3 Problems and issues in CIPAC implementation

In this chapter, the main problems and issues involved in the creation and implementation of CIPACs are identified and discussed. It is based upon more than half of the CIPACs depicted in the previous chapter,¹ and primarily on two sources of information:

- the CIPAC Library Questionnaire (CLQ), and
- relevant project literature (reports, articles, papers, information sheets).

23 libraries (of 38 that had been contacted)² returned a CLQ and for 20 CIPACs some kind of project literature was available. Because of the overlap between the two types of sources the following account is based upon statements on 28 CIPACs.³ In addition, some information from CIPAC web-pages was used, as well as some other literature.

3.1 Reasons for establishing CIPACs

The results of the CLQ confirm that the four aspects listed in the questionnaire, namely

- cost-effective / moderately priced method
- relatively fast way of converting a card catalogue
- savings in space (getting rid of card cabinets)
- universal access to the catalogue via Internet/WWW

were indeed the most important reasons why the respective libraries chose the CIPAC approach for the conversion of their card catalogues. It seems that cost, speed and universal access were about equally important (applying in most cases), whereas the space saving factor – although sometimes a crucial aspect – was, by and large, slightly less significant.

Another motive that was mentioned repeatedly is catalogue preservation, especially when old catalogues existed only in one copy and the digitization of the cards was also regarded as a measure for data security. Only in a few cases each of the following reasons were mentioned: Improving catalogue searching; the only choice in the case of

¹ 50 CIPACs plus the "future" CIPAC of the Federal Maritime and Hydrographic Agency of Germany (for which a LCQ with useful information was received).
² As mentioned in Chapter 1, only institutions with CIPACs known of by mid-2001 were sent a CLQ. The questionnaire is included as Appendix B1.
³ For details see Appendix B2.
handwritten catalogues; public relations; good experiences of other libraries; digitization as the groundwork for a "real" conversion.

3.1.1 Cost

Undoubtedly, the creation of a CIPAC is less expensive than "real" retroconversion, but the difference in cost is not so easy to determine. Practically all sources claim that CIPACS are considerably cheaper; in some cases they were described as "the only affordable" or "the only financially feasible" way of converting the card catalogues into an online format. When comparisons with other conversion methods were made, the estimates varied between "twice the price" (of using unskilled staff or students) and "ten times the price" (of a professional retrospective conversion).

<table>
<thead>
<tr>
<th>Location</th>
<th>Institution</th>
<th>Cost (€)</th>
<th>Cost factors included</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vienna</td>
<td>Univ. Econ. L.</td>
<td>7,267.-</td>
<td>not specified</td>
<td>CLQ</td>
</tr>
<tr>
<td>Zurich</td>
<td>Central Lib.</td>
<td>25,613.-</td>
<td>'complete solution' (not specified)</td>
<td>Anon. (1997b)</td>
</tr>
<tr>
<td>Bmo</td>
<td>Moravian Lib.</td>
<td>7,632.-</td>
<td>scanning, hardware, software, external staff</td>
<td>CLQ</td>
</tr>
<tr>
<td>Prague</td>
<td>Nat. Library</td>
<td>11,346.-</td>
<td>scanning, implementation</td>
<td>CLQ</td>
</tr>
<tr>
<td>Prague</td>
<td>Parliamt. Lib.</td>
<td>13,609.-</td>
<td>'complete solution' (not specified)</td>
<td>CLQ</td>
</tr>
<tr>
<td>Berlin</td>
<td>Central Lib.</td>
<td>11,504.-</td>
<td>scanning, software, internet connectivity, security filming, setting up server</td>
<td>Rönsch (1998), CLQ</td>
</tr>
<tr>
<td>Frankfurt</td>
<td>HeBIS-Retro</td>
<td>12,526.-</td>
<td>scanning, OCR, categorization, project management, quality control, database loading,</td>
<td>Dugall (2001)</td>
</tr>
<tr>
<td>Halle</td>
<td>Univ. Library</td>
<td>5,420.-</td>
<td>not specified</td>
<td>CLQ</td>
</tr>
<tr>
<td>Hamburg</td>
<td>Mar. Agency</td>
<td>11,003.-</td>
<td>scanning, indexing, ‘additional cost’, software</td>
<td>CLQ</td>
</tr>
<tr>
<td>Heidelberg</td>
<td>Univ. Library</td>
<td>5,537.-</td>
<td>scanning, indexing</td>
<td>Dörpinghaus (1998)</td>
</tr>
<tr>
<td>London</td>
<td>BLPES</td>
<td>11,136.-</td>
<td>scanning, additional server capacity, inhouse staff time</td>
<td>Price (2000)</td>
</tr>
<tr>
<td>Florence</td>
<td>Marucelliana</td>
<td>10,866.-</td>
<td>'complete solution' (not specified)</td>
<td>Lunati (2001)</td>
</tr>
</tbody>
</table>

Table 3-1: Comparison of CIPAC costs, per 100,000 cards (in Euros)

Some more concrete cost figures are also available, both from the project literature and the CLQ. However, these figures are rather difficult to compare, because they vary with regard to (a) the cost factors covered (e.g. in-house costs are often not included), (b) kind of CIPAC (e.g. a binary search system does not need an index), (c) currencies (some of which may have fluctuated over the years), and (d) time (e.g. scanning may have become cheaper during the past few years). Generally, they are not very precise either (e.g. in most cases it remains unclear if VAT – which also differs from country to country – is included or not). Nevertheless, in order to provide at least some kind of
overview all these figures were converted to Euros\(^4\) and standardized for a CIPAC size of 100,000 cards. The results of this attempt are given in Table 3-1 that also indicates which cost factors were covered by the respective figures.

In spite of all the shortcomings mentioned above, the table does indeed reflect the cost differences between the relatively cheap KatZoom system (Vienna), the systems with indexes or virtual drawers, and the rather expensive Spider system (Zurich). It also leads to the assumption that the actual cost for an average partial index or virtual drawer system can be estimated at somewhere in the region of €11,000.- for 100,000 cards, or 11 Eurocents\(^5\) per card (VAT not included).

