability and redundant architecture. It allows for easy addition of favorable price/performance servers to deal with increasing

volume and speed demands. All this makes it ideally suited for large-scale deployments.

3. Data quality

The larger the data set, the higher the chance that data contains errors and inconsistencies. Because data quality issues arise for a variety of reasons, a good federated solution must cover a wide range of capabilities for dealing with each.

WCC's ELISE platform offers:
> **language independence**
  The ability to ingest data "as is": ELISE matches in any language and character set.
> **semantic search**
  Improved search accuracy through understanding the intention and contextual meaning of terms.
> **a variety of algorithms**
  ELISE easily deals with omissions, typos, transpositions, multi-cultural name variations, and more.

Inconsistent data formats and fields are another common issue when searching or merging data from disparate databases. For example, some databases store name data in two separate fields while others use a single field. ELISE uses similarity algorithms to handle missing, additional, or reversed name components and determines the impact on the overall match probability.

4. Search result quality

Even if the other challenges are met, search result quality determines whether you will find what you are looking for or not. Search results must be crystal clear – both transparent and unambiguous. WCC's ELISE delivers meaningful search results even when working with questionable data quality, huge volumes of data, and massive numbers of users. The results are highly relevant and fine-grained thanks to its unique fuzzy logic capability. First, ELISE normalizes and cleanses data. Then, it uses de-duplication to produce uniform, complete, and reliable data entries. Its real-time matching capabilities can find similarities among words, names, concepts, and numbers, with complete control over rules, weights, value ranges, and other factors.

FEDERATED SEARCH OPTIONS
FOR DATA ACCESS & SECURITY

Deployment Examples and Use Cases

Organizations dealing with remote, disparate, or distributed databases need versatile and flexible solutions that simplify and optimize the search and match process. WCC's ELISE is built for flexibility: it fits seamlessly into any number of setups. The platform is specifically purposed for identity data and applications. Therefore, typical ELISE installations have smaller footprints and can be deployed with more agility than larger MDM solutions that provide comparable capabilities. Below are five examples of how ELISE provides identity data management functionality in remote and federated environments.

1. Classic Federated Search

Step 1: The central application sends a query to the ELISE federation component.
Step 2: The federation component then queries multiple databases.
Step 3: Local versions of ELISE search in their respective local database.
Step 4: Scores and unique identifiers are returned to the ELISE federation component.
Step 5: The federation component combines the scores according to customizable fusion algorithms and returns the customizable threshold to the central application.
Step 6: There are protocols in the application layer for providing details about matches that meet certain criteria.
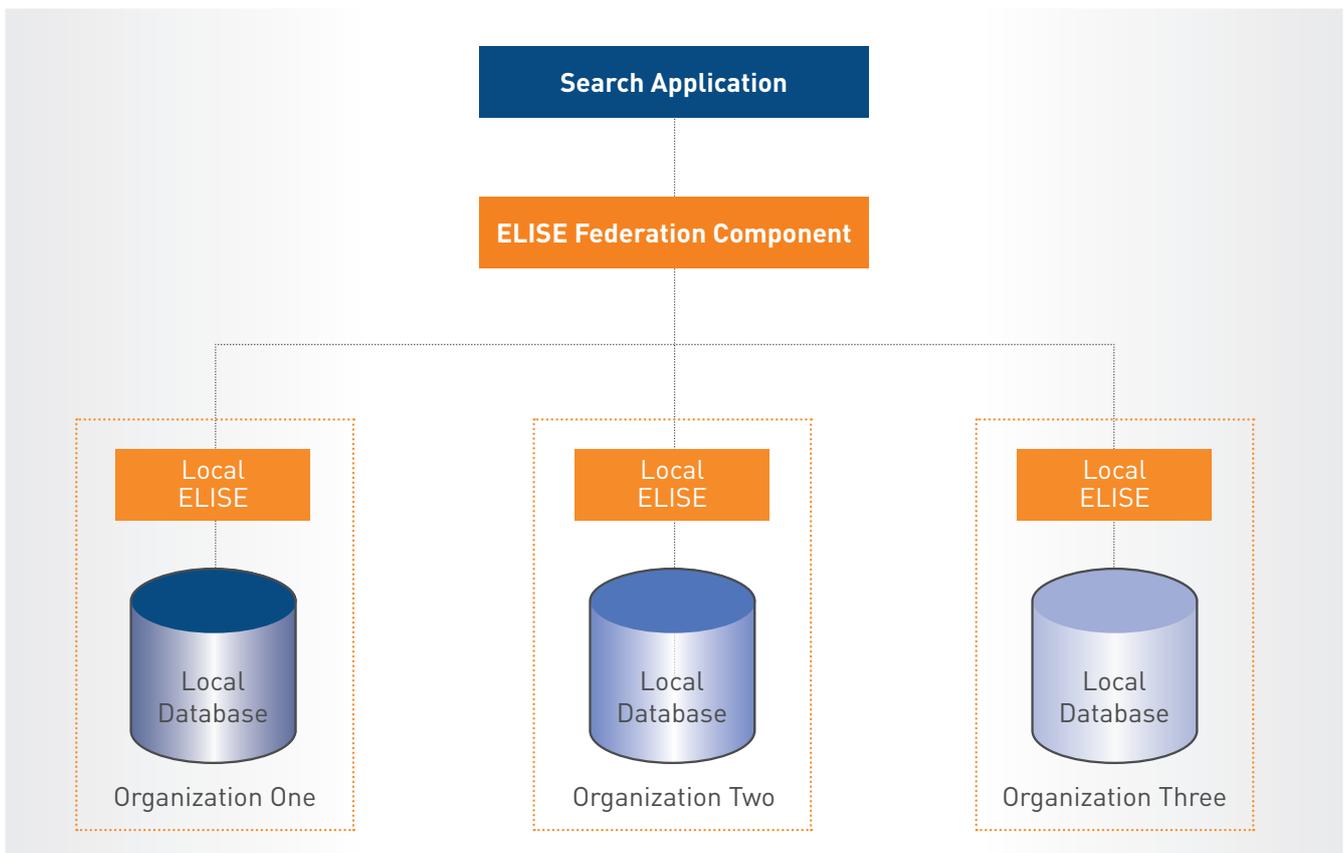

