June 2006

Predicting High-Circulating Titles for Public Libraries

Bob Molyneux

SirsiDynix, bob.molyneux@sirsidynix.com

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Molyneux, Bob (2006) "Predicting High-Circulating Titles for Public Libraries," *Against the Grain*: Vol. 18: Iss. 3, Article 17.
DOI: <https://doi.org/10.7771/2380-176X.4965>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Country Western and Line Dancing.

Country western dancing includes a number of traditional and new dances associated most commonly with Country western music. On the other hand, **Wikipedia** defines Line Dancing as "formation dance in which a group of people dance in a line formation or in lines, and they all execute the same dance moves individually." Although some of it's 1990s disco faddishness is gone, Line Dancing is still a popular type of dancing worldwide as is Country Western Dancing. (See also Square Dancing and Clogging.)

<http://www.cwddance.org/> — While not an extensive site, this is the home of a major C&W Dance association.

<http://xenon.triode.net.au/~dragon/ldance/history.html> — A really well-written history of line-dancing.

<http://geocities.com/Heartland/Prairie/7415/> — Part of a larger dance site, <http://geocities.com/Heartland/Prairie/7415/> sponsored by the **United Country Western Dance Council** this page is valuable in that it contains videos of the main line dances done at every level of proficiency.

Clogging: High stepping fun that's not just for hill-folk! This folk dance, like Square Dancing, came from English and Scottish roots. While principally associated with Appalachia and Bluegrass Music, Clogging has become popular worldwide.

<http://members.aol.com/mdevin/clogtext.html> — Part of a California Clogging Website, this page will provide the patron with a history of this dance style.

<http://www.clogdancing.com/> — A major site for all manner of information on Clogging.

<http://www.doubletoe.com/> — A magazine exclusively devoted to clogging.

Exotic Dancing: G-strings and Gyration? Nope, not in this article! Ok, I'm just old-fashioned, so you'll need to find these resources on your own. Exotic Dancing is clearly a dance subculture with ancient origins and many modern manifestations and variations but, I did not feel that most librarians would be comfortable exploring the Websites I could provide by themselves or with a patron so, I left this one to your imagination. Just go to **Wikipedia** and type in "Exotic Dancing."

Latin Dance: Salsa is more than a Condi-ment. From the Tango, to the Rumba, to the Salsa, Latin Dancing has taken the world by storm with its sultry moves and exciting music.

<http://www.salsacentro.com/> — A well-organized site run by a professional dance teacher devoted to the Salsa style of dancing that has its roots in the Afro-Cuban tradition.

<http://www-staff.it.uts.edu.au/~don/pubs/latin.html> — A site from the **University of Sydney** offering an excellent history of Latin Dance.

<http://www.latin-dance.net/> — Excellent site for dance club access.

Morris Dancing: An ancient form of

Street-Dance with Bells, Sticks and Hankies. The origin of Morris Dancing may be lost in the mists of history but this traditional folk dance of the British Isles is alive and well today.

<http://www.morrisdancing.org/> — A major connection site to worldwide Morris Dancing.

<http://www.ucolick.org/~sla/morris/whatisit.html> — A small page with history and a link to a great bibliography

<http://u.webring.com/hub?ring=morris> — Just what it seems, a Web-ring on all things Morris.

Slam/Punk Dancing: Making the most of pain in the Mosh Pit. <http://members.aol.com/rik0lar/moshing/mosh.htm> — This is a good introduction to the "Mosh Pit" world.

<http://www.gothpunk.com/howtos/how-to-dance-punk.html> — This site teaches the fundamentals of punk dancing. Ouch!

Square Dancing: Dosado to your Heart's Content. Square Dancing evolved out of the structured Morris dances of the British Isles and France into its own distinctive dance style.

<http://www.dosado.com/> — This extensive site is well-organized and will connect the patron with Square Dance Clubs worldwide and of every possible variation.

<http://www.square-dancing.com/> — Another site with connections to local Square Dance clubs.

<http://www.squaredancefestivals.org/> — Just what it sounds like: Square Dance festivals.

