A model for a library-management toolbox

Data warehousing as a tool for filtering and analyzing statistical information from multiple sources

Markku Laitinen

National Library of Finland, University of Helsinki, Helsinki, Finland, and

Jarmo Saarti

University of Eastern Finland Library, Kuopio, Finland

Abstract

Purpose – Quality management, and its associated evidence-based ideology, have influenced library management from at least the beginning of the 1990s. Both concepts emphasize that decisions have to be based on facts. Libraries have a long history of using statistics in data collection due to international agreements and standards. Although there is much data available in the libraries, perhaps this collected data have not been utilized as effectively as possible. This paper aims to discuss the possibilities of combining different types of statistical data in order to measure the efficiency of the work being done in libraries and what possible impact they can have in their parent organizations. It also aims to use the concept of data warehousing as a tool for joining different types of statistical data in the analysis.

Design/methodology/approach – This paper is a case study with statistical analysis of library data collected and a conceptual analysis of possibilities in the data warehousing of the library and library-related data in order to improve its analysis.

Findings – The authors suggest that in order to improve the analysis of the library statistical data towards the actual effects the library has, for example, to the learning and research done within university, one must start to develop statistical data systems that enable the combination of different types of statistical data. This needs co-operation between different vendors and government officials in defining the statistical data collected.

Research limitations/implications – The paper is based on Finnish experiences and the data warehousing model should be tested empirically and developed further due to the challenges in the conformity of the data collected.

Practical implications – The paper presents ideas and possibilities to evolve the library statistical data analysis.

Social implications – The paper presents ideas for national and international bodies in how to develop statistical data systems and their interoperability.

Originality/value – The paper promotes the idea of combining library and other statistical data collected in order to provide better tools for library management.

Keywords Evaluation, Impact assessment, Data warehousing, Data management, Statistics, University libraries, Finland

Paper type Case study

1. The core processes of the university and evaluating library's impact on these processes

The ideology of quality management started to impact on the public sector in Finland during the 1990s (see Poll and Boekhorst, 1996) which is a good summary of the development in academic libraries. Some of its concepts caused changes in the way that management works, e.g. evaluating progress and setting goals, have been adopted from the private sector. One good example was the so-called "New public management" movement, which started using private sector instruments within the public sector:"[...] instruments of such policy interventions are institutional rules and organizational routines affecting expenditure planning and financial management, civil service and labor relations, procurement, organization and methods, and audit and evaluation (Barzelay, 2001, p. 156)".

One reason for this change was the fact that academic libraries faced increasing competition due to the change in the way that scientific information is disseminated. The move from printed journals to e-journals happened within a mere ten years. At present, the same change is underway with books. Nowadays more and more titles are available in digital form as a result of back-catalogue digitization and increasingly more and more new titles are being published in this electronic format (Thompson, 2005). It has been estimated that the change to a digital book-culture shall occur in about the same time as happened with the journals, i.e. a decade.

The simultaneous changes within higher education have meant also that the service providers within the university, including libraries, have had to face many new realities (Saarti and Juntunen, 2007; Saarti *et al.*, 2010). One can argue that the traditional way of demonstrating library's results has been statistically oriented: the more you have – in terms of books, journals, premises, etc. the better you are. This rhetoric is changing towards the need for effectiveness, cooperation and allocating resources to top-level research and certain strategic focus areas. Today, it is not enough for the library simply to quantify the extent of the resources or how they are used; one also must be able to show that this investment and availability of resources produces better results in terms of research and education in the university and that these services are being delivered in an efficient manner.

This means that libraries must diversify their indices with which they are evaluated. The traditional library statistics represent a sound basis to do this analysis, but more is needed. We need more tools to describe how, in reality; the library services can influence the university's core processes and results. These are needed both at the micro-level – a good example of this is the analysis of Iivonen *et al.* (2009) which revealed how a good library collection is an effective tool for dissertation writers – and at the macro-level – an example of this is the study of Kiviniemi *et al.* (2009) who attempted to analyze the contribution of the university libraries to the Finnish national economy.

The most crucial change within the management ideology of the academic libraries during the past few decades has been a change from institution based thinking towards a service and user-oriented thinking. Of course this has long roots – we can even track them back to Ranganathan's principles that emphasize user orientation, but one can see that the digital revolution that has changed the ways that scientific documents are disseminated today as being the most important reason for this change in perspective (Kolderup Flaten, 2006).

