Establishing digital collections in a scientific research library network: part one of a case study from CSIC, Madrid, Spain

This article describes the process that CSIC (Consejo Superior de Investigaciones Científicas) followed from 2001–2003 in developing a new centralized model to enable more effective acquisition of its traditional information resources (print and databases). It illustrates how, at the same time, it managed the migration of its collections from print to digital. The CSIC Network, the scholarly communications background, determining factors forcing the change, the changes and solutions found, the new collection model, the technical management of resources and their access, the new collections and the investment levels required to sustain them are all discussed.

AGNES PONSATI

MERCEDES BAQUERO
CSIC Libraries Co-ordination Unit
Madrid, Spain

The CSIC Library Network

The CSIC (Consejo Superior de Investigaciones Científicas) is a multisector, multidisciplinary public research entity attached to the Spanish Ministry of Science and Technology. It is a scientific institution which collaborates with the state, *comunidades autónomas* (regional authorities) and local authorities, as well as with other research institutions (universities, public and private research entities). It also supports social and economic organizations, both national and foreign, in the development of research programmes and in the provision of scientific and technical advice and support. It is made up of 100 institutes, each with a specialized library supporting research. The CSIC research areas are biology and biomedicine, food technology, materials technology, physical technologies, chemical technologies, agricultural sciences, humanities and social sciences, information sciences and natural resources.

These specialized libraries are organized into the CSIC Library Network situated in 21 cities of 10 autonomous authorities. Of these, 17 belong to centres based in different universities. As a whole, they are a highly specialized library resource of more than 1,400,000 monographs, and more than 44,408 periodical titles (but through duplication,

some 76,157 print subscriptions) as well as maps, photographs, archives and manuscripts. The collection of scientific journals is particularly remarkable for its size, its level of specialization, the wide chronological coverage and the considerable level of investment which this has entailed for the institution.

The Network uses the Aleph 500 library management system and has created one of the largest automated union catalogues of scientific information in Spain (CIRBIC, Computerized Catalogues of the CSIC Library Network). At present, it has 97% of its monographic collections accessible by the online catalogue, 100% of the journals, as well as a growing catalogue of maps, archives, authority records and electronic journals.

Background: the journals crisis

The key factor which influences the management of CSIC is the fact that its organization is dispersed throughout Spain. It was already quite clear in the mid-1980s that centralized collection management policy was essential to enable the automation of the library system. This lead to the creation of the

union catalogue. It is only recently, however, that it has been considered desirable to have a central-ized system for other aspects of library manage-ment, such as acquisitions.

In 2001, it became clear that the existing decen-tralized acquisition of the collection was unsustain-able and could not be allowed to continue for much longer. The exponential growth in cost of the journals was exacerbated by the high level of duplicate subscriptions, held in the CSIC libraries throughout Spain. Because of this decentralized model, CSIC could not take advantage of econ-omies of scale as a strategy to counteract the pricing polices of the majority of scientific pub-lishers. The existence of a number of cost centres had lead to a lack of co-ordination concerning the management of the journal collection for many years and this, in turn, had produced the unco-ordinated growth of the collection with a high number of duplicates. To all this, it is necessary to add the cost of administration in time and human resources which handling acquisitions in more than 100 different locations entailed.

From the early 1990s, the so-called 'journals crisis' was deeply felt in our institution, which was already suffering an overall budget crisis. This had already rung the first alarm bells on the impos-sibility of keeping up the journal collections in our libraries as they stood. Since then, the libraries have gradually begun to cancel titles and look for other means of financing apart from the ordinary budget in order to be able to keep up their collec-tions in face of pressure from researchers. In 1994 the managers of the CSIC Network warned that a change in the collection management model was urgently needed. This change involved looking for a single solution for all the libraries and the cen-tralization of budgets in order to establish a certain level of economy of scale. However, in spite of the urgency of the problem this call still did not arouse sufficient interest in the institution; it was not considered to be an institutional priority at the time and the problem only became worse.

Determining factors for facing the change

At the end of the 1990s the situation was really critical. The stagnation of the budget for purchas-ing journals lasting several years lead to a wide-spread situation within the CSIC Network where the ordinary budget could meet the cost of no more than 60% of annual subscriptions. By 2001, the low point had been reached – budgets were frozen and many of the alternative ways of financing, which had been organized over time, could not match the required funding.

A management policy for the collection, which was economically, administratively and techni-cally dispersed and unco-ordinated, had lead to a bleak scenario. The decentralized purchase model had produced fragmentation of the problem and, although the CSIC was a large client, it was dis-persed and without strength. It lacked an overall view of what was bought and this led to a failure of awareness of the seriousness of the problem.

Furthermore, there was little willingness in the libraries to accept the idea that it was unavoidable that the process of organizing the management of the collection entailed the elimination of a percen-tage of duplicates. Moreover, up to this time, the digital collections had not had the chance to be introduced and to be considered as a reliable infor-mation resource whose management had to be tackled with an institutional view without further delay.