<http://callerlab.org/> — Square Dancing is dependent on a caller to instruct the dancers as to the sequence of moves. This site an international Association of Square Dance Callers.

<http://www.valleytrailers.org/terminology.html> — Part of a Los Angeles dance club site, this page is about square dance terminology.

<http://squaredancing.com/> — This site links the patron to all manner of Square Dancing products and services.

Swing Dancing: the Forties lives on in Dance. <http://www.centralhome.com/ballroomcountry/swing.htm> — A great history of Swing dancing.

<http://www.westcoastswingamerica.com/> — <http://www.bigcityswing.com/> — <http://www.atlantaswingdancers.com/> — Three Regional clubs with links to resources.

http://www.brown.edu/Students/Swing_Club/videos.html — A good source for online "learn to swing dance" lessons.

Tango: the Dance of Passion. <http://eng.tango.info/> — An extensive site with loads of links to everything related to the Tango.

<http://members.ping.at/kdf-wien/tango/> — Another site with an incredible list of links related to Argentine Tango.

Tap Dancing. <http://www.tapdance.org/> — A major site on the world of Tap Dance with every sort of link. Did you know there is an official tap dance day on May 25th?

<http://www.usatap.org/> — The **National Tap Ensemble's** comprehensive site which is highly recommended.

And finally:

Canine Freestyle Dancing: Tango with your Terrier. When a dear friend and colleague told me about freestyle dancing with her dog, I was amazed. Dog dancing is, however, recognized as a sport by the prestigious **Westminster Kennel Club** with demonstrations at their events. Beginning in the late 1980s, to date there are 17 official clubs all over the world and an estimated 7,500 human practitioners with at least an equal number of dogs.

<http://www.worldcaninefreestyle.org/index.html> — The **World Canine Freestyle Organization, Musical Dog Sport Association and Canine Freestyle Federation, Inc.**

<http://www.musicaldogsport.org/index.htm> — <http://www.canine-freestyle.org/> — These are large organizational sites with the history, events and links to learning materials. 🐕

Predicting High-Circulating Titles from page 87

rials. It is not easy to imagine how this test could be set up, but if it could (or even should) be implemented, it would allow us to further disaggregate user experience to build an even stronger understanding of who checks out what and why.

The Future

It will be profoundly useful to predict which books will circulate at a given library. Of course, not all of a public library's collections are expected to be high-circulating items, as the example of reference and *Hamlet* above make clear but what is purchased to circulate should circulate.

We're trying to find new models to predict

how to spend our dollars in the best interests of our users. There is even potential here to discover the future of collection development and, indeed, invent it, as well as a fiduciary responsibility to spend public money responsibly. That's a great thing. Toward those ends, we will continue to dig deeper into these big questions as we discover the ultimate potential of the **Normative Data Project**. 🐕

Endnotes

1. <http://www.spss.com/clementine/>
2. http://www.sirsidynix.com/Resources/Pdfs/Company/Abram/20060321_CIL.ppt
3. **John Pruitt and Jonathan Grudin**, "Personas: Practice and Theory," <http://research.microsoft.com/research/coet/Grudin/Personas/Pruitt-Grudin.pdf>.

Review Essay: from page 86

tal sculpture of Easter Island resembles sculpture of Central and South America more than the sculpture of the rest of Polynesia, and South America is closer with favorable ocean currents in between. Archaeological evidence has since demonstrated that the initial occupation of Polynesia began over two thousand years ago, providing time for the unique features of its culture to have developed in isolation with influences from the east and west. It is certain from ocean currents that the Maori came to New Zealand from, central Polynesia, and the two groups had probably been separated for about six centuries. Little else is certain, and it is extraordinary that after settling most of the Pacific Oceans, the Polynesians are not definitely known to have colonized either Australia or South America. Genetic studies promise to provide additional evidence, but all of the types of evidence will have to be fitted together into a coherent whole for the problems relating to the origin and development of Polynesia to be wholly solved, and among the most important evidence will always be the systematic information that was first provided by the Cook expeditions.

