Abstracting & Indexing at ProQuest

for the SLA Education Division

presented by
ProQuest Training & Consulting Partner
aimee.leverette@proquest.com
Aimée D. Leverette
ProQuest Training & Consulting Partner
Master of Science in Information Sciences – Univ of Tennessee
Master of Education – Colorado State University

Start Date: January 31, 2005

Home Office: Fort Collins, Colorado

Territory: AK, CA, CO, HI, ID, MT, OR, UT, WA, & WY
British Columbia & Alberta

📞 303-328-1589
✉️ aimee.leverette@proquest.com
Objectives of this Presentation

1. To present an overview of abstracting & indexing (A&I) at ProQuest.
2. To provide insight into the skills and responsibilities of a ProQuest indexer.
3. To provide definitions and descriptions of abstracting and of the ProQuest controlled vocabularies and authority files used to index.
4. To illustrate the day-to-day processes and procedures that allow an article to get “A&I’ed”.
Table of Contents

ProQuest Editorial Operations

What Does a ProQuest Indexer Do?

The Indexer’s Responsibilities

Abstracting

Indexing

ProQuest Controlled Vocabulary

ProQuest Authority Files

What gets A&I’ed?

Traits of an Indexer

Training a New Indexer

Conclusion
Main Editorial operations are in Louisville, Kentucky.

Louisville handles the receipt and processing of the majority of content loaded to ProQuest databases such as *ABI/INFORM Complete*, *Accounting & Tax Database*, *Banking Information Source*, *National Newspapers*, and *Research Library*.

A brief history of Louisville:
- Location of ABI/INFORM from its earliest days in the 1970s, by a company called Data Courier, owned by the prestigious Courier-Journal newspaper.
- Purchased by UMI in Ann Arbor in 1986.
- Databases began with a presence on Dialog, expanded to CD-ROMs in the 1980s, and moved to the Internet in 1995.
- Evolved from 40 employees doing A&I on paper for hundreds of publications to a current level of 100 employees doing original A&I or QA on thousands of publications.
- Employees have an average of 13 years’ service, with several in the range of 20-30 years.
What Does a ProQuest Indexer Do?

Our Indexers have 2 main jobs:
1. Creating abstracts and indexing for articles.
2. Performing quality assurance on indexing that has already been done by a human or through automation.

Tools Available:
1. An editorial system in which to create and search for records (i.e., state-of-the-art software).
2. A controlled vocabulary (thesaurus) and authority files for indexing company, person, geographic, and product names.
3. Editorial guidelines telling them how to choose index terms, how to compose an abstract, how to add new company or personal names to the file, and so on.
4. Access to the hard copy of a publication or to the text (ASCII or PDF) of articles on ProQuest or elsewhere, including the publisher’s website.

They Also…
- Help troubleshoot thesaurus searching issues.
- Keep Product Managers, Marketing Managers, and Trainers informed of new vocabulary terms and search techniques for finding content—information that can be critical in closing a sale.
The Indexer’s Responsibilities

Indexers for a specific database or subject area may be responsible for:

- Keeping track of how many articles they index.
- Completing A&I within expected time frames.
- Keeping track of missing publication issues or missing articles in an issue.
- Understanding the subject matter they index, especially if they specialize in certain types of material.
- Proposing new subject terms, new types of indexing, or new publications that will provide added value to the product.
- Knowing how to search their database using the company’s platform.
- Understanding how customers search and what they expect when they search—which translates into knowing how to index to meet those expectations.
- Knowing who the customer base is.
DEFINITIONS

• ANSI definition of an abstract: "an abbreviated, accurate representation of the contents of a document."

• A concise presentation in "the author's own language, of all the points made, in the same order as in the original" document, without any added interpretation or criticism by the abstractor.

• An abstract is "intended to stand as a readable and complete item in its own right."

• Publication, or Author, abstracts are used whenever possible; these are marked in ProQuest as “PUBLICATION ABSTRACT.”