Figure 1: Classic federated search

2. Remote Search
Civil Identity Use Case: UNHCR

In this case, the goal was to biometrically register large numbers of refugees efficiently and fast. Accurate records on all refugees were needed to ensure fair distribution of daily rations. Because the refugee camps were in remote locations without amenities, registration was done locally. Once a day, each local database was synchronized with the central database in Geneva.

Both the central database and every local system have an installation of ELISE. The central ELISE database holds all the organization's records. Each local subset contains data from a specific location.

First, new data is entered into the local system and de-duplicated by the local ELISE. Then, within a fixed timeframe, the new data is collated and sent to Central ELISE. Here, de-duplication of the new data in relation to the entire dataset takes place. Central ELISE then sends the cleansed data (and information on any discrepancies found) back to the subsets. If Subsets A and B share coverage over an area or region, both send all data to the Central ELISE. After de-duplication, Central ELISE sends the cleansed data (and info on any discrepancies found) to both subsets.
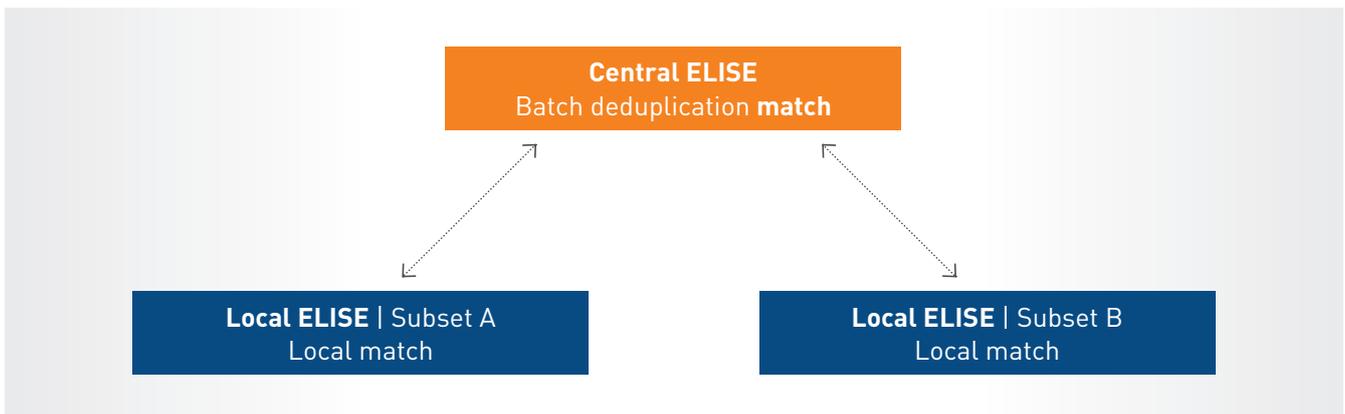

Figure 2: Remote database search

3. Migration and Search across Disparate Databases
   Law Enforcement Use Case: Finnish Police

The goal was to modernize a legacy system comprising many different, disconnected databases containing potentially relevant information.

