Identify This! Identify That! New Identifiers and New Uses

January 11, 2012

Speakers: Roy Crego, Janifer Gatenby, and Chris Shillum

What is the ISTC?

• A global identification system for textual works
• ISO Standard (21047) since March 2009
• 16-digit alphanumeric code
  - Digits 1-3 = Registration Agency Identifier
  - Digits 4-7 = Year of registration
  - Digits 8-15 = Title number
  - Digit 16 = Check digit

Example: A03-2011-000010B-1

ISTC Key Features

ISTC
The Identifier for Textual Works

Roy Crego
11 January 2012

Collocation
• Groups products containing the same textual work

Hierarchical Structure
• Uses Source ISTC to tie derivative products (abridged audio, Spanish translation, 2nd edition) to parent work

Central Registration
• The same ISTC is used on same text from all publishers across all countries
ISTC International

• 10 ISTC Registration Agencies representing 6 languages to date:
  - Boekenaankerk (Flemish)
  - Bowker (English)
  - BTFL (Canada) (French)
  - CCR (Chinese)
  - Cercle de la Librairie Electre (French)
  - FIIP (Russian)
  - MVZ (German)
  - Nielsen New Zealand (English)
  - Thorpe-Bowker (Australia) (English)

• Central ISTC Database (STRS) went live in 2009.

• New ISTC-International Site (www.istc-international.org) launched July 2011 with public ISTC lookup.

The Need?

• Over 3 million ISBNs were assigned in the US in 2010.
• E-books, reprints, and self-publishing are adding to that number daily.
• Metadata for similar products needs to be linked

The ISTC Solution

• ISTC designed for use by computers on the web
• ONIX for ISTC format (ver. 1.01)
  - ISNI Support within ContributorIDType tag
• ONIX for Books support
  - Ver. 2.1 WorkIdentifier
  - Ver. 3.0 Work Relation Code (01=contains work; 02=source work)
• MARC Discussion Paper (2010-DP04)
  - Consensus to use 024 and 787 fields

Find Same Text in Many Formats/ISBNs

Breaking Dawn
ISTC A03-2010-0000000F-F

- 9780316178136: Hardcover
- A03-2010-0000000F-F
- 9780316178166: Paper
- A03-2010-0000000F-F
- 9780316178284: E-Book
- A03-2010-0000000F-F
- 9781616720203: Audiobook
- A03-2010-0000000F-F
Find Derivative Versions of the Text

The Red Badge of Courage (1895)
ISTC A03-2010-0000093-C

• Find Translations and Title Variations

Rollout

• Bowker is storing ISTCs in Books In Print for distribution when “critical mass” is achieved.
• Feedback needed on what is “critical mass.”
• Publishers, libraries and retailers all need to encourage the use of ISTC in product records.
• ISTC Registry data available (by license) through web call.

Additional Information:

Bowker ISTC Agency
www.myidentifiers.com/istc/main

Roy Crego
Product Manager
ISTC, ISNI, DOI
(908) 219-0240
roy.crego@bowker.com

International ISTC Agency
www.istc-international.org

ISTC Search Page
Specifications
User Manual
ISNI: Scope, Participation and Operation

NISO Webinar “Identify this, identify that: New identifiers and new uses

January 11th 2012  Janifer Gatenby, OCLC EMEA

At January 2012

• ISNI is an ISO Standard ready for publication (FDIS)
• ISNI-IA incorporated in the UK in December 2010
  • ISO contract agreed; in process of signature
• OCLC (Leiden) appointed as the ISNI-IA Assignment Agency
  • Created the initial ISNI-IA database (2011)
  • First million ISNIs assigned and being diffused to data contributors
  • Ongoing assignment operations in test, to start Q1 2012

The ISNI-IA Founding Members

• CENL (Centre of European National Libraries)
  • 48 European national libraries, represented by BL and BnF
• CISAC (International Society of Authors and Composers)
  • 225 societies from 118 countries (87% musical composers)
• IFRRO (International Federation of Reproduction Rights Organisations)
  • 135 organisations from 74 countries
• IPDA (International Performers Database Association)
  • 37 societies from 28 countries
• OCLC
• Proquest/Bowker (BIP)
Why have ISNI?