Libraries are facing new challenges that are mostly centered on their management. For example should one have large printed collections or can one rely solely on digitized materials; should one have collections at all or is the solution here to act as a

middleman between one's own patrons and collection owners? The ultimate question is: is a library needed within an academic institution?

These questions all have several optional, perhaps equally valid answers and thus the library's management must start to make strategic decisions so as to lead the library in the correct direction. In order to make as sound decisions as possible, the management needs facts and information on which to base these decisions. The university's basic products are academic degrees and scientific publications. The processes leading to these results involve all of the various actors within a university and thus it is of the utmost importance that the library is able to show its input to those results – in both qualitative and quantitative terms.

In the following section we shall discuss the role of the management in building the tools needed for evidence-based management that represents the background ideology behind the quality management approach. In addition, we will try to make proposals for how to create a toolkit for library management involving a many-faceted approach for measuring and analyzing the library processes.

2. National statistics as a base for quantitative evaluation on the libraries

Though library statistics are just one part in the total process of measurement, they have a major role in the generation of the facts needed as the basis for planning. This is why it is important for those working in the management of libraries to acquire the ability to understand statistics. Delivering the annual statistics is a major effort for libraries, but after the statistics have been collated, the numbers are often forgotten. This "statistical illiteracy" of the librarians is generally well known and for years there has been discussion about the need to improve their skills (Ambrožič, 2003).

In Finland the statistics describing the operation of libraries have been collected for more than 100 years, and they have an established position as an important source of information. The annual statistics of the libraries of Finland are systematically collected in the joint-statistics of the libraries. There are two databases: the database for public libraries is the responsibility of the Culture and Media Division of the Ministry of Education (1999-2010), and the database for academic libraries is collated by the National Library of Finland (Research Library Statistics Database of Finland, 2002-2010). These databases are open to all interested parties through the Internet; they can freely be browsed without the need for a password and are available in the two official languages of Finland (Finnish and Swedish) plus English.

Both the academic and public libraries of Finland prepare excellent statistics which contain a substantial amount of recent information and a wide set of indicators illustrating their work. The indicators are automatically compiled by the computer systems.

The statistics are collected according to the standard ISO 2789. The statistics of the scientific libraries provide a general view of the resources, collections and services of Finnish scientific libraries. The statistics contain information about the national library, about the university libraries, about the Polytechnic libraries and about several special libraries, starting from the year 2002. The statistics on the Finnish public libraries in their present format originate from 1999. The older statistics have been published in printed form.

The information for the annual statistics is directly entered into the database in the libraries where there is an individual in charge of statistics. Part of the information saved in the database is directly transferred from the library databases.

The essential key figures describing the resources of the library, library use and library collections are automatically produced on the basis of the data in the statistical database. Some of the indicators are based on the international standard ISO 11620; some have been developed for the needs of the Finnish scientific libraries themselves.

Figures 1 and 2 show the change of the basic resources and their use in Finnish university libraries between the years 2002 and 2009. The following parameters have
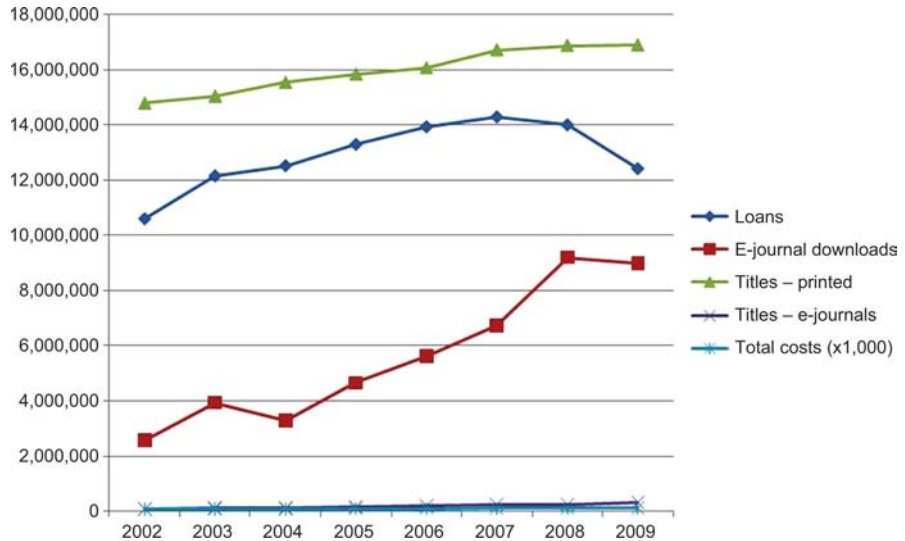

