
Developing an ISO 2789 compliant database for South African academic library statistics

Deborah Becker

Deborah Becker
Cape Peninsula University of Technology
beckerd@cput.ac.za

Abstract

The use of statistical data in libraries is important in management and advocacy. Efficient and effective tools are essential in order to make better business and service decisions, to make the library more visible and increase user satisfaction. In the management of libraries operational statistical information is used to make important development and policy decisions. In the budgeting process the request for financial resources is often based on the need for statistics that show trends and performance in service delivery.

Over a period of three years a statistics database was developed and tested by staff at Cape Peninsula University of Technology (CPUT) Libraries, Western Cape, South Africa and then extended to include interested libraries belonging to the national body, the Committee of Higher Education Libraries of South Africa (CHELSA). These two phases of the initiative represent two levels of database usage.

This article outlines the basic principles of the two separate levels of the database as well as some more advanced features such as the alerts systems and available reports. It also provides an overview of the structure of the database and how the various data elements are linked within the database.

Keywords: Academic libraries, Statistics, Database, South Africa

Introduction

The use of statistics data in academic libraries is important in management and advocacy. There are many reasons why higher education libraries gather, store,

analyse and provide access to data about their users, operations, personnel, services and collections. These may be to determine trends within a library over a period of years or to benchmark with other similar libraries. Statistics can provide data upon which to base business and service decisions, to make the library more visible and increase user satisfaction. In the management of libraries, operational statistical information is used to make important development and policy decisions. In the budgeting process the request for financial resources is often based on the need for statistics that show trends and performance in service delivery.

In some countries, libraries have been collecting and recording statistical data at a national level for many years. These include the United States of America, Australia, Britain, Germany and Finland. Viewing of the statistics is sometimes restricted to members who subscribe to the national database (for example in the case of Australia and Finland (National Library of Finland 2002)). In some cases statistics have been recorded nationally for more than 10 years.

Library statistical data when compared with that from other institutions providing similar services assists academic libraries in identifying areas in which the library is well resourced or areas that require further improvement. They can also be used to show the return on investment on funds received from the mother institution in terms of library use, the strengths and weaknesses of collections, age of equipment, and adequacy of seating space and provision of library and information services. At present South Africa lacks a central database that reflects activities in academic libraries and the development of such a database and planning tool will go a long way towards improving the way libraries request funding and other support from mother institutions and government.

Cape Peninsula University of Technology (CPUT) Libraries have realised that the development of a national higher education libraries' statistics database will be a milestone in the development of higher education libraries. In order to achieve this, CPUT Libraries proposed the development of an automated statistics database (Becker 2010) for CPUT and South African higher education libraries.

A pilot project (phase 1) would be run at CPUT libraries and then rolled out to other interested academic libraries in South Africa (phase 2). As it is envisioned that the database would form the foundation for a national database of library statistics in academic libraries in South Africa, it should also conform to existing protocols, for example ISO 2789, which provides standards for the collecting and reporting of statistics for the library and information services.

Background

In 2010 a brief was received to develop a basic database for the collection of statistical data generated at CPUT Libraries. An investigation was done into the data to be collected and based on this information an outline was developed for the database (Becker 2010). Although it was recognised that MySQL (a structured query language used for managing relational database systems) would be the best platform for the database, a prototype was developed using Microsoft Access (now deactivated). The reason for this decision was that the database could be set up fairly quickly and easily as the software was available and staff had a working knowledge of the software. The initial database was developed by library staff as an example of how the database could be structured and how data could be captured.

The prototype was demonstrated at a meeting of the Committee of Higher Education Libraries in South Africa (CHELSA) in 2010 and instruction was received to develop a basic database for the recording of statistics for CHELSA libraries that were interested in participating. Members of CHELSA had already developed a document, Measures for quality in SA HEI libraries (CHELSA 2006b), based on measures collected by the Association of Research Libraries (ARL 2013), the Council of Australian Libraries (CAUL) and the Society of College, National and University Libraries (SCONUL), outlining the statistics which should be recorded and which would be used for benchmarking academic libraries in South Africa. The document was prepared by the Quality Assurance Subcommittee of CHELSA.

