Searching for Names with ENS (IBM InfoSphere Global Name Management Enterprise Name Search) Version 6.0

Prepared by: Richard Strangfeld

Edition

This edition applies to Version 6.0 IBM InfoSphere Global Name Management (product number 5724-Q20) and to all subsequent releases and modifications until otherwise indicated in new editions.

1 Table of Contents

1 Introduction

IBM InfoSphere Global Name Management Enterprise Name Search, or ENS, is software that allows sophisticated name searching based on Global Name Management NameWorks to be done at the enterprise level. It uses proven IBM middleware and database components to:

- Perform efficient NameWorks-based searches in large lists of names.
- Distribute those names across multiple servers to allow larger lists, higher performance, and reliability through redundancy.
- Make it easy to manage those name lists and servers.

This document discusses using ENS to search for names. It assumes that someone has already installed and set up an ENS system, as described in a separate document titled "Setting up and Managing ENS".

1.1 Overview – ENS System Architecture

An ENS system consists of a number of servers with "searcher" components. Each searcher holds an instance of NameWorks with some collection of names. Working together, these searchers provide searching in a potentially very large combined list of names, and can support numerous users searching at one time. Another server-side component called a "dispatcher" coordinates the work of the searchers.

The ENS user who wants to search for names does so in a search client program. This can be the browser-based search client included in ENS, or some custom search program specific to the user's organization.

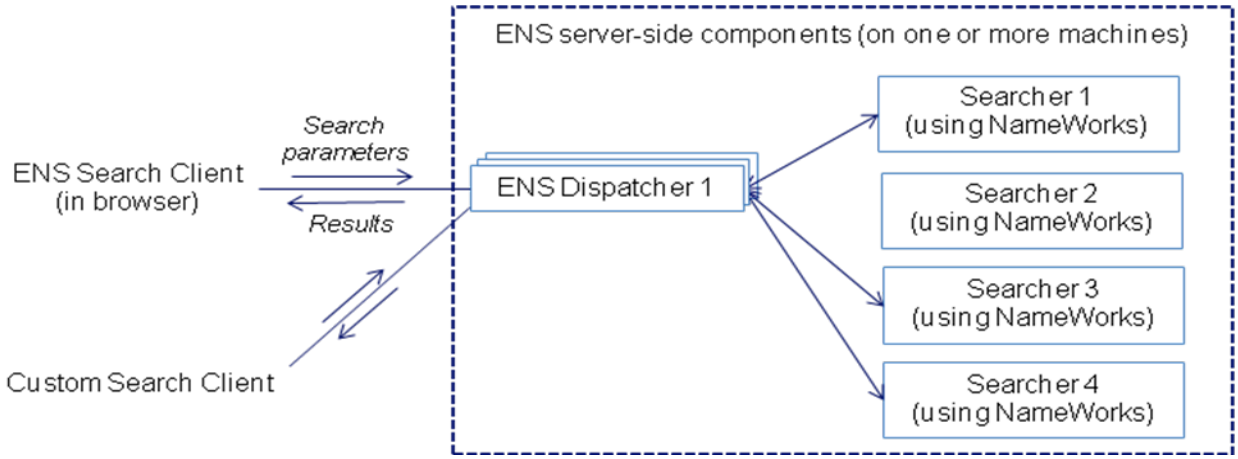

Figure 1 - Simplified client view of ENS Architecture

In either case, the client program makes web service calls to perform searches. At the back end, a dispatcher component handles that call and delegates the request to some set of searchers, then collects their responses and returns results to the client.

On the server side, the ENS system may consist of a single dispatcher and searcher on one machine, or of numerous dispatchers and searchers on many machines. This can vary widely based on the requirements for scale (number of names and number of clients) and performance. Servers and components can be added and removed as needed, even in a running system.

From the client's point of view, it looks the same regardless. The fact that the server side has one machine or many is not important, or even noticeable, on the client. The client just makes search requests to a dispatcher and gets name results.

1.2 Loading Names into ENS

Names may be added to ENS in two different ways: with the NameLoader program, or with the addName web service.

The NameLoader program is described in the "Setting Up and Managing ENS" document. The addName web service is described in the "Web Service API for ENS" document.