Figure 1.
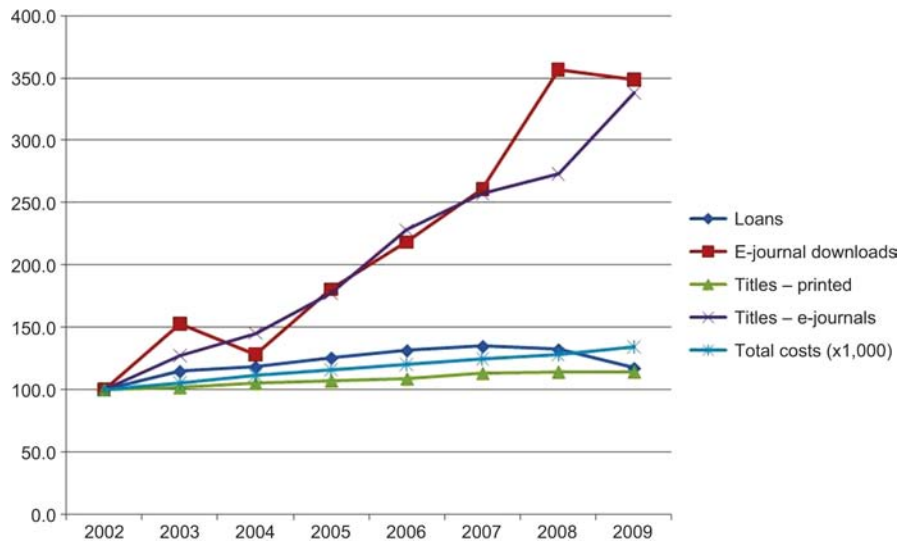The development of the basic resources and their use in Finnish university libraries 2002-2009

Figure 2.
Change in the basic resources and their use in Finnish university libraries 2002-2009

Note: 2002 = 100

been used as indicators: number of loans, number of e-journal downloads number of printed resources (books and journals), number of e-journal titles and the total costs.

One can clearly see the rapid evolution of e-collections and their use. It seems that at the same time libraries have continued to fund printed collections although their use has started to decline. Thus, one can state that trend analysis is vital for library management. If libraries were to undertake a timely trend analysis, they could target their resources more precisely.

3. The challenges of data harvesting
In addition to library statistics, the library management needs statistical data from other sources, e.g. from study administration, research administration research funding, publication databases (impact factor data), and governmental sources, e.g. from databases maintained by the ministries of education, that hold statistical data on universities and different fields of education.

During the past decades, so-called data warehousing and harvesting tools have become available for this type of merging statistical data from different sources. In the following section, we shall discuss the possibilities of utilizing these types of tools within the library sector.

3.1 Future vision of a toolkit for library evaluation management
Since the library is a basic service unit within its parent organization, it is of major importance that the library is reckoned with already at the stage when the library management is negotiating for its budget with its main financial supporter and other stakeholders. In this respect, it is crucial for the library to gather information and to measure parameters that would be deemed useful from the point of view of the decision-makers who need library information. In this way, the evaluation of the impact and effectiveness of the library would become a natural part of the annual evaluation.

Irrespective of the physical differences and the type of statistics which may be very different between the library sectors, certain principles are identical for all types of libraries. On this basis, we have created definitions of the basic indicators or products which can then be compared – "a common language" had only to be found. Of course, it may be necessary to decide whether some library sector-specific indicators related to productivity and effectiveness need to be developed.

Thus, we have compiled a proposal for a uniform set of indicators that would meet the need of the libraries, and that would be suitable for use as a tool for management as well as providing information for use by libraries themselves. This set contains certain basic indicators taking into account the needs for reporting, information and benchmarking.

The matters to be taken into account were:

- how the libraries influence the operation of their frame organizations;
- input-output;
- resources; and
- impact and effectiveness of the library in the society.