However, during 2010 the CPUT Libraries underwent an audit. A guiding document for the collection of data for the audit was the CHELSA Guide to the self review of University Libraries (CHELSA 2006a). A number of requirements in this document were not in the Measures for quality in SA HEI libraries document developed by CHELSA. For its own purposes, CPUT Libraries decided to incorporate elements of the CHELSA (2006a) guide into the database so that the data would be more easily available for future audits. While developing the database, it was also noted that an International Organisation for Standardisation guideline called ISO 2789 and published in 2006 existed, which outlined the standards required for the collection of statistical data. It was therefore decided that the database should be evaluated against the standard to ensure compliance with international statistical data collection standards.

The 2010 brief and development of the database

In 2011 the library managed to secure funding to hire a MySQL programmer for four months to assist with the database development. The project appeared fairly simple; however, due to the complexity of some of the data to be captured and the enhanced approach taken by CPUT Libraries, the project soon became more complex and time-consuming than first imagined.

Initially, the focus was to develop a statistical database for CPUT Libraries with a planned future development to extend the database to include CHELSA Libraries. Each institution would have full access to CHELSA-level data (both for inputting and viewing of reports) as well as the option of using the institution-level database for the collection of their own internal statistics. The CHELSA-level data would be based on the items listed in the Measures for quality in SA HEI libraries (CHELSA 2006b).

It was recognised that the database should be structured to allow library staff to input data for their own institution and to view data relevant to their institution as well as data necessary for benchmarking across all participating libraries. The integrity of the data needed to be maintained as well as confidentiality of institutional specific data. Three key factors were identified as essential to such a database. These were Access control, Alerts and Reports.

Although most statistical data is captured annually, it was decided that staff should be able to input their statistics into the database monthly. The database would in effect replace the paper-based system of data collection currently employed at CPUT Libraries.

The purpose of the database

It is envisioned that the database will serve a number of purposes, namely:

- To create a friendly digital environment for the registering of statistics by staff (replacing the current manual system)
- To make library statistics easily accessible to staff for use in reports
- To have controlled access based on need and authority
- To provide a central point for accessing library statistics
- To ensure staff have access to accurate and accepted library statistics
- To store data in table formats
- To archive the statistics to ensure that they would be available for future trend analysis

- To link various tables together to draw more complex reports

With the extension of the database to include CHELSA libraries, the purpose would be extended to include:

- Benchmarking with one or more selected institutions
- Trend analysis

In the context of the higher education sector libraries the advantages of the database would include:

- An automated system for collecting statistics at regular intervals (at national and institutional level)
- Simplifying the preparation for Quality Assurance reviews
- The provision of accurate and current data to assist library managers with decision making, strategic planning, library research and budgeting

Design criteria

The database was designed using MySQL and php (a scripting language used for web development as well as general programming) and consists of two main levels of access, namely, that of Global Administrator and Institutional access. The Global Administrator has the ability to add institutions and institutional administrators as well as update the forms for the collecting of CHELSA data. Institutional access includes administration functions assigned to the Institutional Administrator as well as data entry and retrieval functions. The database is currently hosted by CPUT and is kept behind a firewall with access only to the web-based functionality available to Institutions.

Access to the database

In order to access the database, a user has to be registered. Registration includes personal information such as e-mail address, name and a password as well the selection of an institution and branch. The entered data is registered in the Persons table and can only be accessed by the registered user.

The selection of an institution and branch determines the access rights of the person logging in. The institution selected would refer to the library of the institution as defined in ISO 2789:2006 (E) 3.1.6 and the branch as defined in ISO 2789:2006 (E) 3.1.2.

Once logged in, the user would only be able to access the data relevant to the institution and branch selected during the registration process. Access would be further restricted by roles and access permissions allocated to the user. The system administrator could also determine whether or not the user has further access to the registration screen in order to update details or not.

Database structure

In order to accommodate different institutions and their branches, three tables were set up in the database, namely: Region, Institution and Branch. Due to CPUT Libraries' requirements two further tables allowed the listing of Faculties and Departments. The Region and Institution tables can only be accessed by the Global Administrator, currently the Quality Assurance Librarian at CPUT Libraries. The Regions were based on the nine provinces of the Republic of South Africa. Institutional access gives the Institutional Administrator full access to all tables relating to the institution, including the setting up of branches and, if required, faculties and departments.

