

Using the Identify Database for Self-Affiliation: A Guide

Uses for the Identify Database

More and more Ringgold customers are using the Identify Database as an authority file for a variety of innovative purposes. In addition to disambiguating institutional subscribers and licensees, it can be used to affiliate authors to their academic institutions, society members to their institutions, conference attendees, researchers etc.

It is becoming more common for this to be done using real-time self-affiliation, inviting customers/authors etc to select their own affiliations when they log in, register, buy content, or submit articles. The Ringgold ID can then be incorporated into data repositories across the business, enabling the capture of clean data to present a more accurate picture of contacts from specific institutions. Further, this real-time approach eliminates the need for a retrospective clean-up of your data.

Case Study: ORCID

ORCID has been using the Identify Database to normalise the educational and employment affiliations of researchers since 2013. By using Identify when researchers create a new ORCID record or add information to an existing record, problems with free-text entries are avoided: misspellings, ambiguous acronyms, and versions of the same institution in multiple languages. ORCID reviews free text entries periodically, and uses Ringgold's Validate service to obtain new Identify records as needed. The ORCID integration uses predictive type-ahead searching, and attaches the Ringgold ID, institution name and location to a user's ORCID record. ORCID uses a locally-loaded copy of Identify in their implementation. More than 300,000 ORCID records now contain Identify-supported affiliations.

There are two ways to incorporate *Identify* into an interface which captures data on individuals:

- using Ringgold's Application Programming Interface (API), or
- hosting the Identify Database internally (local load).

This document provides information and recommendations for each path of integration.

Option 1: Using the API

Ringgold has developed a series of API calls which can be used by the client to interrogate Ringgold's copy of *Identify*. Please refer to the current version of the [Identify Database Web Service](#) schema for complete details.

- Very simple code can be used to plug the API into an existing registration page
- Uses Ringgold search algorithms
- Ringgold's copy of *Identify* is constantly updated, so the client is accessing the most up-to-date version of the database at all times
- Eliminates the need for in-house database maintenance and warehousing by the client.
- Download restrictions exist, but they can be increased if needed

All metadata (including ISNIs) is returned as xml, so the client can add and display whatever information they want on their own interface pages. There is a demo portal for the current API, SOAP 2.4 at <https://idproduction.ringgold.com/webservedemo24.html> which can be used to test the various methods of obtaining information. A license key can be provided on request.

There are currently a number of services provided by the API to match specific client needs:

Search by Ringgold Identifier

This query will return the single institution with a specific Ringgold ID – with name, city, state, zip code and country information. It can be used to obtain organisation details when a Ringgold ID has been supplied, for example with an order or transfer from a third party. Records in the academic sector will return classifications details, and for records at Level -1 downwards, these will include subjects.

Find Institutions

This finds records on a full or partial name, using several search fields, and can be used to find a Ringgold ID to add to internal systems, as a definitive Identifier of a particular organisation, or to affiliate an individual to an organisation. There are a number of options available here:

- Search Key – the search is a simple pattern match. For example, when the Search Key is 'Portland State', this will match to 'Portland State University', but will not match to 'State University of Portland'. URL/email address can be used in addition to the name.
- A wildcard (%) can be added, which will produce a match to 'Portland%State' of any records with those combinations of letters in that order, including 'Portland Estates Elementary School'.
- This option also allows for searching with City, State, Post Code and Country (ISO 2 only), or combinations of these.
- The number of results returned using this option can be set using the Page Factor setting.
- Results are returned in strict alphabetical order
- Ringgold aims for 100% uptime for the server behind the API.
- These searches return Ringgold ID, Name, City, State, Post Code and Country

Find Institutions by Keywords

This service uses multiple search patterns with a matchScore basis, using tight, medium or loose Boolean matching. The keyword search covers the Institution Name, Alternative Names, Location fields (city, state, post code, and country), and URLs.