- To differentiate one Will Smith from the other
- International and Cross Disciplinary in Scope

Thanks to David Grundy, ALCS

Typical result of a web keyword search on Will Smith retrieves noise titles does not retrieve titles by the Fresh Prince

More use cases for ISNI

- Rights management societies often receive payments then need to research to find the rights holders
  - As digital grows, need is greater and requirement is international
  - Need the system but are bound by data privacy
  - Can use VIAF as interfacing public domain data
  - Shared work load = efficiencies + higher quality
- Researchers need identifiers
  - Reputation management
  - Grant applications

Dorothy Wood, the Scottish lady who writes craft works did not write about Danish immigrants to South Australia
More use cases for ISNI

- European Arrow Project
  - Envisages using ISNI with ISTC for registration of digitisation rights information

- Supply Chain
  - NISO I2 committee recommended the ISNI Assignment system for all institutions in the Digital Supply Chain

- Music Industry
  - Need for a unique global level identifier shared by record labels and exposure sites

ISNI Scope

- International

- Cross domain
  - Including creators and other contributors in all disciplines (authors, editors, translators, illustrators, composers, actors, performers, artists)
  - And all organisations involved in the distribution and supply chain of created works (publishers, intermediaries, retailers, database vendors, libraries)

- Must have centralised registration
  - Global network of registration agencies
  - Cannot use ISBN model of allocation of ranges

ISNI’s relationship with ORCID

- ISNI-IA is advocating one shared scheme
  - Confusing to have 2+ identifier schemes appearing at the same time for the same identities
  - Dilutes effect of linked data
  - Corrections easier to administer with one scheme
  - Cross domain identification - e.g. writer of scientific articles also book author and song writer
  - ISNI linking identifier in ISO (ISBN, ISTC, ISAN, ISWC ++)

- Interoperating systems
  - Method to be negotiated; aim for SYNERGY
  - ORCID’s focus is on end user input; ISNI’s is on registration

The ISNI database

- Free public access to core metadata of Assigned ISNIs:
  - http://isni.oclc.nl via 4 indexes
  - SRU: http://isni.oclc.nl/sru?DB=1.2/ (XML, RDF coming)

- Member access (RAG and Data Contributors)
  - Access to full metadata, except confidential data
  - Access to all records, assigned and unassigned via 20 indexes
  - Ability to add and modify own data

- Administrator and Quality Team
  - Full access, additional indexes
Leveraging high confidence data from different domains

- British Library
  - JISC names (research grant data), UK theses, ZETOC (Possible)
- CISAC
  - IPI - International Party Identifier (87% musical composers), 2 million records
- IPDA (International Performers Database Association)
  - 500,000 performer records
- IFRRO (International Federation of Reproduction Rights Organisations)
  - Including access® (Canada), ALCS (UK), CEDAR (Netherlands), CEDRO (Spain), Librius (Belgium), Prolitteris (Switzerland), VGWort (Germany) (135 orgs from 74 countries)
- Proquest/Bowker
  - BIP (books in print), Theses and Scholar Universe, American professional societies

Base cross domain file of the ISNI database, building on work already done

An ISNI is assigned where:

- Metadata from 2 or more independent sources matches with a sufficient level of confidence (match confidence)
  - Matching adapted from VIAF matching
- Or metadata has 3 or more VIAF sources
- Metadata is complete and unambiguous (special rules for sparseness and for common surnames)
- All records have a data confidence level
  - Indicating closeness of contact with party behind the identifier
  - Rights management societies have highest confidence
ISNI Assignment - Possible matches

- where the matching score is between the no match threshold and the match threshold
- where the incoming record have different local identifiers for the same source (could signal either a merge or a split)
- where the incoming and match record disagree on birth or death date
- where the incoming and match record share the same local identifier but their similarity is below threshold

ISNI Database Quality

- Team at Bibliothèque nationale de France and the British Library (Anila Angjeli, Michael Docherty, Nicole Druet, Andrew MacEwan, Richard Moore and Alison Wood)
  - Manual checking of statistical samples, resulting in numerous matching algorithm improvements
  - Review by data source - input to data policies
  - Establishing the percentage of data errors & characteristics
  - Resolution of queries from RAGs and general public
- Program analysis
  - Creation of anomaly checker, e.g. publishing before 10 years old
  - Diffusion of corrections

Centralised Registration / Diffused Collection and Management

- Network of nodes and expertise
  - RAGs for gathering, completing, assessing the quality of input data, responding to disambiguation responses
  - Verification Agencies and Reference Databases for input to disambiguation, matching, assignment and correction
  - RAGs and Verification Agencies / Reference Databases for diffusing ISNIs and promoting usage
  - Assignment Agency stores URLs in database and sends notifications of changes and corrections
  - All ISNI members responsible for Quality; special role for the Quality Team

From RAGs:
- ISNI requests
- Batch load
- Web and Atom Pub API
- ISNI enquiries
- Service issues to Service Desk

To RAGs:
- ISNI enquiry responses
- ISNI Responses
- Reports and Notifications
- Service Desk responses
- Relay Q issues to Quality Team
From General Public:
• ISNI enquiries (Web and SRU)
• Data quality issues from isni.org sent to Quality team
• At least one RAG will accept public requests on the ISO RAND principle

Modules:
• Web enquiry (filtered)
• Downloadable search box
• SRU (filtered)
• Suggestions via isni.org

In conclusion
• ISNI-IA
  • Not for profit, incorporated in the UK, unprecedented cross domain alliance
  • Funded the creation of the database and assignment system
  • Assignment Agency and RAGs also on RAND cost recovery
  • Ongoing costs are modest - no permanent staff, permitting price per ISNI to be as low as possible
  • Emphasis on registration and data quality, building on existing data
  • Data privacy is respected while core data is open
  • diffusion and usage of assigned ISNIs is free and is encouraged

Problem
Estimated 27 million researchers world-wide including private, government and research institution, not students

Many researchers have common names, change names, or are otherwise difficult to uniquely and unambiguously identify
ORCID will work to support the creation of a permanent, clear and unambiguous record of scholarly communication by enabling reliable attribution of authors and contributors.

