

Using the Data You Have to Create New Data for New Decisions

Christie Koontz

PhD, Associate in Research, School of Information, Florida State University
Tallahassee, Florida USA
ckoontz@fsu.edu



Copyright © 2014 by **Christie Koontz**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

Abstract:

Libraries, gather data on a daily basis to present answers, build knowledge, or solve problems for customers. Why not direct these same processes of research and fact-finding toward organizational problems? Organizations such as libraries hold and have access to a wealth of data that can reveal the needs and interests of customers. Having an estimate of population-served based upon geographic market area, provides accurate basis for creation of a wide-range of valuable customer data.

Introduction

On a daily basis we (librarians) gather data. We do a great job of collecting, organizing and selecting data of every kind to solve every kind of problem for our customers. In the current day climate we must turn these skills upon ourselves, to achieve better management. Why? Libraries (and other nonprofits) specifically need information about their customers to remain competitive and successful. In fact, any agency facing competition must prioritize customer data collection. Customer data allow managers to more effectively market i.e., provide the best product at the best price at the best place with the best communication to promote products. *Changing* customer characteristics and information needs must be identified regularly. Dynamic and ongoing customer research is crucial to provide library services and materials that meet complex needs in ever-changing communities.

Critical Customer Data is Often Lacking

Despite these critical needs, many US library managers may find themselves lacking relevant customer data due to: 1) lack of legal mandate to collect specific types of customer data, resulting in fewer employees dedicated to this task (by contrast with schools, fire and emergency services); 2) inherent and traditional reluctance based upon customer privacy issues; 3) lack of training; and 4) reticence to ask for available customer-related data from local planning departments. Overall data gathering for many librarians requires extra initiative, effort, focus, and sometimes funding. On the latter point—cost--data collection and analysis are the most expensive activity in the marketing research process.¹ So as the need for accountability grows, and budget fights become more intense amongst and between public agencies, accessing data that is available is cost effective and efficient. Prioritizing what data to access is of utmost importance. The first statistic needed in order to know the very most about your library's productivity is an estimate of *population served*. The US Census Bureau offers population data at many governmental boundary levels and is easy and inexpensive to obtain. This paper discusses this statistic and related statistics that can be accessed or developed to optimize reporting and planning and decision-making.

Using Population-Served Data to Describe Market Area

A key distinction in determining the value of the *population-served* statistic, is whether it is drawn from the estimated *geographic market area* (GMA), which is defined as the actual geographic area that customers and potential customers are drawn from, or from the traditional *legal service area* which is the geographic area *ascribed* to a library from which it receives its funds, i.e. a county, city, school district, etc.² At this time all federal public library statistics in the USA are agglomerated and iterated by legal service area. Yet this leads to inaccurate statistics for the performance of a single branch (which is like a franchise.) For example let us say library system A has a legal service area of 150,000, which is the county population. (The county funds the library so its boundaries are used.) The library system A has seven branches, yet there is no market area determination for any of the seven branches. So the parts are judged by the whole. There is no way to dissect or distinguish a branch's performance if no population-served numbers are estimated. Currently this data, GMA estimation, is not calculated or used by the majority of USA libraries due to the widespread acceptance of the legal service area as *population served*.

Yet identification and mapping of the boundaries of the library's geographic market area are the first steps in finding out *population served* and the answers to these important marketing research questions-- who our customers are, what services they want and need, at what best cost can we offer them, and where can these services and materials be best accessed. There are simple methods that can be used to estimate market areas to get more precise performance statistics for a single library (branch).

First and ideally as stated earlier, the geographic market area should be the area from which the library draws most of its users, not an assigned area. Yet identification of true geographic customer boundaries is made difficult due to the facts discussed above having to do with data collection in general, i.e., public library managers have not had or demanded the available methodologies or technology to survey and establish customer residence, and do not have the predilection to go against the traditional acceptance of the legally ascribed service areas--as market areas. An accurate method used by retailers is to plot customer addresses (this could be from circulation or registration, to view dispersion.) Other methods of ascribing market areas include: assigning library outlets certain census tracts; using existing local government planning zones; designating a simple radius as the library's estimation of users served; or using the experiential advice of staff.³ The problem of course with these approaches is that one, people may cross census tract or zone lines and use whichever library they choose (and they do!); and two, a rather weak assumption is being made that the majority of users that live within the designated area, use the library.