2 The ENS Search Web Client

ENS comes with a browser-based client program for searching for names. This client calls ENS web services to perform searches. Organizations using ENS can run searches by means of this ENS web client, or they can write their own search client program that uses the same web services.

The ENS search client has primarily been tested in Firefox (ESR version 45) and Internet Explorer (version 11), though it also runs in Google Chrome.

Figure 2 shows the ENS search web client. It has a search form at the top and a main results grid at the bottom. We'll discuss these in the following sections.

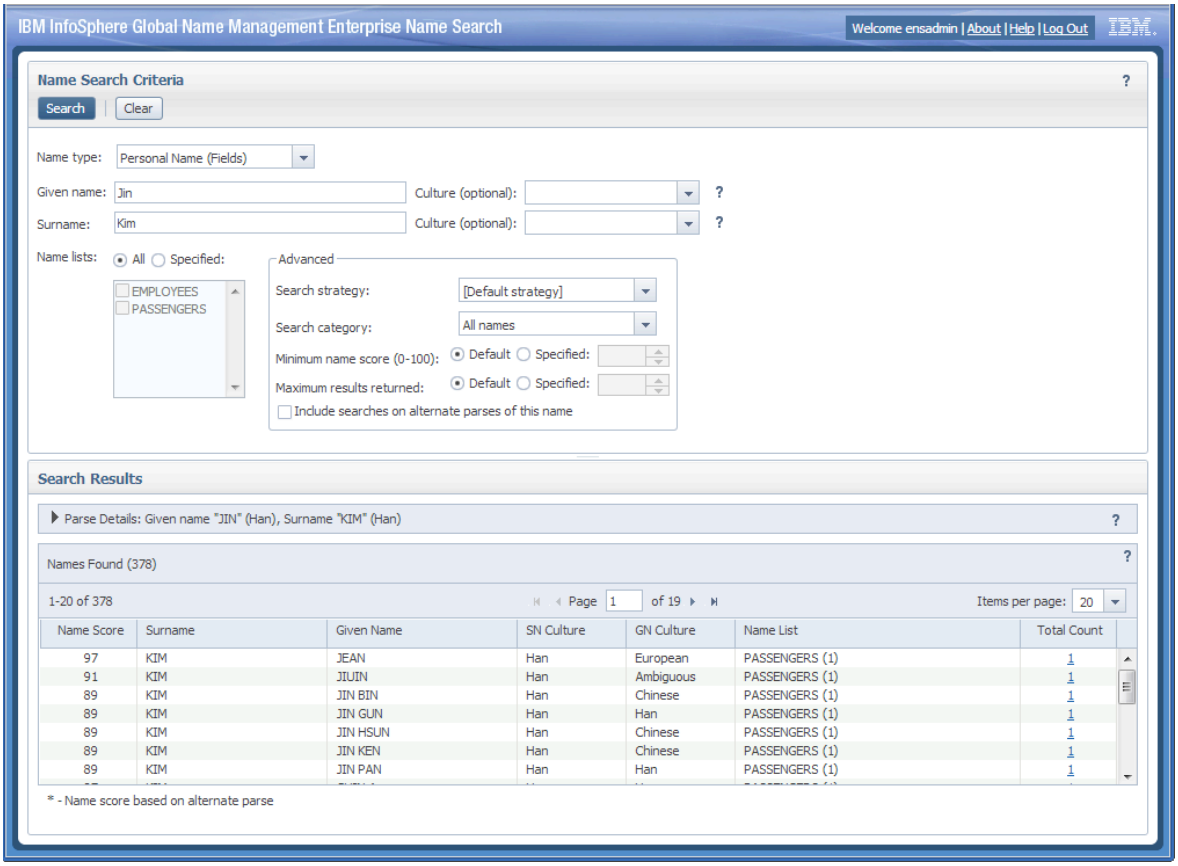

Figure 2 - ENS Search Web Client

2.1 The Search Form

Figure 3 shows the search form in more detail. The circled numbers in this figure are referenced in the text which follows.