![Table 3-2: Cost factors for HeBIS-Retro, per 100,000 cards (in Euros)](image)

Even the costs for the more sophisticated HeBIS-Retro system are about in the same region. The somewhat higher figure includes more cost factors, as shown in Table 3-2. It is interesting to note that the cost figure given for scanning (€0.041 per card) is roughly comparable with the one shown for the Austrian National Library in Table 3-1 (€0.036), whereas Price (2000) reports a much higher amount (€0.099) for the scanning of the BLPES' card catalogue by a UK company.

### 3.1.2 Speed

Many CIPAC libraries felt the need to have all their catalogue records available electronically as quickly as possible. Card-image online catalogues can indeed be created and made available over the world-wide-web in a very short time.\(^6\)

\(^4\) For this exercise, the conversion rates published on 20/03/2002 were used. Figures originally given in British pounds, Czech crowns, and US dollars may have led to slightly higher or lower costs in Euros if currency rates from a different day would have been used.

\(^5\) Approx. 7 UK pence.

\(^6\) For example, the (small) Graz University of Arts Library needed for scanning of its less than 50,000 cards and the development of its simple, home-made browsing software only five months (CLQ). Within more or less the same period of time, the (large) Austrian National Library managed to scan and
Of all activities related to CIPAC creation, *scanning* – even of large numbers – of catalogue cards seems to be the speediest task. The Theological Library at Innsbruck scanned its 190,000 cards within two weeks (CLQ); the Federal Maritime Agency at Hamburg was twice as fast: 350,000 cards were done in only one week (CLQ). Quality control, creating the database, setting up the server, testing the CIPAC – all this takes much longer (e.g. in the case of the Hamburg Agency about half a year). Large institutions such as the Austrian National Library and the Berlin Central and Regional Library scanned their catalogues in daily batches of 40,000 to 60,000 cards (Dikovich & Wilhelm, 1997; Rönsch, 1998). If done off-site, scanning obviously requires more time, as in the case of the Bavarian State Library which shipped over two million cards in five batches from Munich to Berlin (Haller, 1997), or the University of London Library that dispatched its 540,000 cards in batches of 50,000 for scanning (for which five months were planned; CLQ).

Pietzsch (2001a) points out that *OCR processing* is more time-consuming as one might expect. On average, per image (=catalogue card) seven seconds are required for reconstructing the text with OCR software on a Linux system. In the case of the Heidelberg University Library, more than three months (in day and night shifts) were needed for OCR processing of 1.2 million cards. At the Zurich Central Library, 100,000 cards were scanned and OCR read per week (Anon., 1997b), which is about the same speed.

Complex solutions such as *HeBIS-Retro* take more time. The conversion of eight large catalogues (about 8 million cards) took about three years, plus an additional year for a public tender and various test installations (Dugall, 2001). Nevertheless, this can still be considered as fast when compared with "normal" conversion. For example, Wicke (2000) reports that at the Dresden University Library six professional librarians (FTE) converted 266,000 records (350,000 volumes) in seven years, and Sosna (1997) estimated that his 1.5 FTEs at the Czech Parliamentary Library would need more than ten years for converting 200,000 volumes. According to Pietzsch (1998a), a full retrospective conversion (probably recataloguing)\(^7\) of one million cards would have required between 50 and 100 person-years, whereas the conversion of the same catalogue into a CIPAC was achieved within a few months.

---

\(^7\) In the case of old German catalogues there is also the issue of converting PI-based cards into RAK-based records which often cannot be done without having the item being catalogued to hand.
3.1.3 Access

Putting a card catalogue "on the Internet" or, correctly, the WWW, means to make it accessible universally, i.e. independent of location and time (e.g. a library's opening hours). However, not all institutions that created CIPACS had done their original planning with the web in mind. In about a quarter of the cases, solutions on in-house networks were first considered and/or implemented (one library even thought of microfiche first), but sooner or later replaced by web-based CIPAC solutions. When it was first released to the public, the Princeton CIPAC was to be used on dedicated workstations, which were equipped with a special image viewing software, on a campus network (Henthorne, 1995). This was probably the obvious thing to do in 1993/94; however, in the second half of the 1990s the reluctance of some libraries to opt for a web-based solution straight from the beginning is more difficult to understand. Nevertheless, for the majority of libraries the Internet/WWW was an absolute pre-requisite and the only option ever considered.

3.1.4 Space

In many cases the aspect of saving space by removing the old card catalogue(s) was an important reason for opting for a quick CIPAC solution. Most libraries suffer continuously from a shortage of space, and those that actually removed their card cabinets after the CIPACS went online gained at least one large room that could be used for other purposes (often for OPAC workstations). In some cases, the libraries moved to new or refurbished buildings and aimed at getting rid of their old cabinets on that occasion (e.g. Kiel, Dresden, London School of Economics).

What happened to the old card catalogues? In most of the cases they were removed from the reference section but kept in some other place (off-site, stack area, basement, depository, etc.) Only in a few instances the old catalogues were actually destroyed; e.g. at the Berlin Central and Regional Library the cards were "pulped" some time after the CIPAC had been introduced (Rönsch, 1998; CLQ). At the Austrian National Library the removal of the public card catalogues had caused some criticism by users and even the press (CLQ); later a rather unusual solution was found: The catalogue\(^8\) was first exhibited at the Museum of Applied Arts in Vienna; later the newly discovered "object

---

\(^8\) 84 wooden cabinets with over 3,000 drawers
of art" travelled, in spite of its ten tons, to various museums in Austria, Germany and the Czech Republic (Schnelling, 1999).  

3.1.5 Preservation

In many cases, even when libraries wanted to get rid of the old catalogues, the creation of CIPACs was seen as a measure of catalogue preservation. For example, Stoklasová mentions that in the case of the Czech National Library's general catalogue the "cards themselves are historical artefacts and, as such, must be preserved" (1999, p. 8). At the Austrian National Library the only copy of the old subject catalogue had been exposed (unprotected) to the public for many years, so that the CIPAC was considered a security copy (Dikovich & Wilhelm, 1997). At Dresden the last security filming of the author/title catalogue had been made in 1942 so that a new one was urgently needed (Golsch & Simmich, 1999). Several other statements of this kind were found in the CLQs and the literature. Therefore, two (sometimes overlapping) aspects regarding security and preservation can be noted:

- the creation of a digital copy (CIPAC) makes it possible to remove an endangered old catalogue;
- the process of CIPAC creation leads to the availability of a security copy of the catalogue (either on CD-ROM or on roll film).