Changes and solutions undertaken

CSIC introduced a number of changes in 2001 aimed at overcoming the financial and administra-tive difficulties inherent in the existing system. These changes would enable it to reorganize its management of its printed collections as well as take advantage of the opportunities offered by digital publishing

The changes that were put into place were as follows:

- budgetary concentration
- reduction of the administrative purchasing process
- introduction of the first access licences to e-journals for all the CSIC centres
- negotiation and management for the total and legal accessibility of e-journals
- promotion of a policy to eliminate print duplicates
- technical processing of e-journals
- dissemination of knowledge about new re-sources and training for users.

The 2001 strategy had the objective of not losing any more time in creating a digital journal collection which was 'nuclear' and 'collective'. It was designed to allow the incorporation, in the short to medium term, of those publishers which held the greatest interest for a significant number of our researchers and which would also allow access for the whole CSIC community. Therefore the combination of 'interest', 'accessibility' and 'cost' were the key factors in deciding on priorities when it came to purchasing future collections.

Once subscription data had been brought together on a central computerized system, the analysis of the print subscription collection was very important when it came to taking decisions as this reflected the actual demand in our community over many years. The introduction of the digital collections was doubly important as, firstly, it enabled us to meet the demand we had been experiencing for some time and we had not been able to satisfy. It also allowed us to create a digital access culture to the resources and, furthermore, it began to create awareness of the fact that if a digital collection existed, we could start up a policy of cancelling duplicates.

Negotiation procedures were first opened up with the large commercial publishers which in most cases offered a model based on the 'big deal', access to the complete platform in a 'cross-access' system. This made it possible to make the journal collections from Academic Press, Springer, John Wiley, Kluwer, Blackwell and Nature Publishing Group available in a couple of years. We left the introduction of collections from scientific societies such as American Institute of Physics and American Computer Machinery until later.

This incipient digital collection owed its rapid growth to the following combination of factors:

- the pressure from demand, which was beginning to be critical with a lack of certain resources
- the economy of scale resulting from the budgetary centralization, which created important savings in comparison to the decentralized model
- the concentration of contracting procedures for print subscriptions, which made it possible to begin to negotiate combined digital licences, some of which provided a budgetary cushion thanks to the print 'deep discount' in some cases

- an additional investment from the institution which decided to go for a centralized contracting model for information resources.

However, in spite of the fact that considerable ground had been gained in a short time, it was important to point out that the 'delayed' introduction of digital in the CSIC had a high cost because of the considerable burden of having to sit down to negotiate prices for electronic licences with publishers, since these prices were nearly always based on the existing historic collection in order to fix the 'base price'. We could have reduced this cost if the CSIC had taken up the policy of cancelling duplicates years before as this was already stated to be a necessity in 1994.

The new collection model

The desired collection model was a 'hybrid' collection made up of print and digital resources, in which there would be less print for a few and more digital for all by making the most of cross-access options – a collection which, in its print format, would eliminate as far as possible the average number of 1.5 duplicates per title. The aim was to do all this with a non-traumatic transition from the culture of 'print' use to 'digital' use to smooth the way towards a collection which had to be principally digital.

The model chosen to channel purchasing negotiations for the different digital collections was a model based on 'print+electronic', which offered the possibility of cancellation and the existence of a 'deep discount price' on paper subscriptions. We estimate that in the medium term it would be possible to obtain a model based on 'electronic+ paper' with the aim of only keeping the institution's basic and historic collections in this format. Priority would be given to those initiatives which offered complete packages from the publishers at reasonable prices rather than a model which only considered inclusion of access for subscriptions which already existed. A study would be carried out later to decide whether this model was fitting or not by analyzing the use that had been made of these collections throughout 2002 and 2003.

In spite of the fact that a purchasing policy of back-files of the collections purchased has been highly recommended, in the early negotiations

this was not taken up because of lack of funds. This is an unresolved question which the CSIC would have to tackle some time in the future. It was important to follow a purchasing and management policy which was economically sustainable over time and which would guarantee long-lasting access and/or contents.

The negotiation procedures with each of the publishers which we wanted to introduce have been crucial for developing the digital collection. It is acknowledged it will take a long time to finalize negotiations with all the publishers involved, given the scope of CSIC's interests and responsibilities. An effort is always made to compare data offered by the publishers and agents with the idea of tightening up the financial proposals as much as possible. Negotiations were conducted bilaterally with publishers and agents although in some specific cases some deals were negotiated under the option of 'open purchase consortia' with other Spanish library consortia.

Negotiations were always conducted using the following criteria:

- the level of CSIC interest in the collection to be contracted
- the cost involved in the introduction of this digital collection
- the increase in relation to the existing cost in print
- access conditions
- the terms of the licences
- the duration of the agreement reached.

Technical management of resources and accessibility

The creation of the digital collection necessitated new cataloguing procedures in that a new type of resource – electronic journals – had to be made available via the CSIC OPAC. The challenges posed by this type of resource were varied. On the one hand we had to consider:

a) how to promote accessibility to the digital collection as much as possible, to stimulate the highest level of use possible in order to optimize the investment
b) how to achieve maximum integration of the available tools and resources to achieve the most transparent navigation possible.