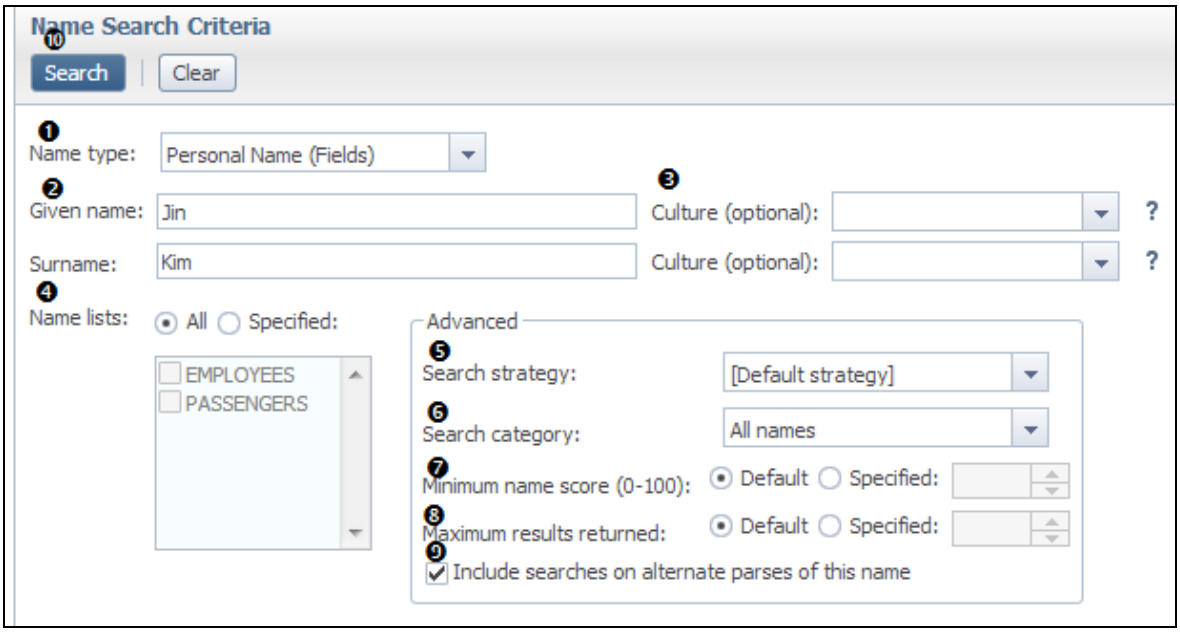

Figure 3 - Search form (showing a personal name with given and surname fields)

2.1.1 Name type dropdown

❶ This lets you choose how to enter the name you're looking for (sometimes called the "query name"). You can enter this name:

- As a personal name provided as separate given name and surname, with optional cultures for each part.
- As a personal name provided as a single field, with a single optional culture.
- As an organization name in a single field, with an optional culture.
- As a name with unspecified category, in a single field.

2.1.2 Name fields

❷ With the "Personal name (Fields)" name type selected, we see separate fields for the given name and surname. These change to a single field for "Full name", "Organization name", or just "Name", depending on the selected name type.

2.1.3 Optional culture dropdowns

❸ By default, NameWorks does its own analysis to guess the culture of a name. If you happen to know the culture of the name that you're supplying, you can optionally override NameWorks' choice in the dropdowns here. NameWorks uses the culture to improve the outcome of the search, by applying culture-specific matching rules.

For a parsed name (separate given name and surname), there are separate dropdowns for given name culture and surname culture. If you select name type "Personal Name (full)" or "Organization Name", a single dropdown lets you select a culture for the name as a whole. If you select "Unspecified", no culture dropdown is provided.

2.1.4 Name lists

❹ An ENS installation can hold one or more name lists. For example, an airline might have one list for passengers, another for employees, another for no-fly persons, etc. When searching, you can choose to consider all name lists or only particular ones. To specify particular name lists, choose the "Specified" radio button, and select the checkboxes for the lists you want.

The checkbox list on this screen shows all of the name lists that are available for you to search. Your access to name lists depends on the permissions given to the user name with which you log in.