As a result of our work, we devised a set of ten indicators:

(1) The estimate "service as a whole" obtained from customer inquiries.
(2) The benefits provided to the customer by the information he/she received.

(3) Library visits/target population or inhabitants.

(4) Library total costs/target population or inhabitants.

(5) The impact of the library in its frame organization/society.

(6) Intensifying of studying and working (an effect indicator).

(7) Loans + literature downloads/target population or inhabitants.

(8) Library material costs/target population or inhabitants (physical and digital material).

(9) Information literacy, e.g. the effect of the teaching of the search and retrieval of information during studying.

(10) The scope and excellence of the collections.

In addition to these parameters, we need input-output indicators that are estimated based on the previously-mentioned topics.

As can be seen, this set of indicators contains the information that must be collected from many separate sources – library statistics, user inquiries, accounting material etc. It is a challenge to transform all this data into a uniform format, a format that permits versatile handling and the widespread use of the information to satisfy the needs of all parties that require information about the library.

3.2 The data warehouse

Nicholson (2006, 791) states that: "Traditionally, library decision-makers examine aggregates and averages to understand their service use. These measures can still all be created with this data warehouse, but it also has the advantage of empowering the managers and decision-makers to ask other questions. If only aggregates are collected and the underlying data discarded (or never made available), the ability to ask new questions is greatly reduced." Nicholson has refined the analysis of the usage of library materials with the aid of data warehouse and data-mining technologies, i.e. so-called bibliomining (Nicholson, 2003, 2006).

Although it needs extra work when establishing the necessary databases – the data warehousing approach gives active tools for a library that make it possible to combine data from different sources as well as enabling the analysis of these new datasets.

From a technical point of view, one solution could be the creation of a data warehouse as depicted in Figure 3. There the information is extracted from the operative systems of the library and is transformed into a format applicable to be loaded into the secondary system where information from different sources is collated and in which history-metadata about statistical data, is also stored. In the data warehouse, it is feasible to generate reports defined with different parameters, for example, graphic presentations or inquiries defined by the user.

Thus this model described in Figure 3 can be viewed as a process whereby the information gathered from different source systems (e.g. the operative systems of the library) is collated as uniform data through the ETL (Extract, Transform., Load) process. As a matter of fact, the ETL is a group of processes whereby the information collected from the operative systems is converted into a uniform form required by the data warehouse. In addition, one can utilize the system with so-called OLAP tools (OnLine Analytical Processing) that produce the desired reports for the users. In principle, every manager should process the issues with this in mind, even though the library would not have a computer system to produce reports.
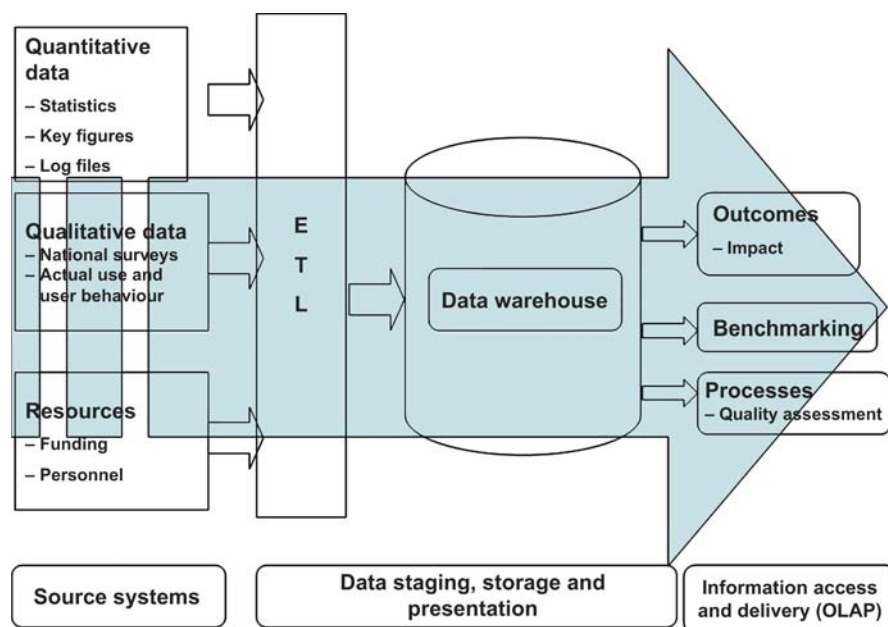
Figure 3.
Dataflow-model for the
library data warehousing

The result is an easy-to-use user interface from which the management of the library can obtain a report about a specific topic or can perform a straightforward, menu-based search. The result of the search may then be opened, e.g. in an Excel table. The tools created for this purpose are available and used in commercial enterprises and to some extent also in public sector.