Cook's own accounts are the most comprehensive of all aspects of the expeditions as well as the most incisive and readable, and his original logs and revised versions of them have survived, which is fortunate since their initial publication involved poor selection and inaccurate editing. The definitive edition based on the manuscripts is the four volumes with a portfolio of maps edited by **J. C. Beaglehole** for the Hakluyt Society from 1955-1967, but as essential as this edition is for reference, it includes more by **Beaglehole** and others than by **Cook**, making it difficult to use and almost unaffordable. No edition includes all that **Cook** wrote and only what he wrote, and a publishing opportunity exists for a complete edition of his text. It needs nothing more, but some well chosen illustrations made during the voyage, some photographs of artifacts that were collected, and a few maps would make a more salable book of perhaps two volumes. All of his text could fit in a single volume if India paper were used and if sailing directions and weather information were printed in small type. A separate volume deserves to be printed with his anthropological accounts for use as a student textbook.

Humboldt

Cook's expeditions inspired many others, and among the most productive was the five-year expedition of **Alexander von Humboldt** and **Aimé Bonpland** to South America. **Humboldt** personally knew **George Foster**, a German naturalist who with his father, **Johann Reinhold Foster**, had accompanied **Cook** on the second expedition and who incorporated much of his father's journal into an unauthorized, but valuable account of the natural history and cultures encountered. **Humboldt** was so deeply impressed that these accounts changed the course of his life.

Humboldt's father prepared him for a career in government, and his earliest studies included economics, political science, and modern languages, knowledge that **Humboldt** later put to good use during his travels, but he was more interested in biology and travel. When 18, **Humboldt** decided he wanted to explore tropical regions of the world, and he began to prepare himself thoroughly. He studied all aspects of science, traveled widely in Europe, and worked as a professional mining engineer with the Prussian Mining Service from 1792-1796. At age 27, he became financially independent, and he hoped to accompany a French expedition to the Pacific, but when the Napoleonic Wars prevented the expedition, he went to Spain and obtained permission to travel throughout South America and Central America. He was a baron whose father had been an official in the German court, and the King of Spain, Secretary of State, and Council of the Indies approved his plans and gave him passports to travel throughout Spanish possessions and instructed local officials to cooperate. He hoped eventually to join the planned French expedition when it reached South America, but when that expedition never materialized, he expanded his travels from northern South America to include Cuba, Mexico, and Peru and consequently extended his trip to a total of five years.

Humboldt was the first naturalist and the first and last foreigner to travel widely in Hispanic America shortly before it ceased to be a Spanish possession, and he made full use of the opportunity. His record of the ways of life, economies, and governments he encountered in many cases provide the best summaries ever made before Central and South America were divided into separate nations, and his travel account was widely published and read for what he recorded about the geography and economies

of the Americas as well as natural history.

At his own expense, **Humboldt** financed his expedition and persuaded **Aimé Bonpland**, a professional botanist, to accompany him. He purchased about forty scientific instruments including a barometer to measure heights, an excellent telescope to examine the southern sky, and microscopes. He carried with him a library of natural history books. To travel in interior areas with so much equipment, he hired numerous pack animals, guides, and interpreters for the Indian languages.

Humboldt and **Bonpland** explored primarily in what is now four countries: Venezuela, Cuba, Peru, and Mexico. Everywhere they went, they recorded information about biology, geology, geography, anthropology, economic

continued on page 88

The APS Journal Legacy Content

OVER 100 YEARS
of Historical
Scientific Research

14 Journals of the American
Physiological Society
www.the-aps.org

Rumors from page 85

through improved self-management and care. **Dr. J. Edward Hill**, President of the **American Medical Association**, made the presentation to **Barbara Carlson** and a delegation of the Coalition's leadership at a reception following the **2006 Libraries and Health Information Forum at the National Library of Medicine** in Bethesda, MD. **REACH 2010** was presented with a plaque and \$20,000 for continuing and expanding its programs. Congratulations, **Bobbie!**
www.nclis.gov/news.htm

And, ta da! **Saving the best for last!** The inde-

fatigable, energetic, and awesome **Jack Walsdorf** will have a new book in the fall — *On Collecting William Morris: A Memoir*. I have seen the prospectus of this engaging account of **Jack's** experiences of collecting **William Morris** again, and again, and again. The book will be published by **The Printery**, a private press founded by **Kay Michael Kramer** in 1970. **The Printery** produces works of historic and scholarly interest on a variety of topics. More soon!