- Client chooses what range of scores to display down to
- A wildcard can be used on the right-hand portion of a search keyword, eg Port*, Univ*
- Maximum of 550 for a success score
- This option also allows for searching with City, State, Post Code and Country (ISO 2 only), or combinations of these.
- Results are displayed in rank order as follows: exact match to the Institution Name, exact match to an alt name, name contains phrase, alt name contains phrase
- These searches return Ringgold ID, Name, City, State, Post Code, Country and main URL

Get Institution Family Tree by Identifier

Using the Ringgold Identifier for a specific institution, this service will return all of the institutions within the *Identify* hierarchy relating to that record.

- The initial results return Ringgold IDs, which can then be looked up individually
- Results are displayed in order of level, with the Ringgold Identifier that specified the search in bold to indicate its place within the hierarchy.

We strongly advise our API users to cache retrieved institutions for a short time to ensure highest performance using the API. Therefore, getFamily only returns institutional_identifiers so the user can look in their cache first before requesting the full institution document with getInstitution.

Option 2: Hosting the Identify Database (local load)

This approach allows clients to host the complete Identify Database locally, and retrieve possible affiliations from this local copy. Please refer to the [Identify Database External Schema](#) for complete technical details.

- Search algorithms can be written in-house, to the client's specifications. Any number of fields can be searched to present the user a limited list of authoritative options from which to select their affiliation.
- Complete database is delivered to the client in an automatic feed each Friday by noon EST. A secure Sharefile protocol is used, and multiple formats are available.
- Locally held data is up-to-date.
- The Identify Database consists of a set of tables containing all the metadata for each record, linked by the Ringgold ID.

For best results when creating your search specifications, Ringgold recommends:

- A minimum number of free text fields are used during data capture
- The use of both the parent and alt-names table (plus ideally URLs and email addresses) when searching should ensure the most accurate results. For example, you might use data

provided by the user in other fields drive which *Identify* records are presented for possible selection; e.g., a user's email address will reveal domain name (user@harvard.edu), which can be matched to a specific institution or hierarchy in *Identify*.

- Selection at the most granular level within a hierarchy will result in more accurate data being captured

Adding new institutions

Although the Identify Database contains more than 440,000 records, there will be times when a look-up is unsuccessful: either the user did not recognize the proper affiliation from the *Identify*-generated list, or they are entering a truly new institution. Once the client has verified that the individual's free text entry does not match anything in *Identify*, there are two ways to obtain a new *Identify* record from captured free text:

Validate

The *Validate* service from Ringgold allows the client to add a few basic pieces of information to our *Identify* production system, and receive a Ringgold ID immediately. The entry is then researched by our experts, and a complete *Identify* record is added to the database. In cases where a *Validate* record is found to be a duplicate, or not valid for some other reason, the client is informed within 24 hours. In the case of a duplicate entry, Ringgold advises the client of the correct *Identify* record to use.

Identify Audit Service

It is possible to send batches of records to Ringgold for audit. During this process, each record will be allocated the correct Ringgold ID, and any organisations not in the Identify Database will be added. This process takes longer to complete (up to four weeks from receipt of the file), and is suitable when Ringgold IDs are not required instantly by the client.

Hierarchies

The Identify Database is organised hierarchically, with parent/child links between institutions with either an administrative or financial tie.

Ringgold recommends having users affiliate with the most granular level possible, which will future-proof your data in the event hierarchical relationships change over time. It is possible to direct users to the right level within their hierarchy using either the local loading solution or the API. The API also has a service which produces the details of hierarchies, Get Institution Family Tree by Identifier. This returns the Ringgold ID and level for each member of a hierarchy, which can then then permit users to select the most granular institution; i.e. a department of study rather than simply the university level record.



Questions?

Please contact us should you require support during your implementation.

North America
UK & Europe

Christine Orr +1 540 359 6620
Diane Cogan +44 (0)7875 992880

christine.orr@ringgold.com
diane.cogan@ringgold.com