The basis of the database is thus captured in a few simple tables, namely: Persons table (holding the personal profile of staff), Regions, Institutions, Faculties, Branches, Units of Measurement, Roles and Data entry cycles.

The Units of Measurement table allows the administrator to define possible units of measurement which can be selected for each counter as well as the number of decimal places that will be displayed. Examples of Units of measurement include FTE, headcount, and so on. Although this table covers global access, institutions may add units if so desired.

The table Roles will be unique to each institution and lists predetermined roles that staff can be assigned in the database. Included in the Roles menu is an indication if the role is supervisory or not and to which other roles that role reports.

Data entry cycles are the years for which data is entered. Each institution can select its own data cycles. The data entered during the cycle will apply to that specific year with annual data being input at the end of each year as suggested in ISO 2789:2006 (E) 5.2.

CPUT Libraries requires certain data to be input at a Departmental level. This table is different to the other tables as it is linked to the Faculties rather than the Institution.

In order to allow for maximum flexibility, a three tier structure was created for determining the organisation and input of data (see Figure 1: *Example of a hierarchical structure*). These were Category, Sub-Category and Counters. The Categories, Sub-categories and Counters can be determined by each institution and can vary from institution to institution. However, in the case of CHELSA they are set up at global administrator level and cannot be changed by individual institutions.

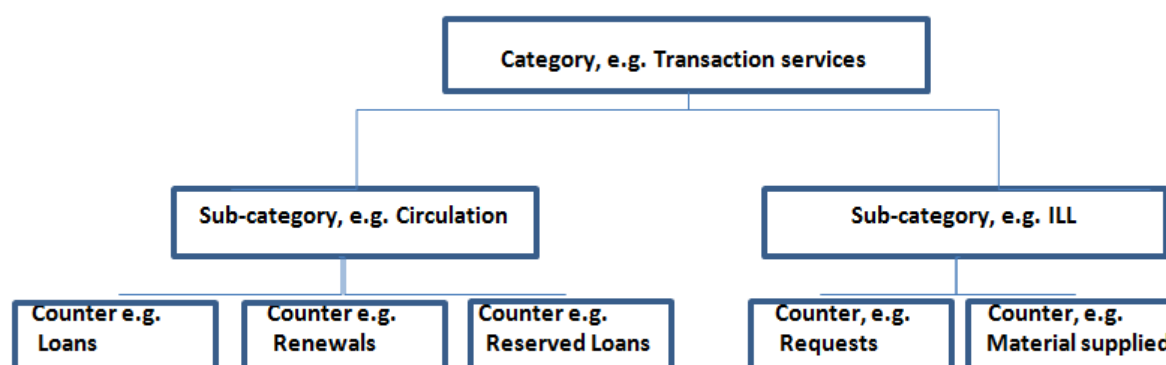


Figure 1: Example of the proposed hierarchical structure

Categories

Categories incorporate a number of sub-categories, each with their individual counters. They cover a fairly broad area, for example Transaction services. When creating the category certain parameters such as display, counter usage and institutional affiliation must be selected.

Generally during data input and reporting only counters are displayed. However, at times it is necessary to display the sub-category in order to make sense of the counters. The decision whether or not to display sub-categories is registered on the categories menu. Each category is linked to an institution.

There are eleven CHELSA categories, based on the arrangement of listed items in the CHELSA document. CPUT Libraries have a further 12 categories based on selected library activities. When accessing the database, staff will see only the categories that apply to them directly.

Sub-categories

Each category is divided into a number of sub-categories. This is done to break the categories into logical and more easily manageable sections. For example, the category Transaction services is divided into 10 sub-categories.

Below is an example of the sub-categories and counters for the category CPUT ILL.

Sub category	Counter
ILL - Requests for books from CPUT	ILL - req. for mat
"	ILL - Int. mat req
"	ILL - Satisfied req
ILL - Copies requested from CPUT	ILL - req. for mat
"	ILL - Int. mat req
"	ILL - Satisfied req

Figure 2: Example sub-categories for the category: ILL

Counters

Counters are the fields which will be displayed when inputting data. The counters menu includes parameters such as unit of measurement and data entry frequency. The Unit of measurement is selected from a drop down menu and includes options such as FTE, headcount, and so on, while Data entry frequency indicates how often the data is expected to be input. Currently, only two options are available: monthly and annually. An option for quarterly, or selected month, was suggested and investigated, but it was determined that this could not be completed within the limited time frame and would be listed as a possible future development.