In the meantime, happy **SLA**, **MLA**, and **ALA** and happy summer. And, be sure and register for the **Charleston Conference** November 8-11.
<http://www.katina.info/conference>

Predicting High-Circulating Titles from page 92

ing of users that drive decision-making for development projects."^{2,3} They allow us to study and understand the different motivations of users and to tailor our services to those different users.

Given that **Personas** provides us another way of looking at users, it may be that we could use them to study actual use if we have volunteers that would let us track how they use mate-

continued on page 84

Desperately Seeking Copyright — Copyright Permissioning Permeates Library Operations

by **Edward Collieran** (Senior Director, Rightsholder Relations, Copyright Clearance Center, Danvers, MA) <ecollieran@copyright.com>

Re-use of published content is central to the life of colleges and universities, and until recently most people on campus relied on librarians to provide both the content and permission to reuse it. But Internet technologies have placed increasing power in the hands of faculty and staff to access and distribute published information. And just as information access has moved beyond the library staff to faculty and administrators, so too has the responsibility for securing copyright permissions.

So we've seen a two-step process in the development of information and library applications and Web services that succeed in the academic market. First, the new product is launched. Then, after its effectiveness and popularity have been proven, vendors integrate copyright permission capabilities within the product.

Most vendors and academic institutions realize that applications which make it easy to find, combine and share vast amounts of copyrighted information must also make it easy to respect the copyrights of publishers and authors. That's why many application vendors and Web services providers work hard to add copyright functionality that holds users within the product's workflow that discourages them from moving to other applications to meet their needs. Delivering automated copyright capabilities supports at least two core objectives of product designers: promote customer product loyalty and enhance customer satisfaction.

The result: colleges can better ensure campus-wide copyright compliance, and fewer library staff resources are spent on the time-consuming and tedious task of securing permissions.

Lets review several significant examples:

Course Management Systems

The **Blackboard Academic Suite** is used by more than 1,200 academic institutions, and last year the **Copyright Permissions Building Block** for the **Blackboard Learning System** was made available. It enables faculty and staff to quickly obtain licenses to post copyrighted text-based content within **Blackboard** when developing electronic course materials. This integration also allows permissions to be routed directly to the librarian, who can apply fair use guidelines or check for existing re-use rights held by the institution before securing the permission. Other course management system vendors are working on similar solutions.

Coursepack Automation Systems

Coursepack production software, such as **Xerox FreeFlow** and **Alto Imaging's ADP** application, allow copy shops and print services to obtain permissions when assembling coursepacks for the printing. Copyright requests are submitted through the application. As the compiled document is assembled, the system displays permissions and royalty fees, and saves the bibliographical information, making it easier to re-secure permissions for subsequent reprintings.

Abstracting and Indexing Services

Scopus is the largest abstracting and citation database of research literature. It recently added copyright licensing capability directly within the search results, so faculty and librarians can quickly get permissions upon finding the content they need.

In all of these ways, copyright permissioning has become ubiquitous within automated library and information delivery applications for the academic world. As Internet and technology solution vendors make content acquisition easier, I expect the trend of placing copyright permissions at the point of content re-use to gain even greater momentum.

"... Internet technologies have placed increasing power in the hands of faculty and staff to access and distribute published information."

Library Applications

Electronic reserve systems, such as **Docutek ERes** and **Atlas Ares**, and interlibrary loan systems, including **OCCLC ILLiad**, have integrated copyright permissions that enable librarians to secure copyright licensing as they prepare content for e-reserve posting or, in the case of interlibrary loan borrowing, after the content has been received.