CREATING THE ABSTRACT

• Abstracts may be handwritten by a member of an editorial staff.

• Abstracts may be copied from the full-text of a publication, whether that exists on ProQuest or another database aggregator’s platform, or on the publisher’s website.

*Long gone are the days when we wrote every abstract manually—competitive pressures mean that CURRENCY is the primary driver for database aggregators.*
Abstract Examples

**Publication Abstract ("Epidemiology and Infection - going electronic after 100 years" by Norman Noah)**

Epidemiology and Infection became 100 years old in 2001. Its powers and influence however do not appear to have diminished, and, as the Senior Editor appointed in its 101st year, I am greatly honoured to be so closely associated with it. Just over 100 years after it was founded, Epidemiology and Infection is 'going electronic'. To mark both occasions we have put together a small selection of some of the most important or thought-provoking papers published in our first century, and have invited eminent epidemiologists and microbiologists to assess and comment on them critically in the light of modern knowledge. Naturally, to mark our coming of age in terms of modern methods of communication also, all the papers chosen for our centenary edition will be available - free - online. I am most grateful to my predecessor, Professor Bo Drasar, for putting together this unique tribute to Epidemiology and Infection, its Editors, its Editorial Board, and, not least the authors of our papers. Enjoy and marvel at our predecessors' wisdom, perspicacity and scientific rigour. [PUBLICATION ABSTRACT]

**Hand-written abstract (Business Week article – “Cocktails Are Next For Cancer-Drug Makers” Aug 30, 2010)**

A diagnosis of AIDS was a death sentence until the advent of drug cocktails in the 1990s allowed patients to suppress the disease indefinitely. Now scientists say a similar combination strategy may change the course of cancer treatment. That is the bet being made at Roche Holding, AstraZeneca, and Sanofi-Aventis, whose latest efforts to develop a new generation of combination treatments are prompting the US Food & Drug Administration (FDA) to rewrite the rules for drug research. Cocktails of experimental drugs would not have been possible just five years ago, when FDA rules required that the merit of each active ingredient be proved before it could be added to a combination. Savings for drugmakers could be substantial if the FDA, as expected, sets up a path to allow companies to forgo differentiating the drugs during the last and most expensive trials needed for regulatory approval. The first combinations available to patients will be mash-ups of existing treatments, such as Roche's Avastin and Bayer's Nexavar.
**Indexing**

**DEFINITION**: The assignment of controlled key terms to an object (e.g., an article, document, or other type of content) to enable all content of the same subject to be linked together by common terminology.

*From the ProQuest Editorial Guidelines:*

> The purpose of indexing is to help database users—corporate librarians, consultants, industry analysts, students, executives, and university researchers, among others—find the information they need in as little time as possible. Indexers do this by assigning specific terms that identify the key topics and concepts presented by the article.

**What kinds of key terms are indexed by ProQuest?**

- Subjects
- Classification Codes
- Company names
- Geographic locations
- Personal names
- Product names

**Other indexed elements of an article:**

- Article (Object) Type (e.g., Feature, Cover Story, Editorial, News)
- Document Features (charts, tables, graphs, etc.)
- Source Type (scholarly/academic, report, trade magazine, newspaper, etc.)
• Used and maintained in Louisville, Cambridge
• Some third-party DBs & thesauri
• Some flat lists, others fully hierarchical
• Some managed more actively than others
The ProQuest Controlled Vocabulary, or *Thesaurus*, is used to index all PQ business, newspaper, and general reference material.

It is a natural language, hierarchical vocabulary complying with ANSI/NISO Standard Z39.19 (*Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*).

Created in the 1970s for the ABI/INFORM business database, then combined with the general reference vocabulary in the 1980s.

