



Identify

An Authoritative Database of Institutional Identifiers

By

Helen L. Henderson
Managing Director, Ringgold, Ltd.

And

Dr. Ralph M. Shoffner
Chairman, Ringgold, Inc.

For Presentation To

DLF/CrossRef
Institution Registry Stakeholders' Meeting

Jurys Washington Hotel, Washington DC

October 7, 2005

www.ringgold.com

1. Introduction

More than six years ago, Helen L. Henderson, Managing Director of Information Power, Ltd., recognized the need for efficient provision of authoritative institutional information in support of the electronic delivery of information. After years of exploring alternative approaches with such organizations as Catchword, the Association of Subscription Agents and OCLC, Helen was persuaded by publishers and standards organizations to bring together a group of interested organizations to propose a new service that would allow authoritative institutional information to be held centrally and to allow that information to be shared reliably and securely. This would reduce the effort of both publishers and users to keep each other up to date with changes in institutional information and credentials needed for authentication. Groups such as COUNTER, ICEDIS and ONIX consider the institutional authority database an essential element to provide consistent and comprehensible measurement of how electronic resources are being used and managed.

In July, 2005, Ringgold, Inc. merged with Information Power and is implementing this proposal, in the form of *Identify*, a service to meet the need for authoritative institutional information.

Identify comprises:

- A database of authoritative institutional information assembled on behalf of a collaborative group of publishers, currently covering over 30,000 institutions in 18 countries;
- A web services application, to support inputting, linking, managing and updating information linked to the database by the administrators;
- Interfaces to the database to allow subscribers and trusted external parties to input, link, manage and update information in the database as well as to add new institutions.

Identify's business model allows free access to the database by the users (libraries, institutions, etc.), and for a general look-up service. The revenue to support the service will come from charging vendors, information providers, or other companies wishing to access the database for downloading with a variety of added value services available for subscription fees or one-off charges.

2. Background

The need to identify institutions is hardly new; taxing authorities have long used identifying numbers to differentiate their tax paying entities. Dunn and Bradstreet established the D-U-N-S Number to improve their financial reporting. D&B reports 66 million currently active companies in their database.

R. R. Bowker initiated and maintains the SAN (Standard Address Number) identifier to signify a specific address of an organization in (or served by) the publishing industry. It is now a USA standard (ANSI/NISO Z39.43-1993) and over 34,000 SAN numbers are active.

The Library of Congress' *MARC Code List for Organizations* has its roots in over 100 years of organization identification, beginning with Charles Evans' *American Bibliography* first published in 1903. LC reports that the *List* contains 27,719 valid identifiers for general use.

We now have an international standard identifier, ISIL, Information and documentation - International Standard Identifier for Libraries and Related Organizations (ISO 15511-2000) - and the Registration Authority started work in 2004. There are 9 countries that have designated registrars (Australia, Germany, Denmark, Egypt, Islamic Republic of Iran, The Netherlands, New Zealand, Norway, Switzerland) and two more are "under registration" (United Kingdom, United States of America).

There are many other institutional identifiers about and one could create still another. However, reviewing the host of institutional identifiers leads us to a different conclusion. We should focus our efforts on improving the communication between publisher, aggregator and institutional consumer, by delivering accurate, inexpensive information about the consumer, including present available identifiers to assist in the identification of the institution.

It should be noted that, in contrast to the international standard metadata for bibliographic data (MARC21), there is no such international standard metadata for organizations. As a practical matter, this means that anyone building a database must define the metadata and the method of representation. It further means that the publishers will have done this same thing and their metadata and their organizational identifications will overlap, but are not likely to be identical. Thus, an *Identify* organization is defined at an elemental level so that it matches but one organization within each publisher's system. Finally, both organizations and publisher pricing policies are dynamic. As a result, there will be continuous change in the metadata maintained for the institution.

There are over 300 publishers listed in the members lists of ALPSP (the Association of Learned and Society Publishers), SSP (the Society for Scholarly Publishing and STM (the International Association of Scientific, Technical and Medical Publishers). The majority of these publish journals which are available electronically and have hosting and subscription services either in-house or outsourced. Until their journals went electronic the only information they needed about their customers were their mailing addresses. Now they need information about who are they, where they are, what their constituent parts are and exactly what their electronic entitlements are.