Link Resolvers

Major link resolvers — **Ex Libris SFX**, **SirsiDynix Resolver** and **Innovative Interfaces WebBridge** — offer librarians and staff the ability to get copyright permissions once they've located text they need in e-reserves and other course materials.

Edward Collieran is the Senior Director of Rightsholder Relations at Copyright Clearance Center (CCC), Danvers, MA, a provider of copyright licensing and compliance solutions. He has nearly 25 years experience in the communications and information industries.

Review Essay: from page 91

All of these travelers produced classic travel accounts, and all of them wrote extensively about the indigenous peoples they encountered as part of the natural histories of the areas they explored. In most cases, the cultures they recorded were changing so rapidly that no better records could later be made. The challenge now is to interpret the best available information without preconceptions.

*Gene Waddell is College Archivist at the College of Charleston and author of **Charles-***

ton Architecture, 1670-1860 (Wyrick & Co., 2003). This is the third of a series of three articles about travel accounts and their value for anthropology and as works of literature. The first was "American Life in the 19th Century: Unabridged Travel Accounts by Audubon, Olmsted, and Twain" (ATG, Jun. 2005) and the second "In Indian Territory" (ATG, Nov. 2005). All three articles were based on a lecture given at the Charleston Library Society on 17 Feb. 2005. He has written two earlier articles about architecture for ATG. All five articles have emphasized the neglected potential of primary sources for research and teaching and the need to make the best possible versions more widely available.

Predicting High-Circulating Titles from page 93

However, it may be that by using one of these pieces of software, we could discover even more attributes of high-circulating titles and then begin to build a sophisticated model that could examine new titles as they came out — and predict which will circulate at what kind of library.

Version 3?

Stephen Abram is leading the **SirsiDynix** study of library users in his **Personas** project. **Personas**, as he defines them in the talk he gave at **Computers in Libraries** this year, "... are hypothetical representations of a natural group-

continued on page 87

Predicting High-Circulating Titles for Public Libraries

by **Bob Molyneux** (Chief Statistician, SirsiDynix, 101 Washington Street S. E., Huntsville, AL 35801-4827; Phone: 256-704-7019; Fax: 256-704-7007) <bob.molyneux@sirsidynix.com>

At the last **Charleston Conference**, I discussed our work on predicting circulation patterns of library materials based on data from the **Normative Data Project (NDP)** (<http://www.libraryndp.info/>). We are attempting to predict which titles will have high circulation and which will circulate little. Using that knowledge, acquisitions librarians in public libraries would have more information about what to purchase and what not to purchase. This article gives a short overview of the talk and update where we are now. It turns out the problem is a bit more complex and the value of such a capability greater than we understood at the time.

The **Normative Data Project** is a huge pool of potential discovery. It collects data on various aspects of U.S. public libraries. For example, it includes data from the **U.S. National Center for Education Statistics**, geographical information system (GIS) data from **GeoLib** (<http://www.geolib.org/>), and **U.S. Census** data on all American public libraries. In addition, there are detailed data from the integrated library systems (ILS) of a smaller group of "contributing" libraries. For these libraries, we have additional information on holdings and circulations. We have such information as **Dewey** and **LC** classes, class of borrower (adult, juvenile, etc., but **no** personally identifiable information!), day and time of day of the circulation, title, and format such as book, DVD, and so on).

The first understanding of the problem as discussed in Charleston is: With the information collected in the **NDP**, can we predict titles that will have high circulations? Such a capability would be useful in spending end-of-year money to maximize circulations or in building new public library collections as is done with "opening day collections." Currently these collections are built based on prescriptive lists. But if we can predict which titles will circulate the most, we can be relatively certain that materials bought to be a part of the circulating collection will, in fact, circulate.

Version 1

Right now, using the **NDP**, we can do a good job of predicting a set of titles that will circulate. Demographics of the library's community affect how the library is used. We can match the demographics of a target library with those from the set of contributing libraries based on default market areas around each of the 17,000 public libraries (that is, systems and branches) in the United States. This comparison will identify a set of libraries similar to the target library and reveal what has circulated at those libraries and produce a bibliography of titles and formats that will be sought and tailored to that library's users. We can do this comparison by any of the many dimensions of the data, whether format, type of borrower, **Dewey** class, and so on.