In this set up, there is one main ELISE. Data Set 1 contains data from a legacy database intended to be made inactive. ELISE cleanses and normalizes this data.

Data Set 2 contains data from an external database. ELISE provides a unified search methodology in spite of differences in structure, format, size, and so on.

Data Set 3 contains data from a legacy database intended to remain active and accessible. This data is also cleansed and normalized by ELISE.
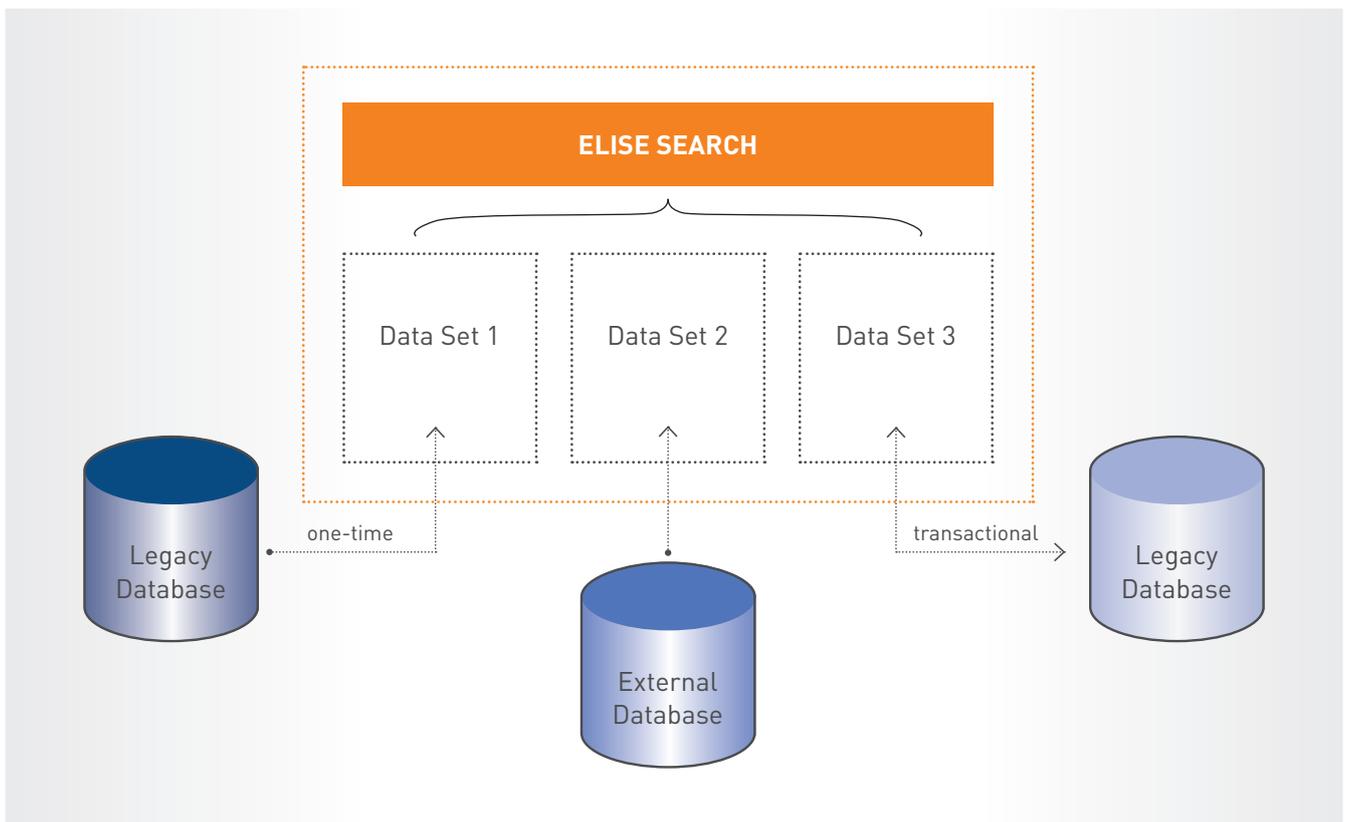

