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

ABSTRACT:
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.

The CSIC Library Network

The CSIC is a multi-sector, multi-disciplinary public research entity attached to the Spanish ministry of Science and Technology and it is established throughout Spain. It is a scientific institution which collaborates with the state, autonomous and loca authorities, with other research institutions (universities, public and private research entities) and with social and economic organisations both national and foreign, to which it brings its research capacity and its human resources and materials either in the development of research programmes or as scientific and technical advice and support.

It is made up of 100 institutes with a specialized library which supports research. These, in turn are organized into the CSIC Library Network made up of 100 specialized libraries distributed between 21 cities of 10 autonomous Communities, 17 of these belong to mixed centres with different universities. In coordination they make up one of the largest contributions to the system of scientific information in the country; highly specialized library resources of more than 1,400,000 monographs and more than 44,408 titles corresponding to 76,157 journal collections as well as other types of documental material such as maps, photographs, manuscripts, etc.

The distribution by subject is in accordance with CSIC research areas; biology and biomedicine, food technology, materials technology, physical technologies, chemical technologies, agricultural sciences, humanities and social sciences, information sciences and natural resources.

The Network manages its services and catalogues with the Aleph 500 system creating one of the largest automated union catalogues of scientific information in Spain (CIRBIC, Computerized Catalogues of the CSIC Library Network).
In 2004 it has 97% of its monographic collections automated and 100% of the journals, as well as a growing catalog of maps, archives, authority records and electronic journals.

The collection of scientific journals is particularly remarkable for its volume, its level of specialization, the wide chronological coverage of its collection and the considerable level of investment which this has entailed for the institution.

Background: the journals crisis

If there is one factor in particular which constantly conditions the management of every aspect of the CSIC this is the fact that the elements of which it is composed are dispersed throughout Spain. While it was already quite clear in the mid-eighties that a sole and centralized management system was essential to tackle the automation of the libraries (the creation of the union catalog), in other aspects of library management, such as acquisitions it is only recently that there has been a view which considered a unitary treatment to be advisable.

In 2001 the existing decentralization in the process of acquiring the collection clearly showed that the existing model was in crisis; that it was unsustainable and could not be kept for much longer. As well as the exponential growth of the costs of the journals there was the extremely high level of duplications as a result of the physical dispersion of the centres.

With this decentralized model the CSIC did not participate in a strategy of scale economy to be able to face the monopolistic situation of the majority of scientific publishers. The spread of costs had lead to a lack of institutional coordination concerning the management of the journal collection for many years and this had produced the erratic growth of the collection with a high number of duplicates.

To all this, it is necessary to add the cost of administrative management in time and human resources which handling acquisitions in more than 100 points of purchase entailed.
The panorama from the early nineties was quite frankly difficult. The so-called “journal crisis” was deeply felt in our institution which was already suffering a budget crisis which had rung the first alarm bells on the impossibility of keeping up the journal collections in our libraries. 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 collections in face of pressure from researchers.

In 1994 the people responsible for the network warned that a change in the management model for the collection was urgently needed. This change involved looking for a sole solution for all the libraries and the centralization of budgets in order to recover 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 got worse.

Determining factors for facing the change

At the end of the nineties the situation was really critical. The stagnation of the budget for purchasing journals lasting several years led to a situation which was quite widespread. With the ordinary budget it was not possible to pay more than 60% of annual subscriptions.

It was necessary to wait until the beginning of 2001 for the turning point which would enable a change of direction. We had reached the lowest point, the budgets were frozen and many of the alternative ways of financing which had been organized over time did not attain the level of prices either.

A management policy for the collection which was economically, administratively and technically dispersed and erratic 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 dispersed and without strength. It lacked a global 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 % 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 informative resource whose management had to be tackled with an institutional view without further delay.

Changes and solutions undertaken

Changes which had already been introduced in 2001 had an effect on contracting resources for 2002. This involved a battery of measures which at first aimed at ordering and systemizing management of the economic and administrative “disorder” by introducing factors of organisation in other aspects of the management of the collection.

At that time it seemed that the digital publishing market already offered a certain level of maturity and the CSIC no longer wanted to let slip by the chance to begin to build up its future digital journal collection at the same time as it reorganized the management of the printed collection. For this purpose the following changes were set up:

a) budgetary concentration
b) reduction of the administrative purchase process
c) introduction of the first access licences to electronic journals for all the centres
d) negotiations and management for the total and legal accessibility of e-journals
e) promotion of a dynamic of eliminating paper duplicates.
f) technical processing of e-journals
g) disseminate knowledge about new resources and users training

The strategy which was thought up in 2001 had the objective of not losing more time in creating a digital journal collection which was “nuclear” and “collective”, which would 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 community. Therefore the combination of “interest”, “accessibility” and “cost” were the key factors in deciding on priorities when it came to purchasing future collections.

The analysis of the collection for which there were already subscriptions on paper was very important when it came to taking decisions (once the subscription data was centralized and computerized) as this reflected the actual demand in our community over many years. Introduction of the digital collections was doubly important as it enabled us to satisfy the demand we had been receiving for some time and we had not been able to satisfy, it 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 and Blackwell, Nature Publishing Group available in a couple of years.

We were leaving the introduction of collections from scientific societies, American Institute of Physics, American Physical Society, Institute of Physics, American Computer Machinery…etc. until later.