3.2 Deciding about the kind of CIPAC

When a library decides to create a CIPAC, immediately the question arises what kind of CIPAC one plans to establish. What features will this catalogue have? What kind of browsing mechanism, what kind of image display, will online ordering be available? What software will be used, and should it be a commercial solution or a self-developed one? All these questions are not only interrelated but also depend on factors such as budget considerations, the kind of catalogue to be converted, and local aspects (e.g. the availability of a programmer in the library), to mention just the most significant ones.

When in 1992 the Princeton University Library was looking for a vendor that could supply a turnkey solution for its planned CIPAC, it was not so easy to find a suitable one (Henthorne, 1995). Even today, the CIPAC software scene is rather scattered; see also section 2.3.3
far too many software solutions for a rather limited market, many of them home-made and/or installed only on one or two sites. Nevertheless, today libraries do have a choice when looking for CIPAC software, even if the number of commercially offered products is limited. In many cases the preference of certain features will still directly imply what software is to be used. For example, previously a library that wanted a probabilistic search of OCR read full-text of the catalogue cards had no choice other than the Spider system; other software of this kind was developed only recently (Pietzsch, 2001b).

The following table shows the reasons for the selection of the CIPAC software as mentioned in the CLQ and/or the literature:

<table>
<thead>
<tr>
<th>Reason</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good experiences of other libraries with this software/this company</td>
<td>6</td>
</tr>
<tr>
<td>System has good features/capabilities</td>
<td>5</td>
</tr>
<tr>
<td>Software developed by other library available at no/low cost</td>
<td>5</td>
</tr>
<tr>
<td>Cost-effective system</td>
<td>5</td>
</tr>
<tr>
<td>Spider software would have been attractive but was too costly</td>
<td>4</td>
</tr>
<tr>
<td>Most cost-effective solution was to develop system in-house</td>
<td>4</td>
</tr>
<tr>
<td>Spider (OCR) not possible because many cards handwritten</td>
<td>2</td>
</tr>
<tr>
<td>System is easy to use, user-friendly</td>
<td>2</td>
</tr>
<tr>
<td>Same system is used nation-wide</td>
<td>2</td>
</tr>
<tr>
<td>Turnkey solution, not much work by library staff required</td>
<td>2</td>
</tr>
<tr>
<td>Developing it together with software house was best solution</td>
<td>2</td>
</tr>
<tr>
<td>Software can be implemented quickly</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3-3: Reasons for the selection of CIPAC software

This table leads to the suspicion that reasons such as low cost and availability at nominal cost\(^{11}\) were sometimes more important than the actual features for searching and navigating. In other cases it seems that the librarians concerned were impressed by a given CIPAC solution and decided they would like to have something like this in their libraries, too. If in the process of software selection any detailed project planning was undertaken this was rarely disclosed. An exception is the Saxony-Anhalt University and State Library at Halle whose project planning document (Schnelling, 1998) shows that the library had looked at some existing CIPACs and was able to describe exactly what software features were desired (p. 4):

- navigation via an index covering every 50\(^{th}\) card
- browsing back/forth card by card
- temporary selection of a different stepwith (e.g. 10 or 20)

\(^{11}\) For example, the fact that in Austria several institutions besides the National Library are using Kat-Zoom was certainly determined by the very low price at which they could obtain this software.
Chapter Three: Problems and issues in CIPAC implementation

- an appropriate module for adding, altering, deleting of images
- a feature for adding to the index
- a feature for the registration of all changes made
- an indication what catalogue is being searched
- the display of approx. four preceding and following index entries for navigation
- a feature for downloading images into a data file
- a printing option (to printer or to file)
- an online ordering feature (incl. printing out of borrower slips)

3.3 Technical aspects

3.3.1 Preparing the card catalogue

With reference to the preparatory work in the library, Rönsch (1998, p. 1566)\(^\text{12}\) in her account on the CIPACS at the Berlin Central and Regional Library succinctly states:

None. The reference catalogue was scanned as it was, straight from current use without any preparation, which means that sorting errors in the card catalogue are also reflected in the image catalogue. The division into three [catalogue sections] was kept.

Presumably, a very similar approach was used in other libraries, too. Even if the cataloguing rules would have allowed to merge consecutive catalogue divisions, this was normally avoided; the same is true for systematic revisions of the filing sequence. However, there are also cases where considerable preparatory work was undertaken before the catalogues were scanned. There are several tasks that need consideration and they can be categorized as follows:

- Improving / completing the existing leader cards for use as a partial index, e.g. when unevenly distributed (as in the case of Princeton; Henthorne, 1995) or when a subject CIPAC required a guide card structure of headings and subheadings (at Halle; Lutze, Schnelling & Worch, 1999);
- Creating indexes on the basis of drawer labels, leader cards, headings etc. (sometimes done by library staff);\(^\text{13}\)
- "Cleaning" of the drawers, e.g. removing glassine covers, re-typing cards that are badly damaged or illegible (Henthorne, 1995);
- Removing duplicate cards (works that have already been catalogued for the OPAC); e.g. at Luzerne 4–5 person-months were needed for the removal of 450,000 cards

\(^{12}\) Originally in German (author's translation).
\(^{13}\) See also section 3.3.5
Chapter Three: Problems and issues in CIPAC implementation

(Niederer, 1999); at Göttingen 810,000 of the 2.3 million cards were removed\(^{14}\) (Buschey, Halle & Harms, 2001);

- Checking the card catalogue for sorting errors (no evidence of realization).

### 3.3.2 Scanning and quality control

Scanning can be performed either by the *library itself* (i.e. its staff and/or additional helpers, e.g. students) or by *commercial firms* (scanning bureaux, software vendors). In the first case it will normally be done in the library, but in the second case it can be done either *on location* or *off-site* (e.g. in the premises of a scanning bureau).