Figure 3: Migration and search across disparate databases

4. Golden Record Creation

A golden record is a single, accurate, complete version of all the data pertaining to an individual. To create a person's golden record, the best and most accurate data about a person must be distilled from all available databases. The ELISE Platform uses de-duplication, multi-cultural name matching, and other identity resolution technologies to do this. ELISE can also compare and rank data from disparate databases. Even if there is no shared key (such as a social security number) in the disparate databases, ELISE will still manage to identify which records pertain to the individual.
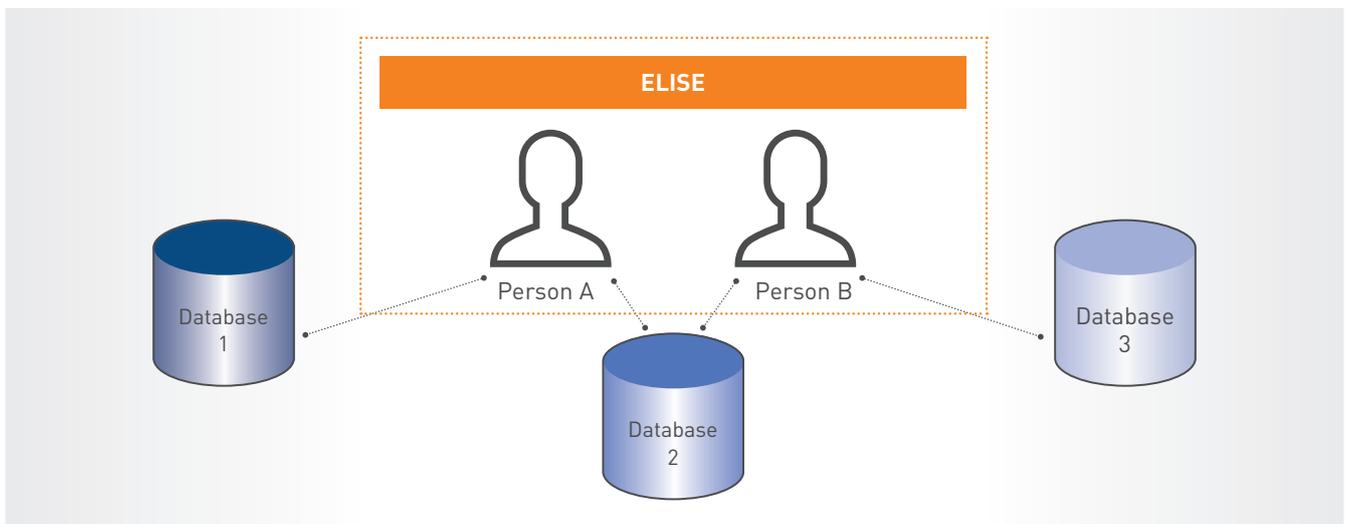

