

# THE USE OF NEW INFORMATION TECHNOLOGIES AT INIST-CNRS

By Christian Lupovici

**Abstract:** INIST is an information and document supply factory. It was founded in 1988 by the French CNRS. It was stated that its information system was to integrate the whole documents processing.

INIST has launched a full automation program which integrates acquisition policy, cataloguing, indexing, catalogue and database production. The information is used also to support the document supply management from the users request in electronic form to the document delivery. Two hundred titles of serials have been scanned to be managed in electronic form. Document transmission will be the next step toward a full electronic process.

## Introduction

The Institut de l'Information Scientifique et Technique (INIST) was created by the French National Center for Scientific Research (CNRS) in 1988. It is thus a fairly new institute although it continues and pursues the work of the two former documentation centers of the CNRS in Paris.

The creation of INIST partakes of the policy to develop, in France, a multidisciplinary and international center for the creation and diffusion of scientific and technical information, associated with a document supply system. For this, it was necessary to create an information production and service chain, using the most recent technology.

At each stage of the production of information, a specific technology has been adapted and integrated into the general information flow, from document acquisition to database production and document supply.

The INIST production chain can be divided into 4 systems which create and

---

Mr. Lupovici's paper was delivered at 57th IFLA Council and General Conference, Moscow, USSR, 18-24 August 1991 and was revised for this issue of INSPEL.

manage their own information while at the same time putting to use information already available in the overall information system.

These 4 systems are:

- document management,
- analytic cataloguing and indexing,
- digital storage,
- order management and document supply.

### **1. Document Management**

The first part of the documentary production chain functions in some ways like a library. INIST uses the modules of a GEAC 9000 bi-processor system, some functionalities of which were specifically developed for INIST.

Approximately fifty terminals are shared between the Nancy and Paris facilities and are connected to the central unit. But, there is no public access to the GEAC database which is for internal use only.

Even if there is no gateway with the suppliers, the Acquisition module is used to build up the base serials candidate for acquisition, pending evaluation by the specialists of the Database department.

Cataloging is done using the UNIMARC format and follows the AACR-2 rules and, for the access points. For the access points, the writing rules of the Bibliothèque Nationale are observed. The cataloging constitutes an important stage of the process for it is the first generic bibliographical data of the INIST documents. This same data is then reused all along the production chain from checking-in, to analytic cataloging of periodical or proceedings articles, all the way to indexing for the PASCAL (science, technology, medicine) and FRANCIS (humanities) databases. This same information also feeds into the document supply system.

Every document is identified by a bar-code which is assigned during cataloging for monographs, or during checking-in for periodical issues. This identifier is used in the lending module, and also as a first part of the address on the digital optical disks for the articles which are digitized.

The circulation of the documents is organized and planned according to the characteristics which are attributed to them in the bibliographic format, in local zones (9xx). Each particular detail concerning its use is coded into a specific zone.

Thus, the day following checking-in for serials or cataloging for monographs, document can be redirected to their next destination.

Circulation of the 400 periodical issues checked into the system each day is as follows:

- 20% will go to be cataloged at the article level for the databases,
- 20% will to to be cataloged at the article level for the databases, and then also digitized.

The remainder is shelved in the stacks.

The proper dispatching of the documents can be checked on printouts, and magnetic tapes with the cataloging data are sent along with the documents which are to be analytically cataloged for the databases and digitized.

All documents leaving the stacks are subject to a lending procedure. In this way, not only is it possible to locate precisely all documents in circulation, but also to check that lending periods are respected, during analytic cataloging, digitalization and indexing.

## **2. Analytic cataloging and indexing**

The analytic cataloging is contracted to a private company, JOUVE, which uses a formatted data entry module, implemented on IBM/PC type stations. This system, called PSILOG, is able to merge the description of each article to the generic description of issues coming from the GEAC system and supplied in UNIMARC format. PSIDOC is a system with windowing features. Each window corresponds to a field or sub-field of the format. The data entry ergonomics is made more efficient with the addition of a SGML type structured format, and the data is loaded in the same format into the information specialists' server at INIST. In this way, the cataloging of articles is shared by the different workstations of the information specialists who do the indexing work on screen. All the workstations are connected by a NOVEL local area network. The indexing work done manually on paper worksheets progressively diminishes giving way to working directly on the screen. A new workstation which will allow for local vocabulary management and indexing control is under study.