ORCID will Meet Needs for Multiple Stakeholders in the Scholarly Communication Chain

- **Academic Institutions**
  - Help track output of faculty and students
- **Funding Bodies**
  - Help perform research assessment of grantees
- **Publishers**
  - Streamline data input
  - Creates author link
    - to publications
    - to collaborators
    - to other forms of communication

By issuing unique identifiers to all researchers, ORCID will facilitate discovery and evaluation for researchers, institutions, scholarly societies and publishers.

- Joins faculty or student body
- Joins scholarly society
- Applies for grant
- Submits manuscript
ORCID History

- ORCID Initiative started in late 2009, based on initial proposal by Nature Publishing Group and Thomson Reuters
- ORCID non-profit organization with a Board of Directors created in August 2010.
- Now over 275 participants from all sectors in scholarly communication
- Launch of service planned for the first quarter of 2012.

ORCID Principles

These 10 Principles were adopted unanimously by the ORCID Board in December 2010:

1. ORCID will work to support the creation of a permanent, clear and unambiguous record of scholarly communication by enabling reliable attribution of authors and contributors.
2. ORCID will transcend discipline, geographic, national and institutional boundaries.
3. Participation in ORCID is open to any organization that has an interest in scholarly communications.
4. Access to ORCID services will be based on transparent and non-discriminatory terms posted on the ORCID website.
5. Researchers will be able to create, edit, and maintain an ORCID ID and profile free of charge.

6. Researchers will control the defined privacy settings of their own ORCID profile data.
7. All profile data contributed to ORCID by researchers or claimed by them will be available in standard formats for free download (subject to the researchers’ own privacy settings) that is updated once a year and released under the CC0 waiver.
8. All software developed by ORCID will be publicly released under an Open Source Software license approved by the Open Source Initiative. For the software it adopts, ORCID will prefer Open Source.
9. ORCID identifiers and profile data (subject to privacy settings) will be made available via a combination of “no charge” and “for a fee” APIs and services. Any fees will be set to ensure the sustainability of ORCID as a not-for-profit, charitable organization focused on the long-term persistence of the ORCID system.
10. ORCID will be governed by representatives from a broad cross-section of stakeholders, the majority of whom are not-for-profit, and will strive for maximal transparency by publicly posting summaries of all board meetings and annual financial reports.

275+ Participants
Newest Participants (Sep-Oct 2011)

ORCID is open to any organization with an interest in scholarly communication

ORCID transcends discipline, geographic, national and institutional boundaries

Board of Directors
Represents a broad cross-section of stakeholders with the majority being not-for-profit

<table>
<thead>
<tr>
<th>Publishers</th>
<th>Elsevier</th>
<th>Crossref</th>
<th>OCLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academia</td>
<td>InPurpl</td>
<td>M4H</td>
<td></td>
</tr>
<tr>
<td>Funding Bodies</td>
<td>wellicomtrust</td>
<td></td>
<td>NII</td>
</tr>
</tbody>
</table>

250 participating organizations as of 1 August 2011
Researchers will be able to create and maintain an ORCID ID and profile free of charge, and will control their privacy settings.

Types of Identity Assertion

- **Self-Asserted Identity**
- **Socially-Validated Identity**
- **Organizationally-Validated Identity**

**ORCID Phase 2 – Disambiguation Concept**

Self-asserted + socially-validated + organizationally-asserted identity = more credible assertion

Kaliya Hamlin: [http://www.identitywoman.net/the-identity-spectrum](http://www.identitywoman.net/the-identity-spectrum)

Geoff Bilder, “Disambiguation without de-duplication: Modeling authority and trust in the ORCID system”: [http://www.orcid.org/node/615](http://www.orcid.org/node/615)
What Makes ORCID Different?

- Only not-for-profit contributor identifier initiative dedicated to an open and global service focused on scholarly communication
- ORCID is backed by a non-profit organization with over 275 participants behind it
- ORCID is backed by many different stakeholders
- Publishers are an important ORCID stakeholder but are just one part
- ORCID is serious about building an open system
- ORCID is the only researcher identifier that is not limited to discipline, institution or geographic area
- ORCID is the one to bridge them all by registering the identifiers of all other relevant standalone services (silos big and small)

Come Join Us

Register as a participant at www.orcid.org

Follow us on Twitter: @orcid

Attend the next Outreach Meeting:
17 May 2012 in Cambridge, MA