The information consumers, and particularly libraries, do certainly have a role in maintaining the institutional information necessary for access, and in many cases they have to register with 100 or more separate services. This means that every time they change any of their credentials (IP address ranges, OpenURL resolvers, etc) they have to contact all the services. There is, therefore, an opportunity to change this relationship from "many to many", to "one to one", by setting up a single source of information on each institution, giving the information needed to identify the institution, its rights, such as subscriptions or licenses, and the correct delivery points, e.g., the institution's IP address ranges.

The needs of the information provider go beyond the atomized information about the individual institution. There are relationships among institutions that must be recognized in order for the

provider to recognize the services that are appropriate. For example, the research laboratories of a multi-national corporation may have different levels of access, while subscriptions may be purchased for the joint benefit of members of a consortium who bear no other organizational relationship to each other.

Working with a group of academic publishers, Ringgold has created its *Identify* database to maintain, search and display all manner of identifiers and all manner of organizational relationships. The database contains authoritative records for institutions in 18 countries (USA, Canada, UK, Ireland, Australia, New Zealand, Netherlands, Finland, Sweden, Norway, Denmark, Germany, Switzerland, Austria, France, Italy, Spain, India, China and Japan), and includes academic, government, corporate and independent organizations. We will continue to maintain and extend this database by providing product and services valuable to publishers and other information providers.

There is a debate to be had over the use of the DOI as an identifier for institutions. Certainly, we could assign DOIs, or we could use DOIs assigned elsewhere. But if they are assigned elsewhere, there will need to be sufficient metadata associated with them to discern where in the network of organizations the individual DOI is applicable. As with any other organization, maintaining the integrity of our information requires that we use an identification number that is under our control; however, that does not reduce the desirability of communicating via the identifiers maintained by other organizations. When we encounter an organization lacking one or more appropriate identifiers, we would be happy to report the institution information to the issuing agency, or to give the institution the links needed to register.

3. Benefits

The services offered will provide an opportunity for both information providers and users to reduce the administrative burden of maintaining reliable information relating to the sales, provision and management of online products

The benefits for publishers accessing a central database of accurately defined and uniquely identified institutions include:

- ◆ The ability to keep up to date with changes in customer data, such as authentication range changes, mergers and acquisitions
- ◆ The ability to provide customers and potential customers with instant and automatic holdings reports, access rights reports, and tailored quotations, e.g. via a web interface
- ◆ The potential to eliminate the common phenomenon of institutional renewals being unrecognised as such and therefore appearing as new subscriptions, whilst the old ones falsely appear to have lapsed
- ◆ More accurate information for marketing departments on their customers and on their penetration of particular markets
- ◆ The ability to easily merge/purge customer/subscriber lists when companies or journals change hands

- ◆ The ability to assess the degree of market overlap between potential business partners, perhaps via the services of a trusted third party
- ◆ The ability to provide an intelligent and informed response when an aggregator partner (per the terms of a bilateral agreement) requests approval of a particular sales deal and asks for a suggested price
- ◆ The ability for customers to enhance the new value created by COUNTER compliant usage reports. Adding up all the PPVs, other document delivery, (and maybe unsuccessful requests for articles in journals not licensed) at institutional rather than library level could generate new licensing opportunities for publishers
- ◆ List rental opportunities (“I want all the libraries on your list which are not currently on mine”)
- ◆ Marketing response analysis opportunities – comparing institutional affiliations of individual prospects on rented lists with institutional affiliations of subsequent new personal or institutional subscriptions or purchases
- ◆ Easier ability to track institutional end-users of consolidated subscriptions, especially those where the agent does not deliver orders via ICEDIS structured FTP with Type 2 addresses incorporated in the complete record
- ◆ When customer applies for trial or subscription, ability to insert its identifier and automatically retrieve the stored credentials for checking
- ◆ Offer customer customized instant trials with choices based on the credentials held in the system.

The benefits for the user community include:

- ◆ A central point to lodge information about the institution which it currently has to supply to all of its information providers
- ◆ Potential to eliminate incorrect reporting of subscription renewals and cancellations
- ◆ A centralized reporting system which can provide an accurate and up-to-date list of all subscribed eResources with additional, consolidated information
- ◆ Potential to provide authoritative lists of subscriptions by publisher (particularly important when journals/journal bundles move to a different publisher)
- ◆ Ability to easily view subscriptions at both a library and a consortium level
- ◆ Developing services which would measure holdings, access rights, and usage against a mean for one’s market sector, as defined by size/type/country.