*It is based on Library of Congress Subject Headings, but is not structured like LCSH.*

Thesaurus subject areas examples:

<table>
<thead>
<tr>
<th>Subject area + approx. # of terms</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business, Accounting, &amp; Industry  ~4300 terms</td>
<td>Economic conditions  Recessions</td>
</tr>
<tr>
<td>Education ~450 terms</td>
<td>Teacher education   Core curriculum</td>
</tr>
<tr>
<td>Government &amp; public policy ~850 terms</td>
<td>Health care policy</td>
</tr>
<tr>
<td>Humanities ~1000 terms</td>
<td>Philosophy   Talmud   Social research</td>
</tr>
<tr>
<td>Medicine ~1600 terms</td>
<td>Evidence-based medicine   Health care industry</td>
</tr>
<tr>
<td>Science, math &amp; technology ~2600 terms</td>
<td>Materials science   Mathematical analysis</td>
</tr>
</tbody>
</table>

Total number of CV terms as of EOY 2013: 18,200+ (11,000+ are preferred terms)

Terms added each year: 75-100

Over 81,000 Cross-references (UF, BT, NT, RT)

Over 4,700 scope notes
Chronic obstructive pulmonary disease

SN: Any lung disease, such as chronic bronchitis or emphysema, causing obstruction of bronchial airflow
UF COPD
BT Disease
BT Respiratory diseases
NT Asthma
NT Bronchitis
NT Emphysema
RT Airway management
RT Lungs

- SN = Scope Note
- UF = Used For
- BT = Broader Term
- NT = Narrower Term
- RT = Related Term

- All terms are exhaustively researched to ensure accurate presentation and to produce the Scope Note.
- Terms are added based on suggestions from indexers, product managers who research the market, and customers who need to find certain kinds of information.
ProQuest Authority Files

<table>
<thead>
<tr>
<th>Authority File</th>
<th>Term Count*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corp/Org Names</td>
<td>415,129</td>
</tr>
<tr>
<td>Personal Names</td>
<td>461,711</td>
</tr>
<tr>
<td>Geographic Names</td>
<td>35,470</td>
</tr>
<tr>
<td>Product Names</td>
<td>38,498</td>
</tr>
</tbody>
</table>

- Used to index PQ Central, Research Library, and Newspaper databases.
- Corp and Pers names added to authority file & records every day. All Corp names include NAICS industry codes.
- Standardized format of company and people names goes back 30+ years when all searching was done on Dialog.
- Use of cross-references is critical to enable optimal use of autocategorization:
  - Exxon Mobil Corp
  - UF Exxon Corp
  - UF Exxon Mobil
  - UF Exxon Mobil Corp.
  - UF Exxon Mobil Corporation
  - UF ExxonMobil Corp.
  - UF ExxonMobil Corp

- Authority file lists only exist as “Look Ups” in Advanced Search
- CSA databases will eventually make use of PQ authority files

* As of EOY 2013
Add, Updates, and Revisions

- We receive data 24/7/365 from all over the world
  - Thus our indexers…
    - Add new data to various databases every day
    - Manage pre-publication data + current publications
    - Add historical data when needed
    - Manage publication details, such as title history, as needed

- ProQuest Platform Full-text Publications Indexed: 28,373 (as of 14 Nov 2014)
  - ABI/INFORM Complete: 6403 (+ A&I only = 7740)
  - ProQuest Health & Medical Complete: 2704 (+ A&I only = 3264)
  - ProQuest Education Journals: 1048 (+ A&I only = 1375)

- ProQuest Central: 21,000+ (Full-text and A&I)

- Added / Updated over 6 million records between 1 Jan 2013 and 11 Nov 2014
Life Cycle of an Article

1. Article comes into PQ via electronic feed or hardcopy
2. Article receives conversion to meet formatting requirements
3. Article logged into Journal Tracking system that tracks publication issues
4. Article receives A&I by manual or automated processes and is quality checked by Editorial
5. Article is released to online product
What gets A&I’ed?

EVERYTHING! Well, . . . sort of. . .
Magazines, academic/scholarly journals, newspapers, newswire feeds, working papers, dissertations, books, videos, and anything else we can apply terms to.

