The LLMC-Digital Newsletter
   Issue No. 12: February 2005

Contents:

Good news on the image quality front

LLMC has just had a very satisfying month. In mid-January we installed two new scanners. One (SMA) will be used for scanning bound books and has an oversize capacity. The other (Staude) is a hi-speed scanner de-signed for scanning loose pages, primarily from disbound books. Neither of these systems was being sold in the U.S. when we made our initial equipment-purchase decisions. Both systems work fine for the purpose for which they were intended. But that's the least of it. As an added bonus, both systems also produce observably better images than is possible from our earlier generation of equipment. It goes without saying that the new images are better than anything we can derive from our fiche.

The reason appears to be that the engineering of the new machines has been re-thought. Our older machines look and work much like our old step-and-repeat film cameras did — lights high and off to the sides, capture mechanism (eye) up above the target materials by three or four feet. Maybe that's why we felt so comfortable with them initially. They looked reassuringly like our old cameras. With the new machines both the light source and the eye are within an inch of so of the target material. Eliminating the fuzziness caused by the intervening air makes all the difference. The new scanners have established a new LLMC standard for image quality. (See endnote 1)

In reaction to this auspicious development, we will be trading in our first generation scanners (Zeutschels) for other needed equipment. (See endnote 2) Since they are only a year old, we expect to be able to get back 60 to 70% of what we paid for them

Good news on the image costs front

Better image quality would have been enough of a New Year's present. But our cup runneth over. In addition to the higher quality, we have also realized a major improvement in our costs per image (CPI). Costs in this area are determined by two factors, thruput and initial-capture quality.

— Thruput is an obvious factor. The more pages an operator can put through in an hour, the less labor cost per image. The average thruput for both the Zeutschels and the new SMA is roughly 400 images per hour (IPH). So in that regard those two are roughly equal. The hi-speed Staude, however, hits breathtaking speeds. Rates vary depending on the physical size of the pages (bigger pages take longer to pixelize). But the average is settling in at about 4,400 IPH, eleven times the thruput of the step-and-repeats. (See endnote 3)

— Initial-capture quality determines how much post-processing work will be needed to clean up and enhance the image to acceptable standards. With the Zeutschels our one-year experience has been that post-processing takes as much labor time

Page 2 in the print edition starts here

(at a somewhat higher skill level) as the initial data-capture phase. Now, with both the SMA and the Staude producing significantly cleaner images at the initial capture phase, post-processing time has fallen by roughly two-thirds.

It will be many months before we have accumulated sufficient data to provide refined figures; but the trend set in the early returns is unmistakable. In summary, our preliminary figures for cost-per-image (CPI) — limited to initial data capture and post-processing, and ignoring niceties like depreciation on equipment, overhead, etc. — are as follows:

— Zeutschel=roughly $0.07 CPI

— SMA=roughly $0.04 CPI

— Staude=roughly $0.01 CPI

These costs compare to the $0.035 CPI for digitizing from our fiche. It will take a year of experience to develop more refined data, but there's no doubt

about it. We have turned an important corner. We have brought the costs for scanning bound books to roughly the same price line as it costs to digitize from our fiche. And the costs for digitizing from disbound materials have gone through the floor.

Our practical response to these developments will be threefold. One, we will attempt to limit our digitizing from our fiche to those cases where we have little or no hope of obtaining original hardcopy for scanning purposes. Two, all future purchases of scanners aimed at bound books will be in the new technology pioneered by the SMA. Finally, so as to maximize the return on our new hi-speed Staude, we will make a major effort to obtain as high a proportion of disbindable materials for scanning as possible

LLMC is looking for discard books

As has been mentioned before in these pages, we expect that the digital era will see a much greater amount of hardcopy de-accessioning than occurred during the fiche period. We hope to take advantage of that phenomenon by persuading some libraries to make their discard decisions in synchronization with our production processes. In short, if you are going to discard anyway, think of our needs as part of the equation.

We are delighted to report a major development along these lines. Wayne State University Law Library will be using the opportunity presented by *LLMC-Digital* to recover the shelf space devoted to those of their state court reports which are, either non-copyright, or printed before 1923. The first shipment of books is already in the mail. We expect to process all of Wayne State's books (an estimated 8,000 volumes) in the next twelve months.