4. Product Description

Ringgold recognizes that the success of *Identify* will depend upon its accuracy and timeliness. At the same time, we now recognize how much effort is required, not only to create, but also to maintain accurate information about organizations. To accommodate the joint demands of accuracy and timeliness once operational, we will accept immediately changes input by the

affected organization, but we will also mark these changes as unverified by Ringgold. Until the changes are verified, the previously verified information will still be available, so that both forms will be displayed.

The following metadata is currently available in the database, and Ringgold is actively reviewing needs for all constituencies:

Institution Identifiers	Ringgold assigned, Government (tax) id, ISIL, LCMOC, D-U-N-S, SAN, ...
Institution Name	Legal name in native language (where appropriate)
Alternative Names	Commonly used names, English language name
City	City of administrative HQ
Region/State	Where appropriate
Zipcode	If available
Country	Country of administrative HQ
URLs	
Tier	Carnegie category, JISC band
Size	number of staff, students, employees number of beds, lawyers (as appropriate)
Credentials	IP address range, Athens domain, Shibboleth attributes
URL resolvers	
Organization logo	Where applicable
Linked Institutions	Name and Nature of Link

The main services available using the **Identify** database are:

- ◆ **Identify** Look-up: this will allow subscribers to look up individual institutions for reference purposes
- ◆ **Identify** Download: this will allow subscriber to download information from the database
- ◆ **Identify** Alerts: this will alert subscribing vendors to changes in institutional information, like authentication credentials (e.g. IP address range changes, moves from Athens to Shibboleth) or changes in name or structure
- ◆ The **Identify** Information Service (IIS) available via an access controlled web interface. This allows only those authorized persons to propose, add or approve information.

Librarians and in particular library administrators will be able to access information relating to their institution(s) via an information service (though it is proposed that such services be offered gratis) which would include:

- ◆ Access to forms to apply for access to their institutional information

- ◆ A “My Institution” access-controlled web interface to review information held on their institution
- ◆ Ability to update, via forms, information relating to their institution (this could be a more restricted group than those allowed to review information held)

There are a number of other groups which could have additional services or additional options within a service based on their type, for example:

- Consortia would be able to subscribe to services that would allow them to access the information relating to their member institutions
- Library suppliers and subscription agents would be able to subscribe to the service to support their CRM and Publisher Relationship services. For example, they would be able to review and manage information based on the relationships between customers and institutions, as set up by the publisher or library
- Integrated Library System (ILS) vendors would be able to subscribe to a service to support their Electronic Resource Management (ERM) systems.

5. Advisory Board

Ringgold’s principal concern with *Identify* is that it be “right”, that is, that it be of the highest quality and reliability. As indicated in the Introduction, Helen Henderson has discussed the needs and potential of institution identification with a large number of professionals in the field. She has received preliminary, affirmative responses from those listed below to participate in an Advisory Board for *Identify*, to help ensure that it meets the requirements of the vendors, information providers and the users:

Librarians:

Jim Mouw (U Chicago)
 Chuck Hamaker (UNC Charlotte)
 Hazel Woodward/Simon Bevan (Cranfield U)
 Christine Fyfe (U Leicester)
 Ruth Jones (British Library)
 Sally McCallum (Library of Congress)

Publishers:

Richard Gedye (OUP & Chair, COUNTER)
 Greg Malar (Rockeller & Chair, ICEDIS)
 Gertraud Griepke (Springer SBM)

Organizations:

Brian Green/Tim Devenport (ICEDIS/EDItEUR)
 Pat Harris (NISO)

6. Timing

This is the first demonstration of *Identify*. We will be presenting it at the Frankfurt Book Fair and at the Charleston Conference.

The *Identify* demonstration database contains real data, but it is a freestanding database. Before going operational, we will redo the table structures of the database to correspond with our other databases.

We are ready now to sign up a few publishers to move *Identify* to production status. We expect to wait until after the holidays to begin intensive work with these publishers on their interfaces and workflows, with the expectation of going operational in the period March 31 – June 30. One aspect of this work that will be carefully delineated will be the interaction of customer support. We expect that the publisher will want us to be the first point of contact in the case of service failures. In that case, we will need to be “in the loop” so that we can assure that the service has been reinstated satisfactorily. In the electronic environment, successful customer service is 24x7x365 and we will provide it.