Designation of the geographic market area (which now of course may include identification of remote users as well) is imperative before employing any marketing research activities. For example and simply, how would a library calculate any per capita service for a branch within the above library system A?

For traveled to services such as a fast-food restaurant or a library, estimating and profiling the geographic market area creates a firm basis for identification of customer characteristics. Just as McDonald's would not assume that individuals within certain census tracts use the restaurants that are selected by McDonald's management--nor should librarians assume customers use libraries preordained by library management or local government. Assuming this data (which is based upon data widely available) are calculated, much can be reaped.

Sources of Data: External

With the statistic *population served* in hand, some of the most important external customer-related data are the demographic description (population characteristics) of the people served, can be appended. Demographics relevant to public library use may include categories such as age, family life cycle status ethnicity or race and education levels,⁴ and linguistic isolation.⁵ A boon to USA library managers desirous of collecting demographics is the recently released 2010 US Census Data which is accessible and widely available.⁶ All local planning departments have this available for public agencies similar to the public libraries. Demographic data supports equitable library service provision by facilitating discovery of potential customers and un-served or underserved markets, ultimately increasing the quality of life for individuals and society. Census data combined with library use data can offer valuable insights. For example, certain groups such as the elderly, juveniles, lower income or minority or immigrant groups—need libraries within their path. Not only can something as obvious as distance or lack of

transportation be a prohibitive factor to library use, but also more subtle cultural barriers can be identified, such as the beginning or the end of a perceived neighborhood. Some groups do not and will not travel beyond that invisible line.⁷

Important data we need may be right under our proverbial nose. These are some secondary data sources often available through libraries (print and online) for research:⁸

- federal government
- provincial/state governments
- statistics agencies
- trade associations
- general business publications
- magazine and newspaper articles
- annual reports
- academic publications
- library sources
- computerized bibliographies
- syndicated services

Internal Public Library Data

Parallel to collection and analysis of customer data from the public library's external environment --the library manager has a gold mine in the internal records which may include address or zip code information and basic population characteristics, as well as type and level of use.

Library inputs derived from internal circulation reports are primarily number of materials circulated, and quantities and types of materials by age. Other internal data records include number of staff, hours open, number of registrants, number computer terminals, and square footage, etc. Traditionally these were the only counts used to assess a library's success. It was only in the early eighties that the library profession developed and implemented the concept of 'output measures'--which rather than simply describing what was in the library, now measures the library's performance to people outside the library (i.e., the public, local government, state library.) These output measures frequently include per capita measures of circulation, reference transactions, document delivery, visits, in-library use, and more recently web hits, number of Internet access hours, and e-reference transactions. The *per capita* count is more accurate if based upon the *population-served* estimate of the market area—not the legal service area.⁹

Counting All Library Uses

Whenever certain uses are difficult to count (i.e., often tied to what the software will allow and what library management demands), library systems often omit these statistics in the use report, relying solely on circulation. For example, a

library with high circulation is less likely to have funding cuts as the circulation figures traditionally represent high levels of use by customers.

And by contrast libraries with low circulation, and higher levels of ‘un-counted’ uses--such as in-library use (including computer uses) by contrast, may appear poorly used, thus more vulnerable to funding cuts. More importantly, research indicates that libraries in lower income and majority minority markets often have lower counts of circulation.¹⁰ This is now increasingly important, as these minority markets become the plurality in the USA. America’s library managers must count all the uses to assure all customers’ needs are met and an accurate profile of customer use is available for each library outlet. May we illustrate?

Just as McDonald’s would not consider counting only the hamburgers that are sold at the drive-through window, (eliminating those consumed in the restaurant from its total sales), neither can a public library afford to count only books that circulate as this would eliminate the count of important uses such as materials used in the library, librarians’ assistance in resume writing and science projects, and computer instruction. So with the GMA e.g., in hand you can have per capita computer instruction.