WHO DECIDES WHAT GETS INDEXED?
Editorial and product managers decide what type of content is amenable to indexing. Editorial works with Content Analysts to set up content for indexing.

HOW DO WE APPLY INDEXING?
Across the editorial teams in Louisville and Cambridge, we index using autocategorization, freelancers, in-house editors, and vendors. With non-periodical content like reports or videos, we often map a provider’s metadata to terms in ProQuest vocabularies.

WHAT DO WE INDEX?
Subjects, Company/Org names, NAICS codes, Classification Codes, People, Locations, Taxonomic terms, and many more.
How Indexing Improves Search—Example

The beauty of indexing is that we can limit a term to apply only on articles that are specifically about that term AS WE DEFINE IT.

- We do this through our thesaurus entries and by training the supporting software to use the term a certain way.

Example: *Big Bang theory*

In ProQuest Central, an exact Subject search – su.exact("Big Bang theory") – yields 2,175 results.*

Under “Narrow results by,” you can see that the top Subjects include Universe, Cosmology, Astronomy, Astrophysics…all science terms.

Now do a free-text search of “big bang theory” – you get 44,021 results.*

In Narrow by Subjects, the top listed terms are Television programs, Television networks, Actors, and Emmy awards, because there is a TV show named “Big Bang Theory” that is not about the Big Bang Theory at all.

A free-text, non-fielded search will not weed out those articles. How much time would it take a researcher to narrow down those 44,000 results to the articles he really wants?

*As of 13 Nov 2014
Traits of an Indexer

ProQuest indexers and editorial staff come from a variety of backgrounds and have a diverse set of experiences. What do they have in common?

- They like to *read* and to *learn* about new things.
- They have a *college degree in a subject* that is covered by ProQuest databases—if possible, they work in those databases.
- They have *strong writing and editing skills*—meaning an eye for detail.
- They are often *perfectionists*—even though we can’t always spend the time we want to perfect our work.
- They quickly become “*subject matter experts*” in the material they index.
- They are *flexible and adaptable* to change—whether that be in processes, systems, indexing practices, or priorities.
Training a New Indexer

- The indexer is introduced to the editorial system.
- The indexer is introduced to the principles of indexing and to the controlled vocabulary.
- The indexer begins by indexing “real” articles, which are closely reviewed and intense feedback provided.
- The indexer corrects his or her own mistakes, which accelerates the learning.

**WHAT WE KNOW FROM EXPERIENCE:**

- It takes, on average, three months for a new indexer to become comfortable with indexing and with the material he or she is indexing.
- Within two weeks you know whether someone is capable of indexing.
- No article will be indexed exactly the same way by different indexers—but there is usually at least one term that needs to be applied.
- Discussion about the indexing of a particular article is common within an indexing staff, and it benefits everyone by raising questions about indexing terminology and the terminology about the subject at hand.
In Conclusion

Abstracting and indexing remain core elements of an online search database:

- Index terms enable the search system to extract key concepts from articles and present them in results so that searchers are always confident they are retrieving relevant results.
- Despite the continued addition of new content, ProQuest continues to provide material that is indexed, through not only human methods but with automated methods as well.
- Abstracts remain important as a means of quickly informing searchers of the relevance of an article to their research.

Future indexers need more than ever to:

- Understand search methodologies.
- Understand ways in which indexing of content can be automated.
- Understand the structure of controlled vocabularies, authority files, and taxonomies.
- Be aware of the many types of thesauri that exist in various disciplines.
Thank You … and for more information …

• More information and videos are available at the ProQuest LibGuides page
  http://proquest.libguides.com/

• Contact Today’s Presenter:
  **Aimée Leverette, M.S. Info. Sci., M. Educ.**
  ProQuest Training & Consulting Partner – U.S. West
  Fort Collins, CO | aimee.leverette@proquest.com | 303.328.1589

• Attend a *Live Webinar* or View a Recording
  http://bit.ly/1xnuOLg

• ProQuest Support Center
  http://support.proquest.com