ISO 2789: WHAT'S NEW IN THE REVISION?

John Sumsion

A Working Group of TC46/SC8 has been working on revisions to the International Standard for Library Statistics (ISO 2789) since 1997. Important changes to traditional statistics are proposed, as well as the integration of electronic materials and services wherever possible. These are outlined. A new categorisation of statistics of databases is proposed. When it comes to statistics measuring usage, in August 2001 there was still some disagreement on which measures are most appropriate and feasible: the points at issue will be discussed. Resolution of these matters is complicated by the different situations in various countries and library sectors. It is important to measure the effect of electronic developments on traditional services.

John Sumsion, Honorary Senior Research Fellow, Department of Information Science, Loughborough University, UK
E-mail: J.W.Sumsion@lboro.ac.uk

Introduction

The previous revision of ISO 2789, the international standard entitled Information and documentation – International library statistics, was completed in 1991.

In the mid-1990s, Working Group 4 (led initially by Charles McClure but later by Geoffrey Ford) spent much time and effort developing a new standard, which was published in 1998 as ISO 11620 *Information and documentation – Library performance indicators*. Even before its publication, it was realised that it was inconsistent with ISO 2789 in important respects, and the procedure to revise ISO 2789 had been initiated. Formally this was entrusted to Working Group 2 of the parent Committee SC8, one of the committees of TC46 in the ISO organisation.

Early in 2000 the USA resigned from SC8, principally because of disagreement with the direction the revision was taking and its timing. A meeting in Berlin (May 2000) was able to note a favourable vote on the committee draft but had many detailed comments to consider. A draft standard was produced in July 2000, for group members to consider with their national standards committees. The USA decided to rejoin the work in the summer of 2001, and also decided to initiate a revision of their own NISO standard.

Following a favourable vote on the draft international standard a final meeting of WG2 was held in Paris in August 2001, with the important participation of a strong USA delegation. Significant amendments were agreed, particularly on statistics to measure the use of electronic and networked materials. It is expected that the revised final standard will be promulgated in 2002.

Objectives and scope

The original objectives of the standard, which remained unchanged, were to facilitate international comparisons and to

provide a model for national standards. New objectives include:

- to update the standard in line with changes in library and information practice and technology

- to cover electronic and networking developments

- to satisfy the requirements of ISO 11620 – Performance Measurement

- to encourage good practice

Unlike most manuals and treatises this standard has to apply to all types of libraries. It has also to be useful in countries with fully developed library systems as well as where libraries are at an early stage of development.

Two major innovations are:

Sampling In many places statistics derived from sampling procedures are now included. Instruction on the detailed statistical sampling techniques to be used is outside the scope of the standard, though the principles involved are laid down.

Grossing Up A common problem with national statistics is how to produce representative results when some data is missing. The need for such calculations, and principles underlying them, though not the detailed techniques, are now included in Annex C of the standard.

Level of detail

The standard does not set out to provide technical detail or sets of examples; nor can it supplant the need for separate instruction manuals. However, definitions are spelt out as fully as is considered necessary and notes are provided to avoid the most common or likely cases of misinterpretation. In this respect it goes into more detail than, for instance, the American ANSI/NISO standard.

Structure of the document

The structure of the main standard is essentially unchanged. For any statistic both the 'Definition' and the 'Count' have to be read, and they are pages apart. The statistical categories are:

- Libraries

- Collections

- Use and users

- Access and facilities (new)

- Finance

- Staff

Annexes

A new feature, copied over from ISO 11620, is the introduction of three annexes. These can be revised and updated more easily and in less time than the main standard, though the procedure is still not quick.

Annex A covers the approach to Electronic Services and Networking – particularly statistics to measure use. This is the most critical area for innovation with much ongoing development.

Annex B gives optional extra categories applicable to some types of library such as:

- Broad categories (fiction, non-fiction, reference, children's) for public library stock

- Subject groups (14) for academic and special library stock

- Acquisitions counted by language and provenance

- Acquisitions by donation, exchange, legal deposit

- 'Short loans' defined (academic libraries)

- Less common material types

Annex C refers to Grossing Up (explained above).

New definitions and counts

Libraries and general

The new definition of **library** is sufficiently wide to cover a library that is entirely digital by including the term electronic collection – electronic resources to which access has been obtained by the library. It is 'to build and maintain a collection and to facilitate the use of such information resources and facilities as are required to meet the informational, research, educational, cultural or recreational needs of its users'.

The term **service point** has been dropped so that there are now only central/main libraries, branch libraries and mobile libraries within each 'administrative unit'. However, there may

be 'external service points' away from library premises.

Types of library are unchanged for national, public, higher education and schools libraries. There is a major change, trialled by LibEcon [1], to classify **special libraries** for reporting under new headings:

– government

– health service/medical

– professional and learned institutions and associations

– industrial and commercial

– media

– regional

– other (including voluntary organisations)

This change promises to bring greater appreciation of this growing sector, whose size and significance is constantly understated through lack of statistical data.