Each counter also includes an explanation which is available to the user when selecting the icon in the Definition/units/roles field on the data input screen (see figure 3: *Example of counters and related fields*). Where available, the definition is taken from the CHELSA document and where not CPUT Libraries' definition is included. Institutions can develop their own definitions, based on their required parameters. The explanation includes the selected Unit of measurement for the counter as well as the Roles which can access the counter.

Inter-library and Inter-Branch loans	ILL	0	
"	IBL	0	

Units: Number Integer

Definition: Report the number of requests for material (both returnables and non-returnables) provided to other libraries. Include originals, photocopies, and materials sent by fax or other forms of electronic transmission. Include patron-initiated transactions. Exclude all transactions involving movement of stock within your own library system.

Allocated roles: ILL

Definition/units/roles



Figure 3: Example of counters and related fields

Menu items

Initially all data was input using the above tiers, but as the reporting module was developed and data was needed in specific output format or calculations were included as part of the reports, it was noted that the structure did not cater ideally for data input for departments, e-resources and some other areas. The result of these variations was that menu options were included to cater for these items.

In the case of e-resources, it was important to include information such as start date, end date and status. Also when drawing the report, calculations for cost per search and cost per full text download were required.

Due to the long list of e-resources, it was decided to divide them into three groupings namely, Databases, E-books and Tools. Tools are seen as resources that will be used only by library staff, e.g. Bookdata Premier or as supporting users while not actually providing data, e.g. RefWorks. It can also be used to register costs for subscriptions which relate to the E-resources budget, e.g. SANLIC (South African National Library and Information Consortium) fees. Together the three divisions should represent a complete list of all e-resource expenditure.

With student training, detailed data was needed especially in view of the next library audit which would concentrate largely on information literacy and user training. It was determined that the type of training, the department for which the training was done as well as the year of study of the student had to be captured. A further complication is the selection of the timeslot, that is, whether the training was done during the day or during the extended library hours.

Currently, the database only allows for the entering of numerical data; however, the addition of alpha-numerical data is seen as a future development.

Links

In order to display data properly as well as ensure that staff can only access allocated areas of the database, the tables and menu items had to be linked at various levels. Figure 4: *User links* gives a diagrammatic explanation of tables to which the user is related.

Users

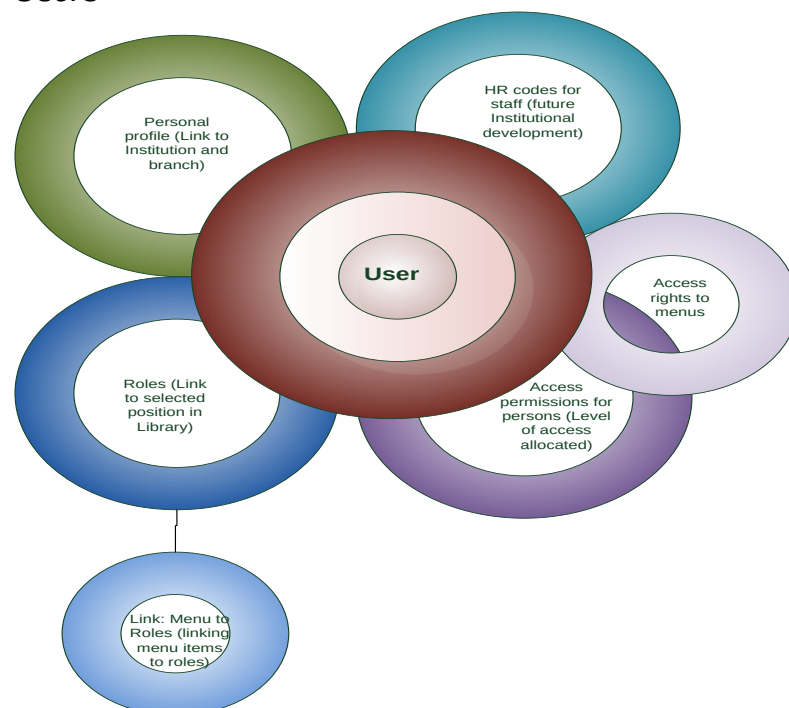


Figure 4: User links

Staff members are linked to their selected institution and branch by selecting the appropriate options when registering on the database. Upon registration all users are automatically assigned the access right: No access privileges.