When the Princeton University Library conducted its pioneering project, six million cards were scanned on the Library's premises by six purposely hired students who worked with three scanners in two shifts from 7:30 a.m. to 11:30 p.m. (Henthorne, 1995). The more recently created CIPACs were mostly scanned by commercial companies, especially in the case of the large libraries\(^ {15}\) – in many instances as part of a package that also included database creation and software. There is no clear pattern visible whether in-house or off-site scanning should be preferred; this obviously depends on factors such as space (is there enough space in the library where scanning can be done without disturbing the users?), time (how much longer will it take when large quantities of cards must be shipped to the vendor's premises?), security (is it the only copy of a valuable catalogue which cannot be given away?), and vendors' preferences (e.g. concerning quality control).\(^ {16}\)

Scanning is normally done with high-speed scanners at a speed of up to 60,000 cards per day, at a resolution between 200 and 400 dots per inch, most often at 300 dpi. For most CIPACs only the front of every catalogue card was scanned, sometimes even if there was also information on the back.\(^ {17}\) The resulting digital images are usually *bipolar*, i.e. black and white (rarely grey-scaled or in colour), in TIFF format, often of the *TIFF G4* (Group IV) standard.\(^ {18}\) According to Dikovich (1998), a 200 dpi black and

\(^{14}\) These cards were identified by means of optical recognition of a special ID number that marked them as printouts from electronic records.

\(^{15}\) An exception is the Moravian Library that purchased its own scanner and hired external operators for scanning their approx. 2.7 million cards (CLQ).

\(^{16}\) For example, of the eight catalogues that the Dresden University Library digitized, six were scanned in Berlin (the vendor's location) and two (the historical catalogues) on the Library's premises at Dresden (CLQ).

\(^{17}\) For example, the Princeton catalogue had information on the back of only 20 percent of the cards, so that it was decided to scan only one side (Henthorne, 1995). Only a few CIPACs feature double-sided card-images, e.g. those at Bologna and Florence (see CIPACs no. 42 and no. 44 in Appendix A1).

\(^{18}\) An explanation of these technicalities can be found in Lee (2001, chapter 3)
white card-image requires 6–7 KB of storage space. The images are normally supplied on CD-ROM (occasionally also on magnetic disk, DVD, roll film).

According to all available sources, quality control is a time-consuming task but crucial for the functioning of the CIPAC, regardless what navigation/retrieval software will be used.19 Today, quality control is usually part of the package offered by commercial vendors, but library staff are often involved, too. Mainly two aspects need to be checked – image quality and completeness (Schäuble, 1996). Image quality refers to the legibility which must be equivalent to the original card. Although at 300 dpi this can be achieved without problems, as a result of the not yet perfect automatic feeding of the scanners a certain proportion of the images will depict only parts of the cards or will be skewed. Completeness means both the correct sequence of the images and the existence of an image for every catalogue card (sometimes cards stick together so that no scan is taken of the second card). According to Köstler & Schäuble (1998) the proportion of defective card-images should be kept under 0.01 percent, especially if optical character recognition will be applied; Dugall (2001) mentions an error tolerance of less than 0.5 percent.

3.3.3 Image standards and web browsers

As mentioned above, scanners normally produce digital images in TIFF format which is the common standard for master images. However, the image formats suitable for web-browsers are GIF and JPEG, so that many digitization projects create GIF or JPEG files from their TIFF masters for the subsequent delivery via the web (Lee, 2001, p. 45-46). In the case of CIPACS there are basically two approaches for the transmission of card-images on the WWW (Braune-Egloff, 2000):

- Conversion of the TIFF images into GIF or JPEG format: This approach has been used in many CIPAC projects, but in the case of card-images the latter formats need more storage space than TIFFs; if the CIPAC system also requires images for a short view of results, cropped versions of these images need to be stored as well);20

- Alternatively, TIFF images can be transmitted and displayed in the browser by means of Java applets or plug-ins, i.e. software that downloads onto the browser and supports not only the display of the TIFF image but also its manipulation (setting the image size, zooming in/out, changing the resolution, the brightness and the contrast,

19 At Princeton, 15 students and 40 library staff were needed for quality checks of every fifth image, and many cards had to be re-scanned (Henthorne, 1995).

20 The Moravian Library found an interesting way to avoid storing cropped images by displaying, for a short view, only the top 160 pixels of each image as the cell background of an HTML-table (CLQ).
rotating the picture, inverting the colours, printing the card-image and downloading/saving it on the user's local workstation). This approach saves storage space and bandwidth, and also helps to avoid problems of synchronizing two image databases in case any changes are made. However, the applets and plug-ins require the use of a recent web-browser version and possibly a number of browser adjustments by the user (activation of Java and JavaScript support, accepting cookies, enabling of printing with Java applets, etc.)\(^{21}\) Some of these plug-ins may also come into conflict with other plug-ins installed on the PC or be incompatible with certain platforms (e.g. MacIntosh computers).

### 3.3.4 Optical character recognition

At about the beginning of the "CIPAC area", Dietze (1995, p. 245) reported from a visit to the Library of Washington:\(^{22}\)

> Experiments involving scanned catalogue cards have not been successful; the tests have been stopped because when applying OCR the correct identification of individual characters proved too difficult. However, the main problem was that the categorization of the scanned cards was too complicated.

For the purposes of CIPACs, the first problem had already been solved in 1995, because the Spider software was being developed at that time;\(^{23}\) the second problem was – again only for CIPAC purposes – solved more recently.\(^{24}\)

As shown in section 2.3.3, only a relatively small number of the present CIPACs are based on BerninaSpider or similar software. Obviously, for these libraries the use of OCR was beyond question. The Berlin Senate Library also found a way of converting its author catalogue by OCR in order to integrate the text, after considerable manual corrections, into its "normal" OPAC (Lux, 1997). However, both from the CLQ and the literature it becomes evident that quite a number of other libraries also experimented with OCR but soon gave up because of poor results. They had found that their catalogues contained a proportion of badly recognizable cards (handwritten, badly printed and partly damaged cards, cards with a variety of typefaces) that was too high for obtaining reasonable results via OCR. In several cases OCR was also described as too expensive – some libraries just did not have the financial resources for this additional step,

---

\(^{21}\) Most Chopin CIPACs that use the Java applet approach come with a technical online help of considerable length that might not be fully understood by every user.  
\(^{22}\) Originally in German (author's translation).  
\(^{23}\) See also section 2.2.4  
\(^{24}\) HeBIS-Retro makes use of an automatic categorization technique; see also Appendix A1, CIPAC no. 25.
whereas others used "OCR" as a synonym for the Spider software which was considered too costly. The Austrian National Library, however, was lucky enough to possess two uniformly typewritten old catalogues (authors and subjects), converted them first to CIPACS and after some time, by using OCR, into one single OPAC. Others mentioned that they had scanned their cards with a rather high resolution (300–400 dpi) in order to make any future OCR attempts easier (e.g. Buschey, Halle & Harms, 2001).