Although this capability is an advance over other methods, there are currently two key limitations to this approach.

One is that a public library will have materials that do not circulate a great deal either because they are reference or because the titles are important for other reasons. For instance, most public libraries will own *Hamlet*, in spite of the title's circulation patterns.

In these cases, we would need to analyze holdings information on the contributing libraries like the target library to discover which titles are in the non-circulating collection and which titles are in the low-circulating collection that still would be important to own. From the holdings at similar libraries, we could build a virtual "core collection" of such items that is based on what *is owned* at like libraries and not from bibliographies of what libraries *should own*. It is hard to say what such core collections will look like. It is reasonable to speculate that most public libraries will own *Hamlet* but only the larger ones will own **Shakespeare's** lesser play, *Cymbeline*. It would seem that the size of the library's population served is an important variable for these both reference and low-circulating titles.

The other limitation concerns the time it takes to compile the **NDP** data because that delay affects the titles that are available. If the **NDP** is three months behind the current titles, any new title from **Stephen King** or **JK Rowling** since the **NDP's** latest compilation will not appear in a list of high-circulating titles although, clearly, such items will circulate. So, such a collection would always miss the latest titles as, of course, would occur with published lists.

Version 2

This latter limitation forced a rethinking of the whole matter, and it is clear that the problem as presented in Charleston — predicting titles that will circulate — is incorrect as stated. The correct way to state the problem is that we should predict the attributes of the titles that will circulate. Authors would be such an attribute in some cases — as the two cited above make clear.

Another probable attribute is the user population's demographic characteristics, which we can estimate in the **NDP**. For instance, one **NDP** user discovered in a set of libraries he examined (where the population had a relatively low percentage of high school graduates) that titles about completing a high-school equivalency were high-circulating items — but these kinds of materials did not turn up in lists of high-circulation items where the population had a high percentage of college graduates. It is scarcely headline making news that a population where many have not completed high school would want more books on high-school equivalency from their library than another library's population where all have completed high school. It is another thing, though, to demonstrate such a relationship with data. That demonstration is evidence that demographics will provide a useful set of attributes in a predictive model.

We might imagine other such attributes, but there would be those we would miss if we set out to list them and, in any case, data mining provides us a method to discover attributes that are both obvious and those that are not obvious. Essentially anything we know about the demographics, past circulation patterns, and the types of books that have circulated may be useful in telling us which attributes of titles will be useful to predict high-circulating titles.

The **NDP** runs on **Microsoft SQL Server**. The new version, **SQL Server 2005**, will have a "predictive analytics" capability based on **SPSS's Clementine**¹ and allow us to test whether this approach will be useful.

If this "predictive analytic" capability yields useful information, then the next decision would be whether to order the full-blown version of **Clementine** or a similar piece of software from another vendor. These applications are expensive and such a step will not be taken lightly.

continued on page 92

Back Talk from page 94

our embracing of the library friendly publishing alternative. I realize that this paragraph can get me in personal trouble so let me assure you all that I do not jump up and down with glee when our acquisitions department sends its annual million dollar check to the Netherlands. But, I am equally non-enthusiastic about the need to subscribe to both library friendly and expensive journals for the same subjects. Consequently, I cannot salute a new call for the subsidy of the publishing enterprise. We already gave to the **SPARC** effort. For me, my job is to give

people reading material and reading material costs money. Moreover, I won't feel better, if in the future, library book budgets are reduced so that the funds can be transferred to the faculty who have to pay the page charges associated with the open access movement.

So, I think the ideas associated with digitizing past and present materials, to preserve our printed legacy, and to negotiate with electronic publishers in a transparent and ethical manner are all good goals/challenges. But the goals to return to the past and resurrect core collection lists in today's electronic environment and to get everyone to contribute to a parallel open access publishing empire are both whacky ideas.