Facilities to be counted include the number of seats, photocopiers, and workstations. Opening days and hours are counted for the central/main library separately from the branch library average. Space is to be counted and split between space for user services, library operations, materials storage and for meetings, exhibitions and events. Counting catalogue records is an innovation.

Use and users

The new standard emphasises use and users. Conventional users are classified as: registered users, active borrowers and active users. Annex B provides optional categories for types of user – varying by library sector. Visits may be counted electronically, by turnstile or by sampling. Remote visits are included – by mode of communication

Statistics of **book issues** are now more closely defined, with new distinctions between ordinary loans (material taken away), on-site loans (material issued for use on the premises), and in-house use (open shelf material used on the premises). The alternative specified is the snapshot count of material – number 'on loan' and number 'in use within the library'. Loans now include CD-ROMs and eBooks.

The count of **information requests** includes a separate total for Internet queries. Annex B provides an optional breakdown by enquiry type.

Staff

Apart from spelling out the FTE (full time equivalent) basis there are two new categories of staff: qualified specialists (accountants, IT, staff management) and non-remunerated volunteers. There are new statistics both on training of users undertaken by staff and also hours and cost of training given to staff.

Electronic services and networking

In the highly dynamic and uncertain situation the approach has been to:

• count what we can count at the present time

• integrate with other statistics – just as electronic services are integrated with conventional in the hybrid library

• concentrate on the basics, recognising that counting use is more difficult and complex

• accept that some imprecision is, in practice, unavoidable but that this does not negate the value of statistics to monitor developments and progress

• build on research work undertaken and ongoing: ICOLC; Bertot, McClure *et al.*; ARL; Equinox; etc.

• at a minimum, to set strategy and targets for ongoing development by both library professionals and suppliers.
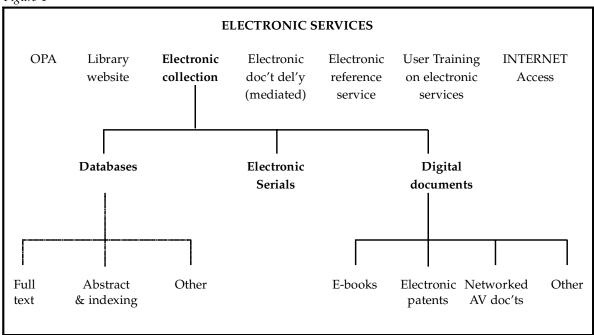
Rather to the surprise of the working group it was found possible to integrate many electronic features into the main structure of the standard: Definitions; Collection(s); Facilities; Expenditure. So there are many electronic features to be found in the standard alongside conventional items.

Much can be inferred regarding electronic developments from basic statistics of numbers of workstations (by type) and other facilities, along with data on numbers connected by network to the library.

It is principally the counting of use and users, the most difficult electronic area, that has been separated out from the main standard into Annex A. This annex is designed for flexibility and to allow relatively rapid update – following the annex structure of ISO 11620.

The distinction between mediated and unmediated access is found to be particularly critical.

Figure 1

ELECTRONIC SERVICES

OPA   Library   **Electronic**   Electronic   Electronic   User Training   INTERNET
         website   **collection**   doc't del'y   reference   on electronic   Access
                                     (mediated)   service   services

Databases   **Electronic**   **Digital**
                              Serials   **documents**

Full   Abstract   Other   E-books   Electronic   Networked   Other
text   & indexing   patents   AV doc'ts

Measuring the activity of library staff is important per se and because it is technically less difficult. However, the volume of unmediated activity in the electronic area is most important, just as it is for access to open shelves in the conventional. The parallel between unmediated access to databases and taking material from the open shelves is interesting. There are prospects of measuring this electronic access and use more effectively than has ever been done for in-house use of conventional materials.

Collection statistics

For all items statistics are required of stock, acquisitions and withdrawals.

Conventional material types are: Books and serials (printed); Manuscripts; Microforms; Cartographic documents; Printed music; Audio-visual documents; Graphic documents; Patents; Other items.

The basic principle is that the library's collection is extended to encompass, beyond the physical, electronic items to which access has been secured by the library for its users. Such access includes that via co-operative arrangements. Material for which free access is facilitated by the library does not form part of its collection – although this activity is recognised separately in use statistics. This relationship, and others, are outlined in Figure 1.

(Note that the term 'document' refers not only to a physical print item but also to electronic/digital collections of information. Provision of interlibrary loans and document delivery are not part of the library collection.)

Databases are defined as 'collections of electronically stored data or unit records (facts, bibliographic data, and texts) with a common user interface and software for the retrieval and manipulation of the data. . . . usually collected with a particular intent and related to a defined topic.' This revised ISO 2789 definition differentiates databases from electronic serials, which are treated as a separate category in accord with long-established professional practice.