The systems administrator allocates access permissions and roles to the user based on the staff member's position in the institution and authority.

These allocations ensure that the staff member will only see data entry screens on which they need to enter data, menu items appropriate to their needs and the reports to which they are allowed access. Further, the above will also inform the alerts system, ensuring that users receive appropriate and timely alerts.

Data

Data is linked to each tier in the structure, starting at the Institution level and each level thereafter: Branches, Faculties and Departments. Counters contain the actual data input and therefore within each tier data is linked to sub-categories and categories as well as Units of measurement and Data entry cycles. As some data entry fields are represented by menu items, data is linked to these as well. (See Figure 5: *Data link*.)

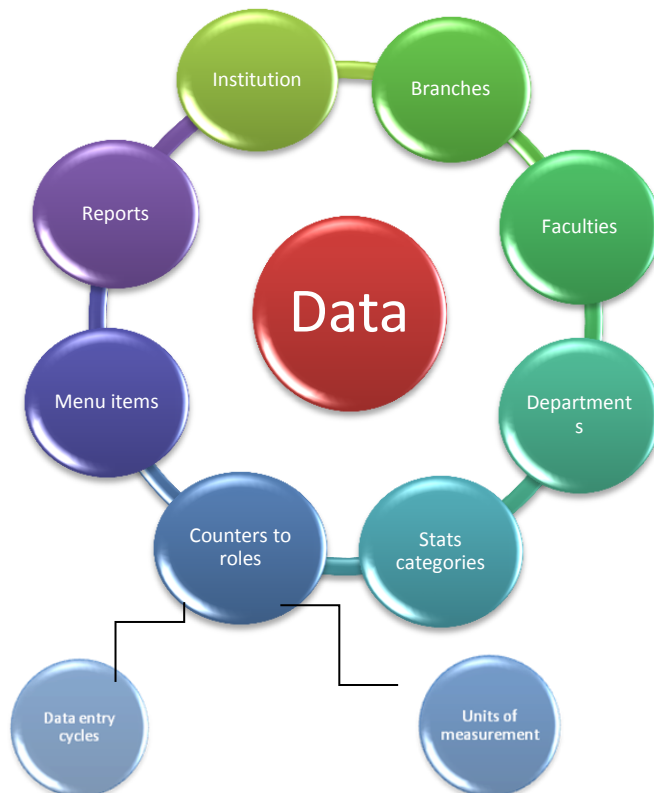


Figure 5: Data links

Automated alerts

Automated alerts were envisioned as an integrated part of the statistical database. Two types of alerts were developed and implemented. The first was an e-mail sent to the system administrator when a staff member registered on the database. The administrator was then able to assign the relevant access permissions and roles to the staff member.

The second alert was with regard to data input. An e-mail alert is generated and sent to all staff with data input access according to a selected schedule. The purpose of the data entry alert is to ensure that staff complete their data entry within a given time frame and that supervisors are kept informed of staff who do not enter data. Alert reports and logs are available and can be accessed by Institutional Administrators.

Data entry

Data entry is a relatively simple input screen. However, before data can be entered a series of selections must be made.

- The year and month for data input must be selected. A year is only selected if more than one data entry cycle is open. This normally occurs at the beginning of a year when data for the previous year is still being input.
- The user is then asked to select a statistical category. The list of categories available is dependent on the role allocated to the user as well as their branch.
- A “Faculty” option will appear if the Faculty is linked to the branch. If not linked, the user will not see this option.

Each data entry screen contains a form for the selected category. The sub-categories and counters are listed on the form as well as a field for adding a value (see Figure 6: *Input screen*). The last column on the input form contains an option to view an explanation of the data to be input into the field.

Sub category	Counter	Value	Definition/units/roles
ILL - Requests for books from CPUT	ILL - req. for mat	0	

Figure 6: Input screen

Due to some counters being listed as monthly and others as annual, the list of available categories will change at the end of the year. All CHELSA data would only be captured annually.

Document management

During the development of the database it was noted that it is sometimes valuable to be able to load a document into the database for future reference, especially if the documents refers to standards or guidelines used in setting up the database. To this purpose a document module was created within the database allowing for the addition of documents to the database. Staff can add document folders in order to group together similar documents thereby making access easier.