Depending on how your system is set up, users might be given a generic "searcher_all_lists" role that allows them to search all name lists, or they might have particular searcher roles (with names like "searcher101") that are associated with particular name lists. This allows fine-grained access control on name lists. Typically, this is managed by an administrator.[1]

If there is only one name list in your system (or only one that you have permission to search), its name is displayed here as read-only text, with no radiobuttons or checkboxes.

2.1.5 Advanced settings

The group box to the right of the name lists has fields for several advanced search settings, described in the following sections.

2.1.6 Search strategy

❺ The NameWorks configuration file in your ENS installation can define one or more search strategies. A search strategy is a named combination of configuration settings that control the parsing and searching behavior of NameWorks. Any strategies defined in your installation's NameWorks configuration file are listed in the dropdown here for selection. The "[Default strategy]" choice that is selected by default uses values selected by GNM, not taken from any strategy, for configuration. These values are culture-sensitive in a way that overrides in defined strategy values are not, so they are often the best choice. See Global Name Management documentation for more information.

2.1.7 Search category

❻ This dropdown lets you choose whether to search for personal names, organization names, or both (all names).

---

[1] For more information on this and other topics related to administration, see the separate "Setting Up and Managing ENS" document.

2.1.8 Minimum name score

❼ Search results only include names with a matching score that is above some minimum name score threshold. By default, this threshold comes either from the selected search strategy or from a default value chosen by NameWorks. You can override the default with your own value by selecting the "Specified" radio button and entering a value in the numeric spinner.

2.1.9 Maximum results returned

❽ This field lets you specify a maximum number of results to be returned. Because ENS combines the search name results of NameWorks searches performed on multiple searchers, potentially derived from different name lists, with a many-to-many mapping between search names and raw source name forms, with deduplication of results from different searchers, and with results listing distinct raw name forms, the effect of a maximum-results specification is hard to predict and may be unexpected. Users are strongly encouraged to use minimum name score instead of a maximum results count as a means of limiting the results from a search.

2.1.10 Checkbox for including searches on alternate parses

❾ When you supply a query name to search for, ENS uses NameWorks to parse and analyze it as described above. If you check this checkbox, ENS will consider more than one parse of the name. If it finds a parse that scores higher (more plausible) than what you provide, it will search for that parse as well as the one you provide.

For example, if you search for the name "Elton John" with this checkbox *unchecked*, the searchers will only look for this name:

1. Given name = "ELTON", surname = "JOHN".

But if you check this checkbox, the NameWorks analysis will decide that "John Elton" is a more plausible parse, so ENS searchers will look for two names:

1. Given name = "ELTON", surname = "JOHN".
2. Given name = "JOHN", surname = "ELTON".

It will then merge the results from the two searches, giving them equal weight.

2.1.11 The Search and Clear buttons

❿ Clicking "Search" causes the search to take place. The client sends a web service request to the dispatcher. The dispatcher uses NameWorks to analyze the query name, then chooses some set of searchers and delegates the search to them. Searchers use NameWorks to perform a sophisticated name search, then use database operations to map the search names found back to source name forms. The dispatcher merges the results from the searchers and returns them to the client.

The results of a search are discussed in the next section.

The "Clear" button clears the name and culture fields in the search form, and clears any displayed results. It does not change the name list selection or the details in the "Advanced" group box. This helps you search for different names without having to re-set your advanced settings each time.

2.2 Search Results: Parse Details

Below the search form is the "Search Results" section. It begins with a "Parse Details" header. This summarizes the parse or parses that NameWorks found for the query name (the name you specified to search for).

By default, the header is closed. Depending on the number of parses found, it looks like Figure 4 or Figure 5.

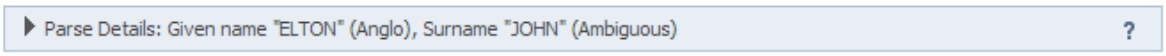

Figure 4 - Parse details header - closed, with a single parse

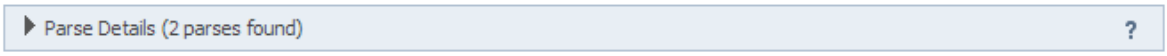

Figure 5 - Parse details header - closed, with multiple parses

Clicking on that header expands it to show a grid detailing the parses found, including cultures. The example in Figure 6 shows the parses found when we specify full personal name "Elton John" and check the "Include searches on alternate parses" checkbox.