Those libraries which may be inclined to join in this effort may be reassured by several factors. One, it will be the digital images of their own books, if accepted, which will be on the screens of their patrons. Two, LLMC is committed to retaining indefinitely the disbound hardcopy for any materials fed through this process as part of the LIPA program discussed below. (See endnote 4) Finally, we are working on "bookplate" design so that in the future every book scanned for LLMC will carry an explicit recognition of the donor library.

LLMC cooperates with the LIPA group

As mentioned in the last issue of this newsletter, (See endnote 5) a number of librarians planned a brainstorming session in conjunction with the AALS meeting in San Francisco in January. Their concern was post-digital preservation of hardcopy, particularly in light of the anticipated flood of discarded hardcopy some predict as a natural result of the migration of so many of our titles to digital. LLMC is also concerned that some minimum number of paper copies of all U.S. primary material be

Page 3 in the print edition starts here

saved. Our concern stems from a general instinct in favor of preservation, but also includes some selfish factors. One, we are convinced that the technology of the future will be vastly superior to what we have today. If that becomes true, then it is probable that someday our successors will want to rescan much or all of our material. Second, we are trying to secure donations of disbindable material. We feel that many librarians will not be comfortable donating their materials unless they are assured that overall preservation for those titles has been assured.

For this reason LLMC was an enthusiastic participant in the LIPA-sponsored brainstorming session, and we left the meeting with an assignment. For the LIPA preservation effort to succeed, it will need, among other things, a recordkeeping system which enables tracking materials down to at least the page level. LLMC already does that for every title which it mounts on *LLMC-Digital.* Over the next four months it will be working with its database to see if that might be modified at low cost to serve LIPA's needs also. The LIPA hardcopy preservation group will be meeting again in San Antonio. LLMC plans to have a prototype database model ready for the group's review and reaction at that time.

Significant cataloging development

Some of the alert catalogers among our users have noticed that strange things have been happening to some of the titles on *LLMC-Digital,* specifically in those cases where the run incorporates title changes. Richard Ameung, our man of many hats at St. Louis University, has been sleuthing the case and has developed an elegant solution, as follows in his words:

"Over the initial period of bringing up *LLMC-Digital,* we've encountered and resolved several thorny display-related cataloging issues. However, one of the areas that is still causing user dismay is the retrieval of an apparent mismatch between the title requested on a record loaded from OCLC into the local catalog and the title retrieved from *LLMC-Digital* when the user clicks on the URL link. For the last few months, I've been engaged in a bit of detective work on this topic. Here are the conclusions to which I've been drawn and a solution that will solve the problem.

First, as has been stated previously, the software maintained by our partner, the University of Michigan, was originally created to handle collections that are largely monographic in nature, while LLMC's backfile includes many serially-oriented titles. Second, LLMC-Fiche traditionally assigned only one control number to an entire run of a serial title, regardless of the number of title changes that occurred. From a vendor or purchaser's point of view, that made sense. People wanted all of the *Michigan Reports.* They didn't care if the first four volumes were really *'Reports of the cases ....'* One good carryover from the fiche days is that, since the digital URLs for a given title are based on this LLMC control number, all the URLs for the entire run will be identical and the run stays together.

To date Saint Louis University Law Library has been cataloging this type of

title using the successive title approach to handle title changes. The problem occurs when successive records for parts of a run reach our friends at Michigan. When Michigan receives these records from OCLC, it "de-duplicates" them based on LLMC control numbers (MARC 037). However, since Michigan can retain only one record, it may not necessarily be the same one from which the local user launched his search, although it will have the same URL. An additional anomaly occurs when all volumes in the title are listed under this one remaining bibliographic record on the *LLMC-Digital* site, regardless of the fact that the last remaining "de-duped" record may only cover a portion of the run. As the French say; 'Que faire?'