In a study preceding the development of the Spider software, the developer group found that for the catalogue of the Zurich Central Library the OCR process resulted in a word accuracy of 67 percent, which means that one in three words in the catalogue sample was incorrectly recognized (Mittendorf, Schäuble & Sheridan, 1995; Schäuble & Sheridan, 1996). The main difficulties for OCR were (a) the large variety of languages of the catalogue entries, with many accented characters, and (b) the large number of proper nouns and abbreviations; in both cases automatic dictionary lookup is not feasible. Pietzsch (2001a, 2001b) mentions as the main problems for OCR (a) heterogeneous font face and font size, (b) amendments made on the cards (handwritten, different typeface), (c) varying degrees of blackness (from card to card, but also on the same card), (d) wear and tear (stains, dirt, mechanical damage), (e) variety of languages. Whereas Schäuble and Pietzsch make use of retrieval software with a high tolerance of errors, Dugall (2001) highlights the importance of quality control and describes various approaches for automatic quality checks and error correction used in the HeBIS-Retro project. He also states that the OCR process is much more difficult than the preceeding scanning step (p. 118).

3.3.5 Manual/intellectual input

Both the KatZoom and the BerninaSpider CIPACS do not require manual or intellectual input for the creation of the respective CIPACS. When designing KatZoom, the Austrian National Library intended to keep things simple and decided to avoid the cost of index creation (Dikovich, 2000), an aspect which was also attractive for the other libraries that subsequently used that software (CLQ). The more sophisticated Spider system by definition does not require any manual work on the part of the library (except the preparation of the card catalogue for scanning).

The libraries using CIPACS with "virtual drawers" had to create indexes of the existing drawer headings and, in some cases, also of the headings of the original catalogue

---

25 See the case study in Appendix B4.
This was obviously not a very costly and/or time-consuming task and therefore also attractive for the institutions concerned. For example, the University of London Library originally considered a more sophisticated index but found that its creation would take too long and cost too much; when the Library later on was given the opportunity to use the LSE’s home-made CIPAC software, a much simpler drawer label index was established (CLQ). Table 3-4 shows a sample section of this index of 970 entries in total.

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Drawer</th>
<th>Cards</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>WADDEL</td>
<td>916</td>
<td>266</td>
</tr>
<tr>
<td>WADDES</td>
<td>WAH</td>
<td>917</td>
<td>588</td>
</tr>
<tr>
<td>WAI</td>
<td>WALKE</td>
<td>918</td>
<td>611</td>
</tr>
<tr>
<td>WALKER</td>
<td>WALLACE</td>
<td>919</td>
<td>607</td>
</tr>
<tr>
<td>WALLACH</td>
<td>WALP</td>
<td>920</td>
<td>517</td>
</tr>
<tr>
<td>WALR</td>
<td>WAM</td>
<td>921</td>
<td>600</td>
</tr>
<tr>
<td>WAN</td>
<td>WARD</td>
<td>922</td>
<td>560</td>
</tr>
<tr>
<td>WARDA</td>
<td>WARR</td>
<td>923</td>
<td>567</td>
</tr>
<tr>
<td>WARS</td>
<td>WASHINGTON [1] CARNEGIE</td>
<td>924</td>
<td>468</td>
</tr>
<tr>
<td>WASHINGTON [8] SMITHSONIAN</td>
<td>WASHINGTON. B</td>
<td>928</td>
<td>395</td>
</tr>
<tr>
<td>WASHINGTON. G</td>
<td>WATKIN</td>
<td>929</td>
<td>329</td>
</tr>
<tr>
<td>WATKINS</td>
<td>WATS</td>
<td>930</td>
<td>386</td>
</tr>
<tr>
<td>WATT</td>
<td>WEA</td>
<td>931</td>
<td>523</td>
</tr>
</tbody>
</table>

Table 3-4: A sample section of a drawer label index

For the CIPACs featuring partial indexes, longer and sometimes more sophisticated files had to be created. In the case of the Chopin systems, this task was often performed by the vendor, especially for the author/title catalogues where the headings of every \(n^{th}\) card (e.g. every 20\(^{th}\), 50\(^{th}\), 200\(^{th}\)) were used as entries; in other cases (e.g. Princeton) the libraries created the indexes themselves. The latter was also true for subject CIPACs for which some libraries (a) keyed in classification schemes (Rönsch, 1998), (b) created new subject indexes to the classification scheme (Lux, 1997; Rönsch, 1998), (c) checked and enriched the index entries produced by the vendor (Braune-Egloff, 2000).

---

26 For example, the CIPAC of the Czech National Library features indexes of the latter kind; see also Appendix A1, CIPAC no. 16.
27 An ASCII version of this index was discovered by the author on the web; when made an Excel file this index is only 102 KB in size.
or (d) created subject guide cards of headings and subheadings which the vendor could then use for index building (Lutze, Schnelling & Worch, 1999). The CIPAC of the Heidelberg University Library – featuring a full (not partial) index – is based on the headings of all 1.2 million catalogue cards which were keyed in by a commercial bureau; by means of appropriate software this index was then also permuted in order to facilitate easier browsing of the "Prussian" headings (Pietzsch, 1998b). The most lavish manual input project was undertaken for the Bavarian State Library's CIPAC.28

3.3.6 Servers, databases, system architecture

In the case of most CIPACs implemented by commercial vendors (Chopin, Bernina-Spider, and many individual solutions by various software houses) the server side of the system is just a "black box". Indeed, the libraries concerned need not worry about how their system is organized internally, all they usually need to do is provide / finance the hardware required for the server (usually a PC or workstation with sufficient memory and mass storage). The vendors themselves seem to prefer not to disclose the technological details of their systems. For example, the Chopin folder29 informs just briefly on the components being used: Microsoft technology for Internet connectivity, ACCESS and SQL-Server for the databases, scripts in ASP and Java for retrieval, and Java applets for visualization. No information on the architecture of the Spider system is available either.30 By contrast, the libraries that developed their CIPACs themselves had to deal with all technological details on the server side, and the two examples in Appendix B3 (Austrian National Library, Heidelberg University Library) may illustrate what goes on behind the scenes.