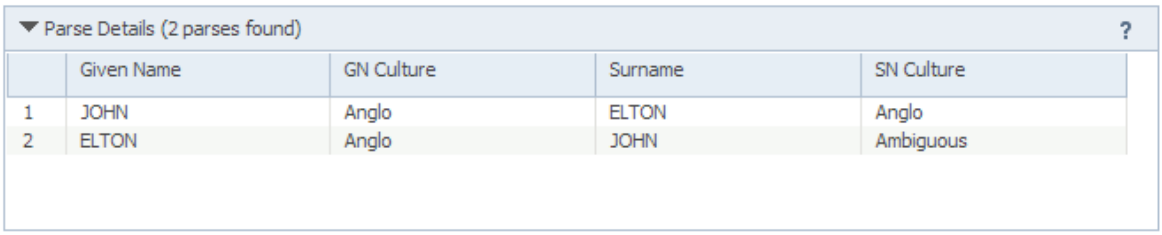

Figure 6 - Parse details header - open

2.3 Search Results: Main Results Grid

The main part of the search results section is the results grid. It lists distinct forms of the source names that were found.[2]

[2] More precisely, the main results grid has a separate row for each distinct combination of category + raw surname/orgname + raw given name + alt parse flag.

This is slightly less granular than the results of the name search web service. In the web service, each result represents a distinct combination of category + raw surname/orgname + raw given name + alt parse flag + surname culture + given name culture + script type. The search GUI merges multiple web service results if they differ only by their cultures and/or script types.

Figure 7 shows the search results grid for a search with several hundred results.

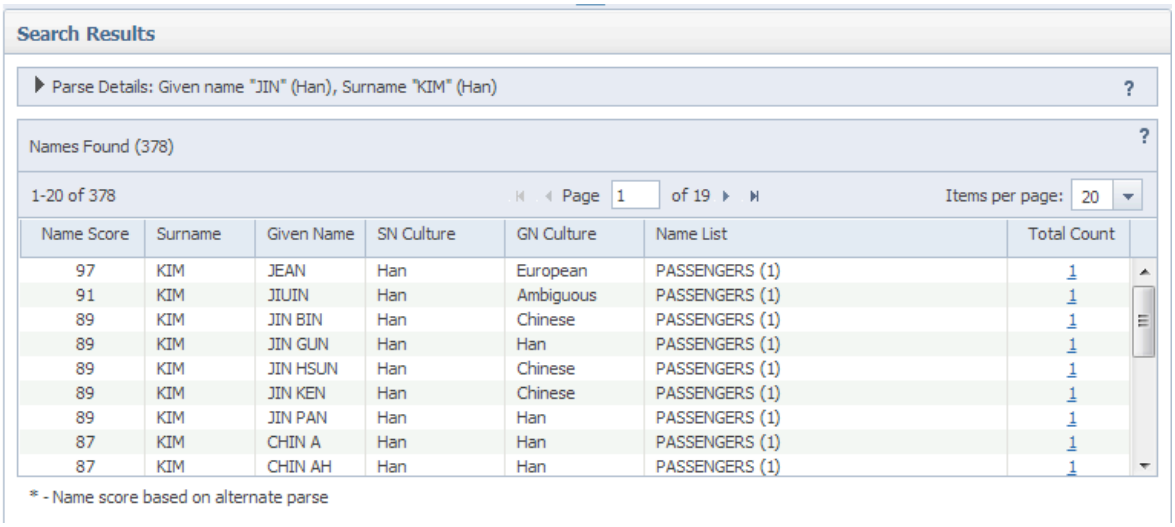

Figure 7 - Main search results grid

2.3.1 Main results grid columns

The columns in this grid are:

Column Name	Example Value	Notes
Name Score	81	Matching score assigned to this search name, where 100 is best. If this result represents a search name that was added to the system as an alternate parse, the name score is marked with an asterisk. Note that this is an alternate parse found on the source name at NameLoader time, not on the query name at search time. If this row represents a grouping of several search results (e.g. by deduplicating results having different cultures), the highest of their scores is shown.
Surname or Organization Name	John	Raw surname or organization name as found in the original source name data. If the original name was provided as a single field full name, that is shown here. The column header varies depending on whether the results include organization names.
Given Name	Tin Hak	Raw given name as found in the original data. This is blank for an organization, or if the original name was provided as a single field.
SN (Surname) culture	Ambiguous	Culture of the surname, either as specified in the original data or as determined by NameWorks when loading the name into the system. If this row groups several search results having different surname cultures, this cell says “Mixed”.
GN (Given name) culture	Chinese	Culture of the given name, either as specified in the original data or as determined by NameWorks when loading the name into the system. This is blank for an

		organization, or if the given name is blank. If this row groups several search results having different given name cultures, this cell says "Mixed".
Name List	PASSENGERS (5) EMPLOYEES (1)	Shows the name lists in which this source name form was found, with a count for each. If more than one name list is shown, this cell shows a summary with a twisty triangle that you can click to see all lines. Only name lists included in the search are listed.
Total Count	6	Total number of instances of this source name form in all name lists searched. In other words, this is the total number of external references to this source name. The number is displayed as a clickable link. Clicking on it opens a Name Details dialog listing the instances found, and showing other details on the name.

Table 1 - Columns in the main results grid

2.3.2 Pagination and sorting

The header on the results grid shows the total number of results found. It also provides controls for pagination. By default, the results grid shows 20 results per page. You can use the pagination arrows or the page-number field to select different pages in the results, and you can adjust the number of items per page.

By default, the grid is sorted by descending name score. Clicking on a column header in the results grid sorts by that column; clicking the same header again sorts the column in the reverse order. If there is a tie between rows in the column you sort on, an implicit secondary sort is done by score, alt-parse flag, raw surname/organization name, and raw given name.

2.4 Search Result Details

If you click on the "Total Count" number for any row in the main results grid, the GUI opens a dialog to display external references and other details about that result.

This is the result of a second web service call made by the web client program. The original search request got back distinct name forms and a count of the number of times they appeared in each name list, but did not include the actual external ids for those instances.

Here, the web client calls the getExternalReferences web service to get those references and other details, and displays this information in a dialog.

2.4.1 Search Result Details– External references tab (personal name)

Figure 8 shows this dialog for a personal name with only one external reference.