The first decision taken was to catalogue the new digital titles (3,500) in the OPAC as independent entries with links to the complete text through the 856 field. The OPAC was not just to be a reflection of our printed collection but rather, it should also show the availability of the digital collection. For this purpose a new interface was also created for our OPAC for specific access to our digital collection (http://aleph.csic.es/F?func=file&file_name=find-b&local_base=electronicos).

Moreover, it was considered fitting to construct an A–Z list of titles (http://www.csic.es/cbic/revelectronicas/ejournals_A.htm) so that the user could also have access to this collection from the Library Network Portal. The entry-points to this collection were therefore varied.

In order to improve the level of integration and transparency between the discovery tools and the aim of the search, it is planned to install a system which supports federated searching and utilizes a dynamic link resolver.

Another difficulty which the CSIC had to face in the integration of the digital collection into its library network was the fact that the community of 5,000 users to which it has provided a service is spread all over Spain. Access takes place from more than 100 service points and is always based on IP address authentication. In order to solve potential access problems for remote access and sub-net users, CSIC, together with the departments which are managed by the Academic Network in Spain (IRIS), set up an authentication server which meant that these users could be given authentication as CSIC users. This service, called PAPI (Access Point to Information Providers), has enabled us to eliminate barriers caused by the inequality of infrastructures and resources in the CSIC.

The current CSIC collection and investment levels

The CSIC collection is multidisciplinary, highly specialized and comprehensive in terms of coverage and length of holdings. It is made up of about 3,545 printed titles (5,465 subscriptions) and an ordinary digital collection of 4,325 electronic titles. The 'core' collection is made up of titles from Wiley, Kluwer, Springer, ScienceDirect, PCI-FT, Blackwell, AIP, APS, ACM, IOP, NPG and Project

MUSE, to which we now add 1,370 titles from collections for individual use in some libraries in the Network.

In 2001, the time at which we undertook the change in management, investment in information resources was at €5,280,381. Three years later, it was at €6,248,127.88, an increase of 18.3%. If we take into account the development of investment in digital collection, the figures are: €264,936 in 2001, €1,274,165 in 2002 and €1,426,077 in 2003, which has meant a percentage increase in the investment of 438%. It must be remembered that, in spite of this sharp increase which was due to the lack of initial investment, at this time the percentage of investment devoted to the digital collection is about 29% and that devoted to the paper collection is 71%. The investment has risen, in this initial period 2001–2004, from 11% to 29%, and has dropped from 89% to 71% respectively.

Results obtained

The introduction of the new management model has allowed a rapid introduction of a digital collection into the CSIC Network. The benefits brought by this are quite clear:

- setting up access to a digital collection which has gone from 0 to 4000 titles in the period 2001–2004
- CSIC now has data which allows it to answer the questions: What do we subscribe to? How do we subscribe to it? Who subscribes to it in our Network? How much does it cost us?
- we have data on the use of our digital collection which is analyzed later on [see part two in next issue] and which is highly useful as a guideline for its growth
- in this period we have been able to obtain a certain level of saving, in particular in investment in the printed collection which we have been able to redirect to investment in the digital collection at the same time as the total cost has been showing sustained growth. This has all lead to an increase in the institutional investment

- in parallel, the 'community' perception of the value of the 'intermediary' from the library as lender of digital services has increased considerably
- the level of satisfaction of our end-user is reasonably good. The community perceives the value of the work carried out and the role played by the library in this effort
- another indicator which we find positive is that the demand for interlibrary lending (a service with a high volume of transactions in the CSIC), which had been growing until 2001, dropped nearly 33% at the end of 2003.

However, it should be remembered that any centralization process also brings certain areas of dissatisfaction caused by a feeling of loss of independence in decision-making and by the impossibility of dealing with all the information needs of a community like ours which is broad, specialized and heterogeneous. We are aware that, in spite of the rapid growth of the digital collection, the levels of coverage are still not ideal. We still have to tackle the incorporation of small collections of titles published by prestigious scientific societies.

All in all, the balance is extremely positive. The introduction of the digital collection has meant an act of political conviction, a change in administrative management, changes in managing the collection, changes in the technical procedures, changes in the mentality of the users and librarians, introducing a new image of the library and its services, changes in the usability of the collection.

Part two of this study will appear in the next issue: *Serials*, 2005, 18(1).

■ **Agnès Ponsati**
Director, CSIC Libraries Co-ordination Unit
Mercedes Baquero
Serials Manager, CSIC Libraries Co-ordination Unit
Unidad de Coordinación de Bibliotecas del CSIC
C/Jorge Manqrique 27
28006 Madrid, Spain
Tel: 34-91-5855076 & 8
E-mail: Mbaquero@bib.csic.es; agnes@bib.csic.es
http://www.csic.es/cbic/cbic.htm