Our first attempt at a solution aimed to determine if the software that Michigan has been using could be re-written to accommodate this rather complex situation, perhaps by retaining all records and de-duplicating based on an OCLC control number (MARC 001) rather than the LLMC control number. It transpires that such a solution would entail a rather large investment in both human and financial resources. In addition, since writing new programs would take some time, the number of titles which would have to be redone

Page 4 in the print edition starts here

at some future stage would continue to grow, putting a premium on an early solution.

A second solution that we explored, and in some cases implemented, was to provide each changed title with its own unique LLMC control number. Sometimes this made perfect sense, when the collective LLMC fiche control number represented a grouping of monographs that should, in fact, be searched one by one (e.g., the Selective Service monograph series). However, when it came to true serials, this approach caused two unfortunate complications. One, it created an ongoing disconnect for the administration of LLMC's assets. LLMC now had to track multiple control numbers for essentially the same material. Institutions wishing to purchase the fiche would use the old control number (e.g., 99-999), while the same title on the website would be tracked by a different set of control numbers (e.g., 99-999a, 99-999b, and 99-999c). Second, we discovered that, while using unique control numbers for each title change resolved the user's dilemma of being referred from one title on the local record to an entirely different (but related) title at the LLMC Digital site, the very fact that they are indeed different control numbers meant that users lost the ability to move through the entire run of the title at the site. They could only browse through those specific volumes associated with that specific title. (See endnote 6)

We have, therefore, devised a new solution that, while perhaps causing purists to cringe, will provide retrievability both locally and on the site for these problem titles. We will continue to use a single LLMC control number for an entire title, continuing the practice of the fiche era. The entire run will be cataloged on a single bibliographic record. The title for this record will be taken from a 'composite title page' created by the cataloging agent (i.e., St. Louis Univ. Law Lib.). This Composite Title page will be digitized and mounted

as the first image for that title on the *LLMC-Digital* site. It will contain all of the information concerning the title history from Volume One on. The bibliographic record will reflect this history by using searchable fields for both earlier titles and related issuing bodies. I have negotiated this approach with the OCLC quality control gurus. In light of the situation as explained, they agree that this is the only reasonable approach open to us.

A word of caution is in order for those who already may have brought OCLC records for *LLMC-Digital* into their local systems. We will be going back and collapsing multiple titles down to one record. When this occurs, OCLC control numbers for the bibliographic records no longer required will be moved into the MARC 019 field of the single record retained. The *LLMC-Digital* website will be updated to reflect the single record retained for the run of the title. We have also set up a mechanism whereby St. Louis Univ. Law Lib. will flags titles for LLMC where future volumes may be expected. In that way, potential future title changes can be accounted for and records updated as needed.

As I have said to the variety of stakeholders involved in these negotiations, (See endnote 7) I realize that this approach may not be ideal. However, it will certainly make all the data retrievable for the user. At this point, that should be our primary concern."

Analog/digital strategy for preservation

Many LLMC-Digital patrons will have heard of the recent statement by the Association of Research Libraries (ARL) entitled "Recognizing Digitization as a Preservation Reformatting Method." This statement was issued in June 2004. It definitely contradicts a policy which has been adopted by LLMC, namely that LLMC will rely for preservation purposes on a dual-medium strategy, the so-called hybrid approach. To summarize, we digitally scan our data, but also rely upon a technology

Page 5 in the print edition starts here

now in place to "write" the digital data to archival quality Silver Halide film. This is not the place to repeat all of our arguments. Suffice it to say that we don't find the ARL statement persuasive. For those who would like to explore this in more detail, the most recent issue of *Microform & Imaging Review* (See endnote 8) carries a symposium on the subject. The ARL statement is reprinted in full, followed by some trenchant observations by leading authorities. We feel that our viewpoint is more than adequately represented.