Reports

Reports are an essential element of the database as they determine what data can be viewed by the user. There are two different types of reports on the system, namely, administrative reports, and end-user reports.

Administrative reports

These reports are viewed by the Institutional Administrator only and assist in providing an overview of the system and the data input.

1. Data entry cycle – this report allows the Institutional Administrator to view all data input.
2. Categories hierarchy – this report reflects the category, sub-category and counter structure of the database. The report gives an overview of the structure and allows the Institutional Administrator to check easily that the correct links have been made.
3. Role allocations – there are four reports in this grouping, each giving the Institutional Administrator a different view of the allocation of roles.
4. Domain administrators
This report lists all administrators and is arranged according to level of privilege.
5. System notes – this is a log of all requests, bugs and completed tasks. The list enables the Global Administrator to keep track of outstanding tasks and to monitor the completion of such tasks. An input screen for system notes allows Institutional Administrators to input requests or report bugs online.

Report designer

The Report designer is not essentially an administrative report; however, it does require more advanced access privileges. The Report designer allows the creating of reports on two levels namely, CHELSA and standard institutional reports.

End-user reports

Access to these reports is based on user privileges. Access is view-only; however, the reports can be downloaded into EXCEL or WORD. Although data cannot be changed, certain data can be hidden if opened in EXCEL.

Discussion

Once the database was developed, CPUT Libraries started training librarians and other relevant staff to input data and by December 2011 all staff were inputting their monthly and annual statistics directly into the database. The

general feedback from staff is that the database is easy to use and much more user friendly than the previous paper-based method of recording statistics. After the successful implementation of the database at CPUT Libraries, roll-out of the database to CHELSA libraries commenced in 2013.

The database is an ongoing project and already a number of possible developments have been identified. These include:

- Inputting of alpha-numerical data
- Inputting using CSV (Comma-separated values) files
- Uploading data from sources such as Aleph, the Library Management System used at CPUT
- Creating on-line forms
- Expanding the data input frequency to include daily and quarterly statistics
- Automatically e-mailing reports to relevant staff on a monthly/annual basis

In order to accommodate the roll-out of the database to other institutions, comprehensive guides have been developed. These include a guide for developers, administrators and a general help option for end users.

Conclusion and recommendations

This article has described the development of an automated statistics database for the South African higher education libraries by CPUT Libraries. However, developments are on-going. As requests for data are received gaps are discovered and the database adapted, for example in areas such as the recording of detailed training statistics.

The ISO 2789 standard was updated in 2013. Such updates and the publication of any new standards would need to be considered and the database checked for compliance.

As all data input into the database can be retrieved as reports, librarians at CPUT Libraries have started using the data in their reports to library management. The database is also being used to provide statistics for the library's annual report and for advocacy to University management as well as for marketing library services to students.

As more libraries use the database, it should be easier to use the available data for benchmarking and advocacy.

References

Association of Research Libraries (ARL). 2013. ARL Statistics : annual library statistics. <http://arlstatistics.org/home>. Accessed 26 June 2013.

Becker, D. 2010. Proposal for the development of a statistical database for the Cape Peninsula University of Technology Library. Unpublished document.

Committee of Higher Education Libraries of South Africa (CHELSA). 2006a. Guide to the Self Review of University Libraries. Draft 4.

Committee of Higher Education Libraries of South Africa (CHELSA). 2006b. Measures for quality in SA HEI libraries. Draft 3. Prepared by the Quality Assurance Subcommittee of CHELSA.

Council of Australian Libraries (CAUL). 2012. Statistics. <http://www.caul.edu.au/caul-programs/caul-statistics>. Accessed 8 March 2012.

National Library of Finland. 2002. Research library statistics database. <https://yhteistilasto.lib.helsinki.fi/?lang=en>. Accessed 17 May 2013.

International Organisation for Standardisation (ISO). 2006. International standard ISO 2789 : Information and documentation – International library statistics. 4th ed. Geneva: ISO. Reference number: ISO2789:2006(E).

Society of College, National and University Libraries (SCONUL). 2013. Statistical Questionnaire, 2010-2011. <http://vamp.diglib.shrivenham.cranfield.ac.uk/statistics/sconul-statistical-questionnaire>. Accessed 17 May 2013.