Customer Data from the Web and Social Media

A recent study of U.S. public libraries indicated that physical visits to library buildings decreased from the prior year for the first time in ten years. However, “visitation remains strong” with a 32 percent ten-year increase overall. Further, the average American visited a public library 5.3 times a year, a ten-year increase of 21.7 percent. Public access computer use continued to be one of the fastest growing services in public libraries.¹¹

Yet we know customers are not only visiting the physical locations of our libraries, but are also interacting with the “virtual” organization through websites and social media. For some users, the online databases and collections are most commonly used, or even the *only* way that they interact with the organization in their everyday life. How, then, can an organization determine geographic market areas or other demographic information for those customers who choose to visit the organization only online?¹²

Visitors to the organization’s websites leave virtual “footprints” in the secondary data of web server logs that capture electronic traces of their travels throughout the organization’s websites—including how the user arrived at the site (e.g., referring page or search engine, keywords searched), as well as time and date of visit, pages visited, length of time spent on a page before leaving the site (the “bounce rate”), and other details such as their web browser type or use of a mobile device. Geodata is also captured through the user’s IP address—a series of identifying numbers that can be geocoded to a location for the user’s computer or Internet service provider.¹³

Librarians have geocoded IP addresses from server logs to discover, for example, how often users access the library's chat virtual reference services from computers in specific on-campus buildings such as dorms, computer labs, and university offices, or even from public computer workstations within the library itself.¹⁴ This geo-data on virtual users can help in mapping the customer market area for the organization's online users. Both free open source¹⁵ and paid software exists for tracking and analyzing web server log data. If the organization does not host its own website and is unable to obtain server log data from its service provider, alternative ways of tracking data also exist, such as using Google Analytics¹⁶ for capturing and assessing how users interact with the organization's website and blogs.

If the organization uses social media sites such as Facebook to interact with the public, additional varieties of user geographic and demographic data are available.

All sources for data (not just web and social media data) will have limitations, which the staff should take into account in assessing results. Data may not reflect all possible user activity but only some activity within a given period; also, not all of those engaging with the library physically or virtually are necessarily customers—some may be other stakeholders such as competitors, vendors, employees, and colleagues in the larger professional community.¹⁷

Research in Action

In the past two decades two USA national projects were put forth to offer databases for free and public access to public librarians for developing geographic market areas and demographic descriptors. These are the US Public Library Database project located at www.geolib.org, and the recent iMapLibraries project www.imaplibraries.org.¹⁸ The latter project was used in a fight to save 22 libraries from closing. Here is the story.

In the summer of 2013, the Miami-Dade Public Library System (MDPLS) in Florida was confronted with a sudden funding crisis. The Mayor and County Commissioners proposed closing nearly half of the system's 49 public libraries.

The news of the Miami-Dade community potentially losing nearly half of their local libraries raised questions about the resources, programs, and services that would be lost at these local branches as well as the potential impact of the closures on the people living in the local neighborhoods.

As the news broke of the potential loss of 22 local libraries in Miami-Dade, the iMap libraries imaplibraries.org project mapped an estimated GMA for each local public library branch within MDPLS using a one-mile and two-mile radius for each library branch. The one- and two-mile radius approaches were used because mapping studies in the past have commonly used service areas of one to two miles

as a maximum distance that people are assumed to be willing to travel to reach a local library.¹⁹ While this is one method of mapping a GMA for planning purposes, and as discussed earlier, it is certainly not the only way.

Demographic reports on the 22 MDPLS library branches were developed. Social media with Facebook, Twitter, and LinkedIn was also used to share the market area descriptions. Many of the MDPLS library outlets served “majority-minority” local communities with over 50% Black residents in the immediate market area while over 50% of the residents are Hispanic in the library GMAs. Two of the libraries to be closed, served areas in Miami Dade’s Haitian community that had the highest levels of linguistically isolated households within the entire county at over 10 percent. Linguistic isolation reflects those who speak a non-English language at home and who also report that they do not speak English very well. Libraries provide local communities with free access to resources in their languages, as well as resources for learning English and for acculturation.

This type of GMA data analysis using mapped library locations with demographic data of the people in the surrounding community becomes even more powerful in telling the story of how libraries impact people’s lives when combined with the library’s own specific data usage reports.