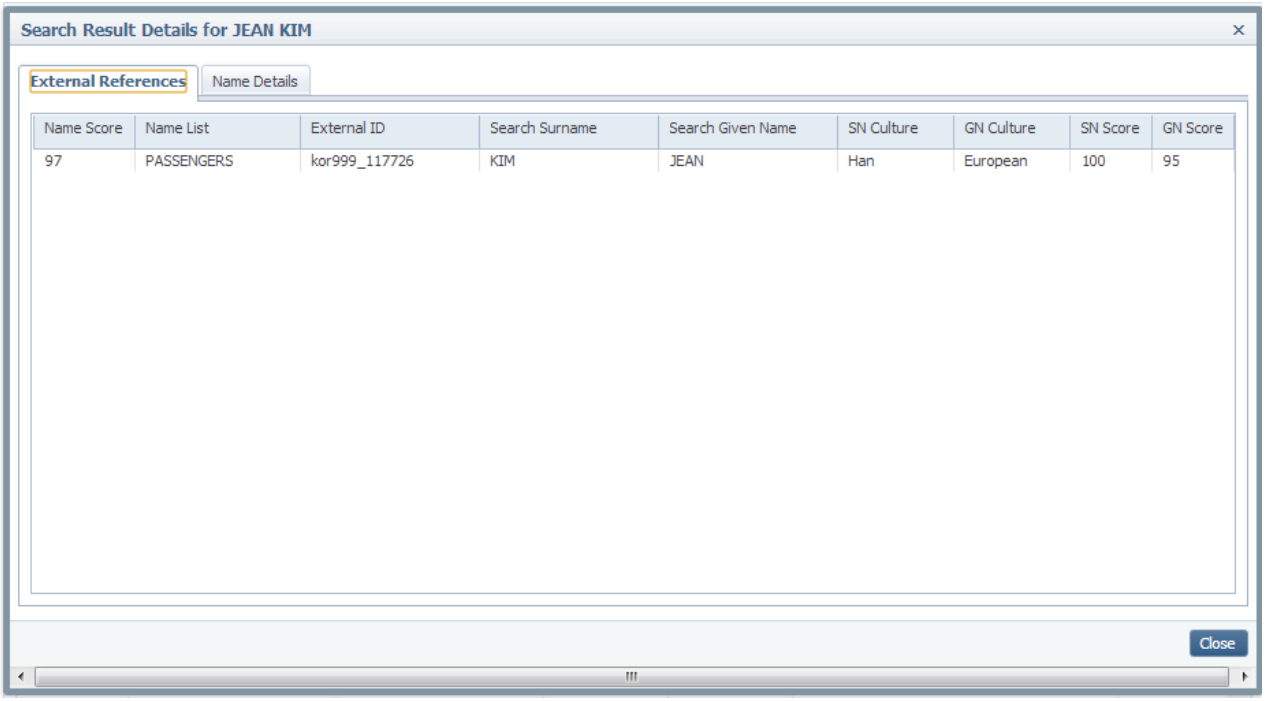

Figure 8 - Details dialog - External References tab – (personal name_

The main tab in this dialog has a grid listing external references, one per row. Alternatively, you can think of this grid as listing source names, raw names originally loaded into ENS. For each one it shows:

Column Name	Example Value	Notes
Name Score	81	Score (for the search name found by NameWorks mapping to this source name). 100 is best.
Name List	PASSENGERS	Name of the name list where this source name (and external reference) is found.
External ID	kor999_117726	ID string provided for this source name when loaded. Typically this is a key into a customer's external database. Sometimes customers also embed some additional data such as date of birth in the ID for quick access.
Search Surname	KIM	Surname (for the search name, as above))
Search Given name	JEAN	Given name (for the search name)
SN Culture	Han	Culture of the surname
GN Culture	European	Culture of the given name
SN Score	100	Matching score for the search name's surname.
GN Score	95	Matching score for the search name's given name.

Table 2- Columns in the external references grid (personal name)

2.4.2 Search Result Details– Name Details Tab (personal name)

The second tab in this dialog shows a property grid with small amount of additional information about the name. This includes its category, original given name and surname, and the value of the alternate parse flag (indicating whether this search name was originally added as an alternate parse of the source name, suggested by NameWorks as a more plausible parse than the one in the input file).

Figure 9 shows this tab:

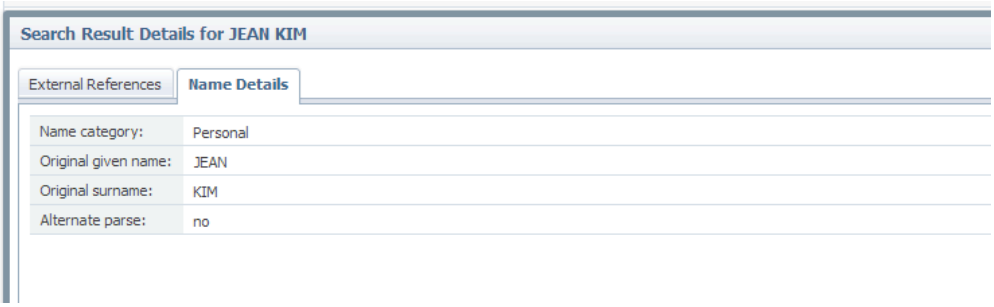

Figure 9 - Search result details - Name Details tab (personal name)

2.4.3 Search Result Details– External references tab (organization name)

Figure 10 shows this dialog for an organization name[3]. It is like the one in Figure 8 but with fewer columns, since organization names are simpler:

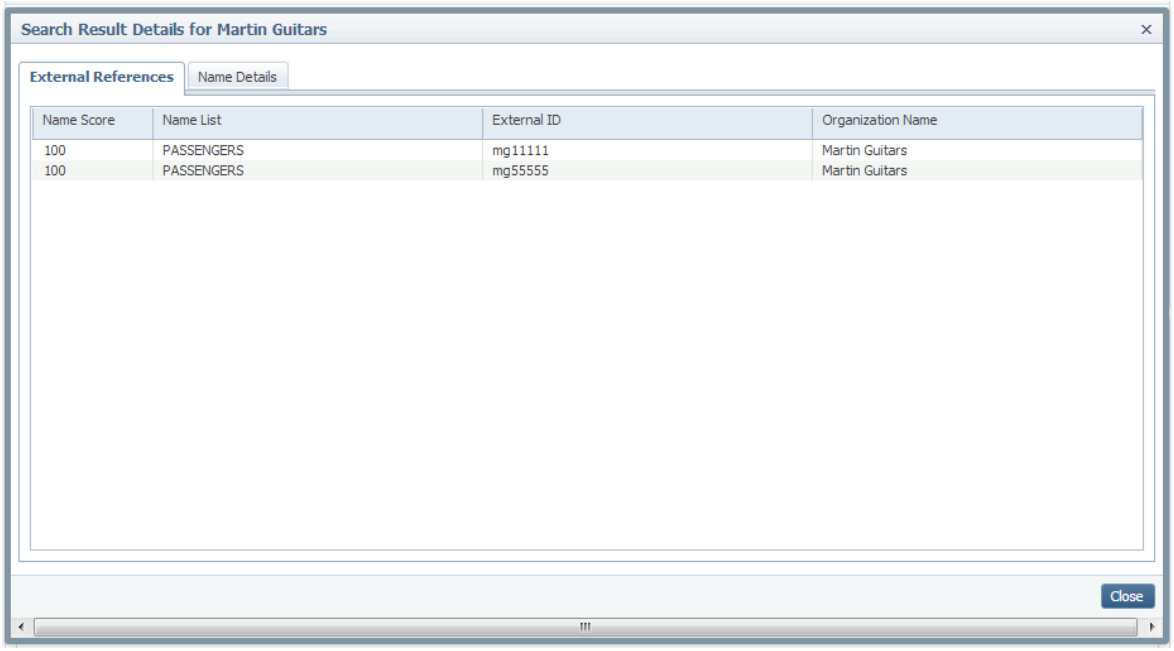

Figure 10 - Details dialog - External References tab (organization name)

[3] This is test data, of course. In real data, an organization would probably not be on a passenger list.

As before, the grid lists external references (or, equivalently source names), one per row. Here, the columns are:

Column Name	Example Value	Notes
Name Score	81	Score (for the search name found by NameWorks mapping to this source name). 100 is best.
Name List	PASSENGERS	Name of the name list where this source name (and external reference) is found.
External ID	mg11111	ID string provided for this source name when loaded. Typically, a key into a customer's external database.
Organization Name	Martin Guitars	Organization name, as originally provided.

Table 3- Columns in the external references grid (organization name)

2.4.4 Search Result Details– Name Details Tab (organization name)

The name details tab has a property grid that just lists the name category, original organization name, and score:

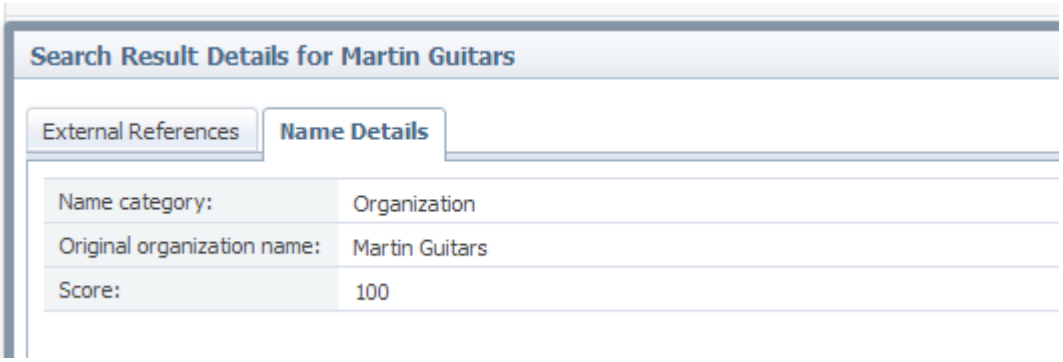

Figure 11 - Search result details - Name Details tab (organization name)