3.3.7 Administrative tools

In the CIPAC context an administrative tool is a software module that enables the library to make various kinds of changes in the card-image catalogue. Not all CIPACs are equipped with such a module;31 this applies not only to simple CIPAC applications32 but

[28] All drawer labels were keyed in by library staff; 21,000 guide card terms were merged into this index by the commercial firm that scanned the catalogue. The vendor was also commissioned to key in the text of all catalogue cards, partly categorized (author, title, year, and call number), partly as free text, with an accuracy rate of 99.95 and 99.5 percent, respectively (Fabian, 1997).


[30] Internal documents describing these systems' architecture in detail will probably exist, but these were not accessible to the author.

[31] See section 2.3.8

[32] E.g. the home-made system of the Graz University of Arts Library (see Appendix A1, CIPAC no. 1)
also to some of the more sophisticated CIPAC solutions. However, most CIPAC libraries that mentioned this issue in the CLQ and/or the literature seemed to be rather interested in such an administrative module. For example, the Halle project planning document contains three relevant requirements (Schnelling, 1998, p. 4):

- modifying, adding and deleting images at a later time must be supported by appropriate graphics tools;
- additions to the index must be possible (in case the Library wishes to index more images or even all images);
- all later modifications must be written to record files in order to make it possible that in case the database is rebuilt from the archive CDs the most up-to-date version can be restored again.

Of the wide range of possible applications of such an administrative module the following ones were mentioned most often:

- changing call numbers and/or locations, either by writing text onto the image (Chopin) or by replacing the card by a newly written one (KatZoom);
- putting "electronic stamps" on cards for which records have already been added to the "normal" OPAC; or (alternatively) deleting such cards from the CIPAC;
- replacing illegible or faulty cards by newly scanned or newly typed cards;
- changing the sorting position of images (in case of errors);
- correction of index entries (including characters not/wrongly recognized by OCR);
- making amendments to the index (e.g. by adding subject headings/subheadings).

3.3.8 Organizational aspects

It has already been mentioned that CIPAC creation can be done completely in-house, or by out-sourcing various parts or even the whole of the project. There are cases were even the CIPAC system as such is operated by a vendor or another institution.

The part most often done by a service agency is scanning (and OCR processing), even in those cases where the libraries undertook most of the projects themselves (e.g. at Heidelberg University). Many other steps of CIPAC projects have been performed by exter-
nal firms as well. In some cases (e.g. Spider CIPACs, some Chopin CIPACs, Berlin Senate Library) the libraries preferred to hire a sole supplier in order to have the whole package – consulting services, project management, quality control, co-ordination, software and systems compatibility, and guarantee of conversion quality – supplied by the same company. It seems that unless such a full package was purchased, Perez' recommendation to use an external library or IT consultant for the validation of project planning and procedures (1998, p. 64), was hardly followed in any of the cases.

3.4 CIPACs and the peculiarities of old catalogues

3.4.1 Physical form of old catalogues

Not all former library catalogues were typed on 7.5x12.5 cm or 3x5 inch cards. The older the catalogue the more likely it will be not only (partly) handwritten, but also

- on oddly shaped cards, slips or sheets (both in horizontal and vertical formats of different size);
- in the physical form of a sheaf catalogue (a batch of slips held together by some binding mechanism)\(^{38}\) or a book catalogue (bound volumes with several or many catalogue entries per page).

Some of these older catalogues may not to be scanned as easily and by high-speed batch scanning techniques as drawers of standard sized catalogue cards. Book catalogues can be processed with special book scanners; the use of both a higher resolution (e.g. 400 dpi) and grey-scaling seem to be advisable (Angelus, Eichhorn-Berndt & Schnelling, 2000, p. 430). The University of Vienna Library's old book catalogue was not scanned from the original but from a microfiche version created in the 1980s (Dikovich, 2000).\(^{39}\)

Concerning navigation, it makes no difference whether a CIPAC is based on a sheaf catalogue or a card catalogue. In the case of book catalogues, things are more difficult, not only because the individual sheets are usually much larger but mainly because one sheet contains several or many catalogue entries. Two solutions exist so far:

---

\(^{38}\) Pieces of wood connected by screws, belts made of woven material, etc.

\(^{39}\) This was possibly a cost-effective alternative but the image quality of this CIPAC must be described as not very good.
**KatZoom**

The program version adapted for the Viennese book catalogue makes use of an index [1] based on the labels of the original microfiches, and works with a division factor of 3. On the selection of an index entry the system displays – horizontally – four cropped images, i.e. the first pages of three divisions of the range selected, plus the last page (Figure 3-1); the user is required to "zoom" into one of the three sections and repeat the dividing procedure until the cropped images of four successive pages appear. Subsequently, the page(s) containing the search term can be displayed in full view.40

![Fig. 3-1: A KatZoom book catalogue search](image)

**Chopin**

A program version was developed for classified book catalogues such as the ones at Halle and Leipzig.41 For these catalogues, the classification schemes were converted to textual databases with links to the matching catalogue pages. The users are presented the main classes and sub-classes and click through these levels of the classifications' hierarchies. Alternatively, they may also perform a keyword search of the class names. When a sub-class is selected, the top segment of the first page belonging to that class is visualized (Figure 3-2); the rest of the page can be viewed by means of horizontal and

---

40 In GIF and/or TIFF format (an appropriate plug-in is required for the latter).

41 See Appendix A1, CIPACs no. 28 and no. 33.
vertical scrollbars. The page section numbers on the left side of the display are used for online book ordering (for the proper identification of the book the users are required to enter this number in addition to the call number).