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

0) The pressure from demand which was beginning to be critical with a lack of certain resources
1) The economy of scale resulting from the budgetary centralization created important savings in relation to the decentralized model.
2) The concentration of contracting procedures for paper subscriptions 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.
3) 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 has 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.

A new collection Model

The sought after collection was a “hybrid” collection made up of paper and digital, in which the balance of what was invested in one format or another would have a greater equilibrium, in which there would be less paper for a few and more digital for all making the most of cross access. A collection which, in its paper format, would eliminate as far as possible the average number of duplicates of 1.5 per title. The aim was to do all this with a non-traumatic transition from the culture of “printed” use to “digital” use to smooth the way towards a collection which had to be basically digital.

The model chosen to channel purchasing negotiations for the different digital collection was a model based on “paper+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” to be able to only keep the institute’s basic and historic collections in this format. Priority would be given to those initiatives which offered complete packets from the publishers at reasonable prices rather than a model which only considered inclusion of access for subscriptions which already exist. A study would be carried out later to decide whether this model was fitting or not by analyzing the uses which had been made of these collections throughout 2002 and 2003.

In spite of the fact that a purchasing policy of Backfiles of the collections purchased has been highly recommended in the early negotiations this was not taken up because of budgetary availability. We knew that this was an unresolved question which the CSIC would have to tackle some time in the future. The important thing was to follow a purchase and management policy which was economically sustainable in 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 introducing the digital collection. It will take several months/years to become a reality owing to the size of the collection, the magnitude of the institution and the volume of the subscriptions. An effort is always made to compare data offered by the publishers and agents with the idea of tightening up the economic proposals as much as possible.

In general negotiation procedures were decided on bilaterally with the publishers and agents although it should be mentioned that in some specific cases some platforms were negotiated under the formula of “open purchase consortiums” with other Spanish library consortiums.

The framework for negotiations was always fixed with a combination of a series of basic parameters: the interest of the collection to be contracted, cost involved in the introduction of this digital collection, increase in relation to the existing cost on paper, access conditions, terms of the licences and duration of the agreement reached.
Technical Management of resources and accessibility

The incorporation of the digital collection introduced changes to the technical management of bibliographic procedures; it involved beginning to process a series of resources catalogographically which was new for our OPAC.

The challenges posed by this type of resources were varied. On one hand we had to consider:

a) how to promote fully accessibility to the digital collection reaching the greatest use 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 label. We did not want a reduced OPAC to be merely a reflection of our printed collection but rather that it should show the real availability of the digital collection.

For this purpose a new interface was also formulated for our OPAC for specific access of 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 entrance ports to this collection were therefore varied and even at the risk of the user’s possible confusion about the way to gain access it seemed to us that at first it was convenient to be redundant.

With the idea of harmonising the routes by which it will be possible access this collection in the future and of offering a search and navigation tool with good functions and levels of navigation it is clear that it is necessary to implant a system which improves the level of integration and transparency between the tools and the final objects. For this purpose the installation of a system which supports federated searches and integrates a dynamic link resolver is planned.

Another difficulty which the CSIC had to face in the implantation of the digital collection in its library network was precisely the fact that the community of users to which it has provided a service (5,000) is spread all over Spain. Access takes place from more than 100 service points and is always based on IP recognition. The question is that many CSIC libraries use varied communication networks and in some cases because of the physical locations of the centres on university campuses they use other sub-communication networks. This situation meant that in some cases there were problems of authentication of the CSIC centres as such by some publishers. In order to solve this type of problem and be able to also offer the community remote access to the digital collection, the CSIC together with the departments which are managed by the Academic Network in Spain (IRIS) set up an authentication server which meant that this group of centres 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.

What the CSIC collection is like at this time. Investment levels

The CSIC collection is multidisciplinary, highly specialized and very complete in a temporal coverage of collections. It is made up of about 3,545 printed titles (5,465 subscriptions) and an ordinary digital collection of 4,325 electronic titles.

The “nuclear” collection is made up of titles from: Wiley, Kluwer, Springer, ScienceDirect, PCI-FT, Blackwell, AIP, APS, ACM, IOP, NPG, Project MUSE to which we now add 1,370 titles from collections for individual use in some libraries in the Network.

The following data can be provided on the level of investment involved in keeping up the collection (paper and digital): in 2001, the time at which we undertook the change in management, investment in information resources was at 5,280,381 €, 3 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 data are, 2001 264,936 €, 2002 1,274,165 € and 2003 1,426,077 € 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 inexistence of initial investment, at this time the % of investment devoted to the digital collection is about 29% and that devoted
to the paper collection is 81%. The investment has gone, in this initial period 2001-2004, from 11% to 29% and from 89% to 71% respectively.

**Results obtained**

Implantation of the new management model has allowed a rapid introduction of a digital collection in our Network. The benefits brought by this situation are quite clear;

Setting up access to a digital collection which has gone from 0 to 4000 titles in the period 2001-2004. Thanks to the procedures for centralizing management we now have data which allows us 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 also have data on the use of our digital collection which is analyzed later on and which is highly useful as a guideline for its growth.

Furthermore, 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’s” perception of the value of the “intermediary” from the library as lender of digital services has increased considerably.

The level of satisfaction of our final 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 the demand for interlibrary lending (a service with a high volume of transactions in the CSIC) which had been growing until 2001 has 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 informative 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.