Of course, merely writing the data to Silver Halide film is not sufficient. That film must also be processed to archival standards, and the process verified through competent testing. For the peace of mind of all of our subscribers, we can state for the record that the *LLMC-Digital* preservation fiche is routinely tested by a recognized authority to ensure proper quality. (See endnote 9)

Trouble with the TVP search

Alert users began to notify us back in December that the TVP search was doing curious things with certain titles. The problem centers in the "Part" feature of the TVP search, and can be illustrated by what happens to the title *U.S. Statutes at Large.* As most of us know, that title is issued in large volumes, which are further subdivided into many parts. But the pagination is continuous, so that page 1,305, of volume 106, for example, may be in, say, Part 5. However, most citations do not include the part numbers. With the books they are not necessary, since the pagination for each part is given on the spine. So also with the fiche; the pagination is given in the header. However, not knowing the part number hurts on *LLMC-Digital.* If one puts in only the volume and page numbers (e.g. Vol. 106, page 1,305) the system will report back that page 1,305 doesn't exist, even though it does. The reason for this is that the "Part" feature defaults to "1" unless some other part number is entered. So the system is looking in Part 1 for page 1,305, and, of course, can't find it.

Unfortunately, we have looked into this and find that there is no quick and easy solution. Michigan has notified us that remedying the situation will take a significant programming effort. We, for our part, have let Michigan know that this is a major priority for us, outranking any merely cosmetic improvements on the site. We will continue to monitor progress on this front and will report regularly in subsequent issues of the Newsletter until the problem is resolved.

In the meantime, you should be aware that this problem exists. If you are using the TVP to look for a specific page, and are not getting a hit when you think the whole title should be up on *LLMC-Digital,* try using the option "View all volumes for this title." This will allow you to see whether you have a "parts problem." If you do, for now you will have to do a bit of sampling among the parts to see where your desired page is. If that sounds like a bit of a kludge, [(See endnote 10)](#) well, we have to admit it is, and we ask for your patience.

Page 6 in the print edition starts here

LLMC-Digital Interface Task Force

We are not yet ready to give a full report on the implementation of the Interface Task Force's recommendations. [(See endnote 11)](#) So this will be in the nature of an interim report.

Many of the Task Force's recommendations involve areas which are completely under the control of LLMC in Kaneohe. Most of those changes already have been implemented. For those who haven't visited the site recently, we invite an update inspection and think that you will be pleasantly surprised.

We are working with the University of Michigan on three remaining issues raised by the Task Force:

— Searching a specific range of volumes, rather than a single volume or the whole title. This capacity already exists on the site, but it is somewhat unclear how to one implement it. We are working with Michigan to make this

more intuitive.

— Some of the normal navigation aids ("e.g., the "Back" button) are missing from the LLMC side of the interface. This is a result of the way in which the Ann Arbor and Kaneohe sites are connected. We are working with Michigan to devise a different connection which will restore some of the navigation tools the Task Force found wanting.

— The Task Force asked that a link to the free Adobe Acrobat reader be provided for users. This is not a difficult matter and that link should be mounted soon.

Persistent URLs Federated Searching, etc.

The technophiles among our members have been asking that *LLMC-Digital* be made compliant with several new technologies related to improved access. People have queried us as to whether our system is compliant for "persistent URL's" (PURLs), "Open URLs," "Federated Searching," and even something called Z39.50. For most of us these arcane terms border on the giddy. But the reality is that our colleagues are exploring exciting new areas, which may eventually be our future. In the abstract the LLMC Board is all in favor of *LLMC-Digital* becoming compliant in any area that promises to improve the site's access capabilities. In reality, every advance along these lines requires expensive programming time. So we will work with our partners at the University of Michigan and chip away at these things. The encouraging news for this month is that Michigan has just notified that they have received funding for and are recruiting a programmer who will be working on, among other things, PURLs, Federated Searching, and Open URLs. They hope to have the new person on board by mid-April.

Search for Content Manager

LLMC itself is recruiting for a middle-level employee to take over many bibliographic duties now performed by the Executive Director. The search is being conducted from the Univ. of Michigan Law Library by the Personnel Committee of the LLMC Board of Directors. Short ads have been put up on the AALL and ALA job hotlines. (See endnote 12) These are supplemented with links to LLMC's main corporate web site for additional information. If you know of anybody who might be interested in an exciting job in a fun place, please do them and us a favor by calling the ads to their attention. Thanks much!