3.4.2 Old cataloguing rules

Whereas in the case of "normal" OPACs the users need not be concerned too much about the underlying principles and rules of cataloguing, the majority of CIPACS are as one-dimensional as the card catalogues on which they are based. This means that, in spite of the online accessibility of CIPACS, their users must have some basic understanding of the headings and the filing order used in the respective catalogues. Many CIPACS offer online help to explain such rules, sometimes at considerable length. However, most old catalogues are based upon rules for cataloguing and filing different from those used in the more recent past, so that the users of a CIPAC divided in chronological sections may easily be confused by a variety of such helpful recommendations. CIPACS in the German-speaking countries are affected worst because in many libraries the notorious "Prussian Instructions" (PI) were used – a set of rules developed in the 19th century, based on grammatical rather than alphabetical principles and originally not made for end-users but for scholarly librarians. The users of the former card catalogues never
understood the PI, the present younger generation of librarians does not know them anymore and it can be assumed that hardly any CIPAC user will wish to understand them. It is certainly a major criticism that some CIPACs are now carrying these idiosyncratic rules far into the online age.\footnote{There are no corporate author entries; works of corporate bodies or those with more than three authors are filed under a grammatically determined title heading (e.g. the book "Hundert Jahre Technische Hochschule Darmstadt" is filed under the heading "Darmstadt Jahre Hundert"); letters such as "I" and "J" are interfiled (even in subject catalogues); the sorting of personal names is incomprehensible without a good knowledge of the rules; etc.}

Simple CIPACs such as those of the KatZoom type are more affected by such old rules than others because their only access points are the letters of the alphabet – then it is up to the user to understand the filing sequence. In the case of systems that work with drawer labels or partial indexes some "repair work" can be done when the system is set up. An example is given by Fabian & Haller (1998, p. 173-174) who describe the making of the drawer labels index for the CIPAC of the Bavarian State Library.\footnote{For this library's old catalogue the Old Munich Rules had been used (similar to the PI).}

- as the letters "I" and "J" were interfiled, drawer labels containing either were also keyed in with the other spelling, e.g. "JMM" (a cross reference for a journal title) was supplemented by "IMM" (in order to find "Imm, Emil"); when using the index the user may learn that these characters have the same sorting value;
- as many personal names were filed phonetically rather than alphabetically (e.g. Schmid, Schmied, Schmidt, Schmitt – all in one sequence), drawer labels containing such names were supplemented by additional index entries (e.g. "Schwarz, Ber..." by "Schwartz, Ber...").

3.4.3 Legibility of old cards

Legibility as a parameter of image quality has already been mentioned above. In this context, there is a second aspect of legibility that may affect some of the German and Austrian CIPACs, because in these countries on some of the handwritten catalogue cards the Kurrent script (Old German Script)\footnote{See also http://www.germanscript.com/history.htm [accessed 27/04/2002]} was used which many of today's librarians and library users will not be able to read. An example of such a card is given in Figure 3-3.
3.5 Presenting the CIPAC to the users

3.5.1 Naming the CIPAC

When the Princeton University Library created the first major CIPAC, this previously unknown kind of reference tool was presented to the users as the "Electronic Card Catalog", but later on its name was changed to "Supplementary Catalog" in order to emphasize the fact that more and more cards were added to the OPAC and some of the information in the CIPAC might be outdated.45

<table>
<thead>
<tr>
<th>Term/Phrase</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>digitized/digital/scanned/electronic/online (version of) card catalogue</td>
<td>16</td>
</tr>
<tr>
<td>(card/author/subject/library/old) catalogue until/before 19xx</td>
<td>15</td>
</tr>
<tr>
<td>image catalogue</td>
<td>10</td>
</tr>
<tr>
<td>scanned/digitized catalogue</td>
<td>5</td>
</tr>
<tr>
<td>web index, online card index</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3-5: Terms/phrases used for naming CIPACs on library web-pages

45 However, "Formerly known as the Electronic Card Catalog" was added as a subtitle, presumably because it was the more meaningful name.
Unlike in the case of "normal" OPACs no standardized terminology for CIPACs has been established yet. Therefore, it is not surprising that on library web-pages quite a variety of different terms is used for pointing/linking to CIPACs. These terms and phrases, as taken from the web-pages of all CIPAC libraries known so far, can be categorized as shown in Table 3-5. Only a few libraries created acronyms for their CIPACs, such as KatZoom (Vienna), DIKAT (Luzerne), DigiKat (Heidelberg), IPAC (Berlin), or Card-PAC (Dortmund). As these acronyms are meaningless without an explanation, they are usually also accompanied by one of the terms or phrases listed in Table 3-5.

As shown in the table, often the term card catalogue is used, together with some adjective that indicates that the digital form (and not the conventional one) is referred to. Another frequently used type of expression refers to the period of chronological coverage rather than to the electronic form (e.g. Author catalogue until 1994). In a number of cases, the rather vague term image catalogue is used, particularly in German-speaking countries (spelled as "Image-Katalog" or "Imagekatalog"), whereas some other libraries named their CIPACS just scanned or digitized catalogues (without mention of the cards), which is not exact either.

Most libraries do not bother to further explain these names, some of which must be rather confusing for inexperienced library users, by giving more information (except maybe in a separate CIPAC online help file). Only in a minority of cases, some sort of helpful mini-description is provided, e.g.

- "an electronic copy of the former card catalogue, comprising digitized facsimiles of the catalogue cards" (University Library, Freie Universität Berlin);
- "just a copy (image) of the conventional original catalogue cards which were scanned" (Leipzig University Library);
- "an online database of images that replicates catalog card indexes to selected library and archival collections"; "... contains a separate image for each catalog card" (Library of Virginia).

A few libraries also hint that their CIPACs are temporary catalogues for the time being, i.e. until all records will have been added to the OPAC. Practically all of them fail to ex-
plain why this conversion method was chosen (e.g. cost, speed), thus leaving their users in the dark about why CIPACs exist alongside OPACs anyhow.

3.5.2 Integration of CIPACs into OPACs and web-pages

Most CIPACs are not integrated with the libraries' "normal" OPACs at all; normally, they only share with them a web-page with links to the various online catalogues (and sometimes databases) provided by the library. Most CIPACs do not even resemble their OPAC counterparts in terms of page design and layout.