This real-world case study demonstrates how integrating geographic market area data for each library branch with the demographic data of the people living within that area, and including branch-level library usage data to show how people are actually using the resources, programs and services at each local branch library is valuable and powerful. Being able to combine advocacy efforts with hard data that shows the impact of libraries on people’s lives can be even more valuable and useful to public officials, opinion leaders, decision makers and the community, especially if the data can be widely disseminated within a critical timeframe, such as through blogging and social media. Public opinion is critical for libraries, as most U.S. libraries operate based on local funding provided by taxpayers in the community.

Summary

Libraries, gather data on a daily basis to present answers, build knowledge, or solve problems for customers. Why not direct these same processes of research and fact-finding toward organizational problems? Organizations such as libraries hold and have access to a wealth of data that can reveal the needs and interests of customers. Having an estimate of *population-served* based upon geographic market area, provides accurate basis for creation of a wide-range of valuable customer data.

¹ Marketprofs.com, “Topic: Research Metrics,”

http://www.marketingprofs.com/ea/qst_question.asp?qstid=8292

² Christie Koontz and Lorri Mon, *Marketing and Social Media: a Guide for Libraries, Archives and Museums*, (Lanham, MD: Rowman and Littlefield, 2014), 81-82.

³ Ibid.

⁴ Christie M. Koontz, *Library Facility Siting and Location Handbook* (Westport, CT: Greenwood Press, 1997), 31-61.

⁵ *Serving Non-English Speakers in US Public Libraries: 2007 Analysis of Library Demographics, Services, and Programs*, ALA Office for Research and Statistics, Office for Literacy and Outreach Services Office for Diversity, Public Programs Office, <http://www.ala.org/offices/olos/nonenglishspeakers>

⁶ “2010 Census Data Product Descriptions,” U.S. Census.gov, http://www.census.gov/2010census/news/pdf/data_products_2010_census2.pdf.

⁷ Koontz, “Library Facility...” 31-52.

⁸ “Research Using Secondary Data Sources,” Steppingstones Partnership, Inc, December 2004, http://www.steppingstones.ca/index.php?option=com_content&view=article&id=64:secondary-research&catid=17:research

⁹ Koontz and Mon, “Marketing...” 79.

¹⁰ Christie Koontz, Dean K. Jue, and Keith Curry Lance, “Neighborhood-Based In-Library Use Performance Measures for Public Libraries: A Nationwide Study of Majority Minority White/Low Income Markets Using Personal Digital Data Collectors,” *Library and Information Science Research* 27 (2005): 28-50

¹¹ “IMLS Issues Public Libraries in the United States Survey FY 2010,” <http://www.examiner.com/article/imls-issues-public-libraries-the-united-states-survey-fy-2010>.

¹² Koontz and Mon, “Marketing...” 86-88.

-
- ¹³ Lorri M. Mon, Bradley Wade Bishop, J. McGilvray, L. Most, T. Milas, and J. T. Snead, "Geography of Virtual Questioning," *Library Quarterly* 79 (2009): 393–420.
- ¹⁴Judy Ruttenberg and Heather Tunender, "Mapping Virtual Reference Using Geographic Information Systems (GIS)," <http://web.archive.org/web/20100610103542/http://helios.lib.uci.edu/question/GIS-ALA2004/>.
- ¹⁵ "Piwik Liberating Web Analytics," <http://piwik.org/>.
- ¹⁶"Google Analytics," Google.com, <http://www.google.com/analytics/>.
- ¹⁷ Koontz and Mon, "Marketing..." 86-88.
- ¹⁸ Christie Koontz, Dean K. Jue, Lorri Mon, "PLUS+Maps: Improving Data Services for Public Libraries A White Paper for the Institute of Museum and Library Services Submitted 19 December 2013 in Partial Fulfillment of IMLS Project No. LG-61-12-0714-12
- ¹⁹ Cole, K. J. and A.C. Gatrell. 1986. "Public libraries in Salford: a geographical analysis of provision and access," *Environment and Planning* 18(2): 253 – 268; Koontz, C.M., D. K. Jue, and K.C. Lance. 2005. Neighborhood-Based In-Library Use Performance Measures for Public Libraries: A Nationwide Study of Majority Minority White / Low Income Markets Using Personal Digital Data Collectors. *Library and Information Science Research* 27(1): 28-50; Sin, S. J. 2011. Neighborhood disparities in access to information resources: Measuring and mapping U.S. Public libraries' funding and service landscape. *Library and Information Science Research* 33(1): 41 – 53.