Annual Billing Cycle

The regular billing cycle for *LLMC-Digital* runs from March 1 of one year to the end of February of the next. This cycle was devised to meet multiple requirements from several of our subscribers related to their differing fiscal and budget years. Invoices for the 2005/2006 cycle will go out on March 1. Payment will be due within the next twelve months; although, if it works for your institution, earlier payments are always appreciated. If you have any questions regarding your billing, please check them out with Debbie Bagwell,

our Business Manager, toll free at 800-235-4446.

Endnotes:

1.) While we had high hopes for the image quality, the main reason we bought the SMA was ergonomic. Our operators were complaining from eye strain due to the bright side lights. The SMA's almost hidden light source has cured that problem.

2.) Being conscious that there is a common law tort termed "slander of product," we hasten to add that there is nothing wrong with the Zeutschels. There will still be a market for them, since they will remain useful for other types of record-retention work done by imaging service bureaus. However, we feel that our work, where the product will be seriously read by many users, should be done to the best image standard currently available.

3.) This compares to the 1,200 IPH achieved by the robotic Kirtas machines, which Google will be using for its big projects. However, remember that the Kirtas is for bound books, while the Staude is for disbound books.

4.) Onewleterur thinking has changed on this. In the past we anticipated that any books disbound for scanning purposes would subsequently be discarded. However, we now believe that this material in effect sets a standard. The images derived will be observable nationwide, and thus will be the basis of comparison for those who may offer "better copies." In addition, the material will have been vetted for completeness and suitability for scanning. So, should rescanning be required, it would be a natural target. As a result, LLMC is now negotiating with one large interuniversity consortium to see if its processed hardcopy, with appropriate packaging and record-keeping, might merit retention in their dark archives.

5.) No. 11, Dec. 20, 2004, p. 5, footnote 7

6.) To provide an example: if 99-999a ends with v. 5, no. 2 and 99-999b begins with v. 5, no. 3, a user who clicks on the URL in the local catalog for 99-999a would not be able to move from v. 5, no. 2 to v. 5, no. 3 on the *LLMC-Digital* site. The user would need to re-run the search either at the *LLMC-Digital* site or in the local catalog. Neither approach seems helpful or, for that matter, likely.

7.) After Amelung explained the full background of the Composite Title Page solution to the LLMC Board of Directors at their recent meeting in San Francisco, the Board voted enthusiastically for the plan's adoption.

8.) Fall 2004, Vol. 33, No. 4, pp. 171–206

9.) One of the essential requirements for creating archival-quality microfilm or microfiche is that the processed film be tested regularly for residual chemicals, particularly Thiosulfate, which may be left on the film. This is called "Methelene blue testing." For the record, all digital-origin microfiche created by LLMC for deposit in our archive at the Harvard Depository are being tested regularly Tests are run for each batch of chemicals used (i.e., roughly once a week), with sample processed fiche being mailed to a respected lab located in Minnesota. After each test we are e-mailed a certificate that no excessive residual chemicals were found in those samples. Fortunately, our new DigiFiche machine does such a good job of cleaning the fiche after processing (love that German engineering!) that the lab queried us as to whether we were actually using the prescribed chemicals. Forget staying within the prescribed minimum range. They couldn't find any trace at all!

Also fortunately, the tests are not that expensive (roughly $12.50 apiece, not counting our labor costs for gathering samples, mailing expense, etc.). The real challenge is making sure that the system is maintained over time, a course to which we are solidly committed.

10.) We are indebted to Jules Winterton, Dir. at the IALS in London, for this terminology, which he assures us qualified technophiles use to indicate "a botched or makeshift device or program which is unreliable or inadequate in function."

11.) First reported on in the last issue of the Newsletter: No. 11, Dec. 20, 2004, pp. 5–6

12.) Those interested can view the short ads on these web sites: 1.) for AALL, go to www.aallnet.org/ and click on "Job Hotline." Then look for the posting for 1.27; 2.) for ALA, go to www.ala.org/ ala/education/empopps/careerleadsb/hotjobsonline/Then scroll to the latest postings. Those interested in the fuller criteria and added application information should check out this link to our regular site: www.llmc.com/content_manager.htm.

End of Newsletter No. 12