In considering the nature of databases the working party realised that conventional **abstract and indexing** search tools had never been separated statistically either from full text journals or from other reference publications such as directories and annuals. When looked at from the outside this seems a surprising weakness in conventional statistics. A subdivision of databases into 'Full Text', 'Abstract & Indexing' and 'Other' seemed very desirable to provide a more complete and meaningful picture. In the event some countries felt there were insuperable technical problems in such a classification – with new products straddling the subdivisions – so this threefold classification is in annex B as an optional statistic. While professional interest is concentrated on electronic journals, the other types of information database need practical recognition.

A related complication that points to the need for lower-level categories arises from current practice of the Association of Research Libraries, North America (ARL) to include 'Indexes & Reference Tools' as a sub-category of electronic serials [2], where European practice (and the revised standard) counts 'Abstract and Indexing' as a bibliographic 'database'. This is one of the few examples to come to light so far of a major inconsistency in current statistics in this area.

Serials are moving dynamically from 'print only' format to combined 'print + electronic' and then to 'electronic only'. The speed and extent of this movement to **electronic serials** is difficult to establish without standardised statistical reporting. ISO 2789 specifies:

–   separate statistics for newspapers and periodicals

–   separate statistics for titles and subscriptions

–   reporting 'print only' serials

–   reporting 'electronic serials', i.e. 'print + electronic' and 'electronic only' formats combined

–   reporting totals, i.e. all formats

Financial statistics

Expenditure statistics for electronic services are further developed at the present time than others. Given the explanations above the revised list of items that follows should present no surprises.

Acquisitions (= costs of all items added to a library's collection for the benefit of users)

a) printed books

b) printed serials

c) non-electronic audio-visual

d) other non-electronic documents

e) databases

f) electronic serials

g) digital documents (excluding eBooks)

h) eBooks

TOTALS: 'a', 'b', 'c' and 'd' when amalgamated give a sub total for 'non-electronic material' – with the rest for 'electronic material'

Other operating expenditure

–   employees (costs of training staff noted separately)

–   external document supply and interlibrary lending

–   collection maintenance (binding, conservation, etc.)

–   premises

–   automation

–   miscellaneous

Capital expenditure This includes the cost of new computer systems (hardware and software) but not their maintenance. Costs of digitisation, when carried out by outside contractors, are included under the 'collection maintenance' heading of operating expenditure.

Electronic use statistics

Annex A of the standard considers unmediated access to electronic data – most importantly to the library's electronic collection. The feasibility of these measures has not yet been fully explored for all types of library, and further development is anticipated. Technical problems, such as access by search engines and cached searches, have yet to be fully resolved.

With these reservations, statistics of use are specified and described under these heads.

Forms of use

Four core datasets are priority items:

– number of sessions

– number of documents downloaded

– number of records downloaded

– number of virtual visits (to the library web site from outside library premises)

Four additional relevant items should be collected where appropriate:

– session time

– number of rejected sessions (turn aways)

– number of searches (queries)

– number of internet sessions

These are required separately for each service as well as summed for all services.

The key definition of '**session**' is 'successful request of a database or the OPAC', but with notes that exclude accesses to gateway pages and the library website (counted as 'virtual visits'). 'A session is one cycle of user activities that typically starts when a user connects to a database or the OPAC.' It ends by leaving the database through log-out or exit.

Location of use

It is desirable to count separately:

– provision inside the library

– provision outside library premises but inside the parent institution or authority (population to be served)

– outside the parent institution or authority (population to be served)

Separate services

While the main focus will usually be on use of categories of the library's own electronic collection, data is also required on the use of the OPAC, on visits to the library web site, and on internet sessions.

Surveys of use and users

Finally, the opportunity to gather information by conventional, and electronic, survey techniques should not be overlooked. Where computer data are not feasible or are incompatible, this need not be the end of the story. Structured questionnaires to users, and interviews, have already yielded many valuable insights. The final section of Annex A points to typical questions that can be posed in both general and specific surveys of user activity in the electronic domain.

Conclusion – Prospects

• the standard provides a hybrid, integrated solution, not a divorce of the electronic from the conventional, and it is vital to retain comparisons with the conventional

• new statistics are likely to show splendidly large figures for serials readership: statistics

are needed to track progress and wider use – and to publicise this

• there are prospects for more and better data to measure the use of networked materials than there are for conventional – particularly on serials where the under-reporting of in-house use has proved an intractable weakness in academic libraries

• the importance of the distinction between full-text serials, indexing reference tools and other databases has been identified: there is a need to count in these categories as well as to monitor innovative products that straddle them

• statistics of expenditure will be the most practicable for some years yet – with better prospects of relating the use of serials to their cost

• the USA is well ahead in this area, but the main pioneering/experimental stage is over: now is the time to get counting, to publish and compare the results. Can resources be provided for active international monitoring and collaboration?

• the revised ISO 2789 is a high-level standard providing a useful overview of the whole situation; there are plenty of detailed problems ahead which are already being worked on!

References

(1) LIBECON 2000, *Library Economics in Europe: millennium study* (URL: http://www.libecon.org/default.asp)

(2) Kyrillidou, M. and Young, M. (eds) *ARL Supplementary Statistics 1999 – 2000*, Association of Research Libraries, Washington, DC.

———