Figure 4: Golden record creation

5. Person-Centric View

In this setup, different data sets on one individual can be created by linking them. This scenario becomes federated when additional information about the individual is searched in an unlinked external database.
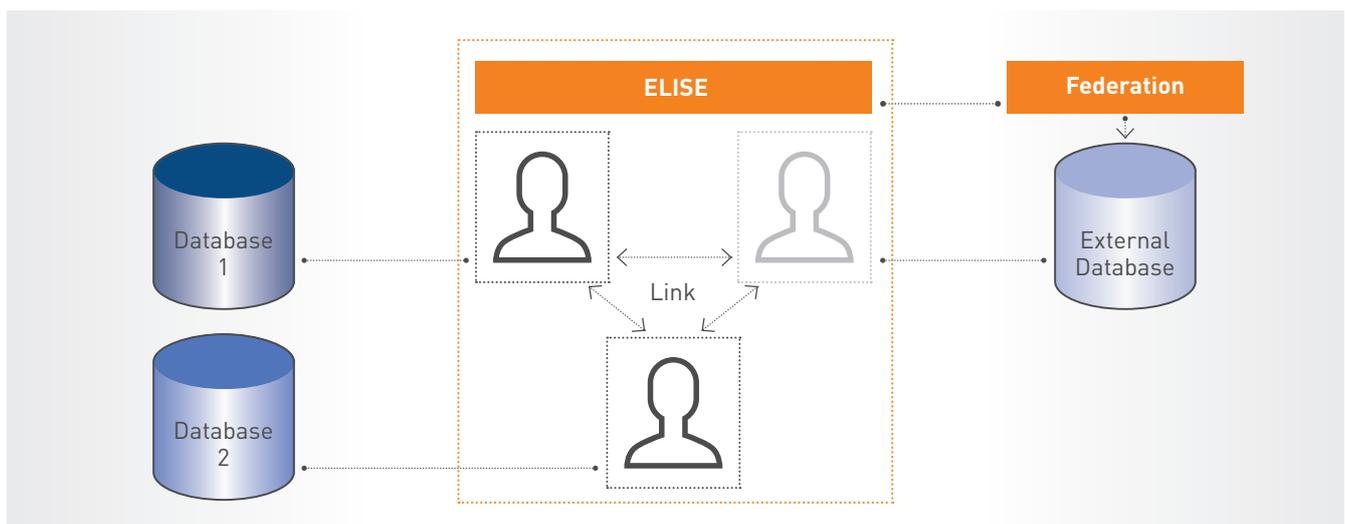

Figure 5: Person-centric view

Accessing data without compromising security

It is time to stop wasting time and resources on ineffective federated search solutions. Moreover, it is time to safeguard the security, ownership, and privacy of identity data.

WCC makes effective and efficient searching possible in any type of enterprise architecture. Bringing together disparate data under ELISE allows for successful matching across the entire data domain, but at the same time enables the original owner of the individual database to maintain control of their data. Merging legacy and new databases becomes effortless. Large repositories, even in scattered data centers, yield transparent results in ranked order and in real-time. Searching has turned into finding.

WCC's ELISE platform is used in various large-scale identity deployments that required remote and or federated search capabilities. It offers a range of applications, including:

> **Multi-modal matching against any identity data: biometrics, biographics, metadata, transactions**
> **De-duplication and uniqueness check**
> **Data cleansing**
> **Migration**
> **Risk assessment and fraud screening**

The way forward

WCC Smart Search & Match offers significant enhancements to both typical and federated search. The ELISE Platform provides a cost-effective and powerful solution for managing identity data in remote and federated environments. It offers de-duplication, golden record creation, fraud detection, and award-winning multi-cultural name matching to ensure quality results.

WCC helps organizations accomplish their goals with fewer resources. WCC prioritizes data security while gathering and collating reliable data from multiple data sources. The benefits are enjoyed not just by organizations, agencies, and departments, but by the individual end users as well.

Contact WCC to learn more about flexible search options and setups for your organization's current and future needs. Review the improvements on your system and processes, and the resources now freed up. Reap the rewards.

About WCC

WCC Smart Search & Match is the world's leading supplier of search and match software solutions and services. Founded in 1996, WCC focuses on two specific solution areas: Identity matching and Employment matching. Its ELISE software platform excels in these areas because it searches and matches data in a unique way that yields more meaningful results than any other software. ELISE is designed to search through vast amounts of data from various sources and return relevant results in under a second. It will search and match data in almost any form, using advanced algorithms, contextual knowledge, and other proprietary expertise. The data can be exact or inexact, structured or unstructured, private or public, and combine multiple modalities, both biographic and biometric.

WCC's long-term experience in developing and supporting employment and identity matching solutions makes it an expert in these fields. WCC understands the business of its customers and knows how to optimize the effectiveness of searching and matching.

WCC's primary customers are government organizations and large enterprises worldwide. In Identity matching, WCC supplies solutions for border management, justice & public safety, and civil identity. In Employment matching, WCC supplies solutions to public employment services, staffing companies and large enterprises for their corporate HR.

WCC is headquartered in Utrecht, the Netherlands, and also has offices in the USA.

WCC Smart Search & Match
Zonnebaan 19
3542 EA Utrecht
The Netherlands
T: +31 (0)30 750 32 00

info@wcc-group.com
www.wcc-group.com

WCC Services US Inc.
228 Hamilton Avenue
Suite 300, Palo Alto
CA 94301, USA
T: +1-888-922 9224

SMART SEARCH & MATCH