4. Conclusions
The libraries have a long tradition of gathering statistical data. However it seems that this data are not used in a rational manner, i.e. for the actual analysis of the library's work and its effectiveness. Based on our experiences, one could state that the national statistical database provides a good foundation for starting this analysis because it enables the definition and analyzing individual libraries' input in a manner that converges towards a more consistent form of data collection. In addition, there is clearly a need for continuous training and communication between data collectors and data providers.

Furthermore, it also appears that this current statistical library database is not sufficient when trying to define the impact of a library on society and for the patron organization. For this purpose, we need new tools, which are able to combine different types of statistical data, e.g. the data of evaluating the university's output, both educational and scientific, data of resources used, and human and other resources. The data warehousing creates the interface for linking and analyzing the different numerical data. It provides new possibilities for accessing different sources of information in order to obtain a more accurate view of what happens when libraries conduct their role within the community.

In Finland, we have started to plan the next version of our national research library statistics database. During this process, we shall take account of the concept of data

warehousing. For example, this will mean that we have to define the interfaces between different databases in order to collate the currently scattered numerical data. However, this will enable a better analysis of the actual effects that the library services have on the research and the academic progress of the students.

References

Ambrožič, M. (2003), "A few countries measure impact and outcomes – most would like to measure at least something", *Performance Measurement and Metrics*, Vol. 4 No. 2, pp. 64-78.

Barzelay, M. (2001), "The new public management: improving research and policy dialogue", University of California Press, Berkeley, CA, (Wildavsky Forum Series, 3).

Iivonen, M., Nygrén, U., Valtari, A. and Heikkilä, T. (2009), "Library collections contribute to doctoral studies: citation analysis of dissertations in the field of economics and administration", *Library Management*, Vol. 30 No. 3, pp. 185-203.

Kiviniemi, V., Laitinen, M. and Saarti, J. (2009), "Are libraries worth investing in?: Finnish university libraries and their effect on the national economy", *Liber Quarterly*, Vol. 19 No. 1, available at: http://liber.library.uu.nl/ (accessed 24 March 2011).

Kolderup Flaten, T. (2006), *Management, Marketing and Promotion of Library Services. Based on Statistics, Analyses and Evaluation. Edit*, Saur, Munchen, IFLA publication 120/121.

Ministry of Education and Culture, Division for Cultural Policy (1999-2010), "Finnish public libraries statistics", available at: http://tilastot.kirjastot.fi/en-GB/ (accessed 24 March 2011).

Nicholson, S. (2003), "The bibliomining process: data warehousing and data mining for library decision-making", *Information Technology and Libraries*, Vol. 22 No. 4, pp. 146-51.

Nicholson, S. (2006), "The basis for bibliomining: frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services", *Information Processing and Management*, Vol. 42 No. 3, pp. 785-804.

Poll, R. and Boekhorst, P. (1996), *Measuring Quality: International Guidelines for Performance Measurement in Academic Libraries*, Saur, München, IFLA publications, 76.

Research Library Statistics Database of Finland (2002-2010), available at: at https://yhteistilasto.lib.helsinki.fi/language.do?action=change&choose_language=3 (accessed 24 March 2011).

Saarti, J. and Juntunen, A. (2007), "From the rhetoric of quality management to managing self-organizing processes: a case study on an expert organization", in Nyce, J.M. (Ed.), *Advances in Library Administration and Organization, Vol. 25*, Emerald Group Publishing Limited, Bingley, pp. 97-112.

Saarti, J., Juntunen, A. and Taskinen, A. (2010), "Multi-faceted measuring of the quality as a tool for quality improvement in the Kuopio University Library, Finland", *Performance Measurement and Metrics*, Vol. 11 No. 1, pp. 39-46.

Thompson, J.B. (2005), *Books in the Digital Age: The Transformation of Academic and Higher Education Publishing in Britain and the United States*, Polity Press, Cambridge.

Corresponding author
Jarmo Saarti can be contacted at: jarmo.saarti@uef.fi

www.manaraa.com