Furthermore, an experiment in the automatic indexing of the words of author abstracts has begun. From digitized documents, a text structure recognition system selects the abstract and optical recognition of character (OCR) is carried out. Each word of the abstract is then indexed, and compared to the vocabulary authority file. The information specialists have the possibility of adding or deleting terms. Once the bibliographic record is complete, with its abstract and key words, it will

feed into the updating base of the PASCAL database. Updating tapes in PASCAL format are sent each month to the hosts. Changing over to the CCF exchange format is being studied.

### **3. Digitization and digital storage**

INIST has selected slightly less than 2000 periodical titles for an electronic document delivery project. The titles concerned are those titles most often requested in document supply and, therefore, represent the international core literature. The selected documents are systematically digitized (scanned), and in analogical mode.

Apart from the technological challenge this represents, several problems must be solved simultaneously:

- a) ensure that these periodicals are always available for document supply and for analysis for the database;
- b) guarantee rapid good quality fully automated reproduction of the documents;
- c) permit transmission of the documents over telecommunication networks so that delivery time can be reduced;
- d) save storage space in the long run.

This digitization began in 1990 and represents a volume of 4 million pages per year. The JOUVE company is also responsible for the digitization operations.

INIST receives two 8mm cassettes daily on which the compressed images in TIFF format according to Group IV CCITT standard are stored. Image resolution is done at 300 points/inch. A magnetic tape with the data management information used to locate the articles in digital storage server database is sent along with the images of the articles. After structure and image quality controls are carried out on Powermate stations, the cassettes are copied onto WORM type 6.4 gigabytes digital optical disks (DON), with a capacity of about 70 000 pages each. The disks are loaded into a Cygnet jukebox equipped with 2 ATG60001 disks drives and which can contain up to 131 disks. The system, supplied by the MC2 company, is controlled by a SUN 4/370 (under (ORACLE/UNIX) workstation linked to the document supply system installed on an IBM 4381.

Three peripheral micro-computers (Powermate 2) equipped with telecommunication boards permit transmission over public telecommunication networks in CCCIT Group IV and Group III standards as well as transmission over the French RNIS (ISDN) network, NUMERIS.

All the units (including the SUN/IBM) link) are interconnected by a Ethernet local area network.

#### **4. Orders management and document supply**

Four phases can be distinguished in the document supply process:

- receipt of the order,
- processing the order,
- filling the order,
- billing.

- Nowadays, automated document ordering is quite common. The transmission of orders can be done either by downloading from the hosts (QUESTEL, ESA, DIALOG ...), or by electronic interlibrary systems such as the PEB in France or OCLC. INIST further allows occasional users to order directly by Minitel (VT 100) and is developing a file transfer procedure for its major customer. The INIST collections, catalog, loaded on the IBM server, serves to locate the documents, while at the same time acting as the OPAC for remote access by the public. It gives access not only to the description of actual documents, but also to the articles in them. Orders on paper forms, by fax or by telephone, are entered in the automated order management system so that everything is uniformly administered.

- The system manages the processing of orders. It sorts them according to the documents shelving numbers, and dispatch them to their location. Priority is given to the digital storage server.

- If the document exists in digital format, it will be sent on the appropriate telecommunications network to a Group IV or Group III fax. If not, it is printed on a highspeed laser printer. If the document is not in electronic format, the request is photocopied on the floor where the paper document is shelved. A special service processes the requests which are sent to it by the system in order to deal with the rejects and the problems of identification of documents.

- The entire document supply process is monitored by the system, including re-dispatching the request to an alternative library, and generating a letter to inform the customer.

- Billing data are entered at the time of shipping for documents sent by mail, and automatically when the order is received for documents sent by fax.

- Monthly account statements are automatically sent to the customers who have a deposit account with INIST.

The system is designed so that, when the requested document is electronically stored, it can be supplied on a telecommunication network, from order to recep-

tion, without any human intervention. Furthermore, the system can produce statistics needed not only for production management but also for collection development.

### **Conclusion**

As it is a new institution, the INIST was able to invest in a computerized and well integrated processing chain, example of the state of the art in 1990.

Already, at this point, plans are under way for new services:

- a table of contents file for the digitized titles;
- integration on the same workstation of on-line searching, document ordering and reception.

Nevertheless, the development of these new services depends on the reliability of the technology used in the production line and on the growth of telecommunication networks, in particular of ISDN's, such as NUMERIS in France, and on their interconnection capability on an international level. It also depends on the development and compliance with standardized formats and protocols. It is for this reason that INIST carefully respects the OSI standards and actively participates to international standardization. One can nevertheless be optimistic because demand is strong, technical solutions are becoming common and more cost-effective.

Christian Lupovici  
Institut de l'Information Scientifique et Technique  
2, allée du Parc de Brabois  
F-54514 Vandoeuvre-les-Nancy Cedex  
France