One of the few exceptions is the Chopin CIPAC at Freie Universität Berlin that was designed to match the "corporate identity" of the University Library as also expressed in the other library web-pages, even if the CIPAC and the OPAC are not integrated yet (Braune-Egloff, 2000). Most other CIPACs were not brought into line with the graphic design of their libraries' web-sites.

Only in a few cases the term "integration" does apply:

- **Zurich Central Library:** From the outset this library planned to offer simultaneous searching of the CIPAC images and the OPAC of the Swiss Union Catalogue (Köstler & Schäuble, 1998). The Zurich BerninaSpider CIPAC is also linked with the Library's automated circulation system.50

- **Bavarian State Library:** The CIPAC is a module of the general OPAC; when the user clicks on a "drawer" icon the combined drawer labels and leader card index appears and CIPAC browsing can start. As the text of the cards was also typed in the reader may, alternatively, search the OPAC and, in case the full display shows a "view card" button, request the image to be shown. The CIPAC part of the system is not connected to the ordering/circulation module, but as all images also have a text record in the OPAC this is not necessary.

A few other libraries also plan some sort of integration, either with the OPAC's circulation module (e.g. Luzerne; Niederer, 1999) or with its user interface (e.g. Freie Universität Berlin, Berlin Central and Regional Library), but by and large one cannot diagnose a trend into this direction.

49 See Appendix A1, CIPAC no. 20 (Fig. A-5).
50 See section 2.2.4 and the figures there.
Chapter Three: Problems and issues in CIPAC implementation

3.5.3 What do CIPAC libraries know about CIPAC users?

Most CIPAC libraries who returned the CLQ reported that their CIPACS were well received by the users (the rest said that their card-image catalogues had only just been released so that no feedback could be collected yet). However, practically all of these statements rely on subjective impressions, on the reactions of individual users, some email messages received from users, and other kinds of selective observation. So far, no attempts have been made to record user reactions to CIPACs systematically (e.g. by conducting a user survey).51

Some libraries reported on various kinds of problems that individual CIPAC users had, e.g. with (old) browsers, plug-ins for image display, slow system and/or network performance, legibility of cards, difficulties with navigation or even with working on a computer in general,52 whereas others mentioned that special instructional sessions were offered when their CIPACS were released. Only one library responded to the author's request (made in the CLQ) for relevant data and/or material; this was the Austrian National Library which provided a portfolio of collected complaint book entries, press cuttings (mostly readers' letters), email feedback etc. From this collection it becomes visible that some users were not happy with the KatZoom type of navigation (too cumbersome, no options for text searching) while others welcomed the accessibility of the old catalogues without the former restrictions (location, time). The majority, however, just lamented – often quite emotionally – about the removal of the card catalogues and made pleas for bringing them back to the Library's reference area.

Even if, generally speaking, CIPAC libraries do not know much about the users of their card-image catalogues, they seem to monitor whether their CIPACS are used at all, because in a number of cases figures were reported on the frequency of use (based on web-server statistics and similar counting mechanisms). For example, often 1,000 users or more search the CIPAC of the Czech National Library per day (CLQ), and 500 the one of the Heidelberg University Library (Pietzsch, 2001b); the statistical data that were used for the diagram in Figure 3-4 were provided by the Austrian National Library.53

---

51 The only library that mentioned a user survey was the Berlin Senate Library; however, in this survey the users rated only "facilities for searching" and "web-pages/OPAC" (CLQ).
52 Austrian National Library/KatZoom: In some short views (cropped images) the headings of the cards are not visible (CLQ); Berlin Central and Regional Library/Chopin: Only every 200th card was indexed; some users believe that an author is not in the catalogue if his/her name does not appear as an index entry (CLQ); Luzerne Central and University Library/BerninaSpider: Professional users (librarians) believe it is faster to search the card catalogue (CLQ) and some readers found the subject catalogue difficult to search (Niederer, 1999).
53 For the purpose of its web statistics, the ONB defined 10 calls of the underlying program scripts as one CIPAC query (Dikovich, 2002).
This diagram shows for both CIPACs (author/title catalogue, subject catalogue) a pattern that reflects the typical ups and downs in an academic year, but, more interestingly, it indicates that the subject catalogue is used quite consistently at about half the frequency of the author/title catalogue.

![CIPAC use at the Austrian National Library in 2001](image)

**Fig. 3-4: CIPAC use at the Austrian National Library in 2001**

### 3.6 CIPACS: Interim or permanent solutions?

Finally, the question arises what future CIPACS will have. Are they only a transient phenomenon or will they last for a longer period of time? In order to find out the view of the CIPAC libraries the question was asked in the CLQ whether they considered their CIPACS as

- a short-term solution (or even a makeshift solution)
- a provisional solution / intermediate stage in a long-term conversion project
- a medium or long-term solution
- a permanent solution

Also, the project literature was scanned for judgements concerning this issue; the results are shown in Figure 3-5.

The short-term category applies to only two libraries that expect to complete the full conversion of their main catalogues as early as in 2003 (BLPES and Austrian National Library). Of the others, the largest group considered their CIPACS as interim solutions for the time their other conversion activities (in many cases already ongoing) will take. However, some of these libraries mentioned that they were not sure how long this will

---

54 Including the Library's vacation period in September.
take and if sufficient funds will be available, so that their CIPACS might become *medium* or *long-term* solutions. As most of those who chose the latter category had a similarly skeptical view one could actually merge these two into one group of libraries who expect that their CIPACS will stay for a while even if they should become obsolete at some time in the future. Some of the larger libraries with several catalogues in CIPAC format mentioned priorities concerning conversion speed so that some of their CIPACS would probably disappear sooner and others later (or never). The relatively small group of libraries that considered their CIPACS as *permanent* were rather confident that this was the best or most realistic solution for their catalogues.

![Fig. 3-5: Estimation of the future of CIPACS](image)

In reality, so far only one library has withdrawn card-image catalogues: The first two CIPACS of the Austrian National Library "lived" their short lives only from 1997 to 2000; then they were converted again and merged into one OPAC.\(^{55}\) Although they still exist somewhere behind the scenes, the public can no longer access them on the WWW. The next CIPAC to be closed might be the one at Princeton; the intention to do so was already announced in 2001 (because of the completion of the Library's conversion project) but has not been carried out so far.

\(^{55}\) See the case study in Appendix B4.