

# Calculating bibliographic Duplicates

Kalkulierbare Dubletten ??

Les doubles bibliographiques calculables

Harald Jele

---

In this project it is shown that under certain conditions the checking for duplicates while loading bibliographic data into an online catalog can be replaced by a statistical estimation. Due to the selected significance level (of 90%) and tolerance bounds (10%) the minimum of the needed sample size is calculated, chosen from the population at random and verified by hand.

The result of this examination is regarded in further consequence as a valid result for the total quantity.

Im Rahmen dieses abgeschlossenen Projekts konnte gezeigt werden, dass unter bestimmten Bedingungen beim Laden von Titeldaten in Online-Kataloge die üblichen Methoden der Dublettenerkennung durch statistische Abschätzungen ersetzt werden können.

Dabei wird aufgrund von Sicherheits- und Genauigkeitsannahmen die Größe einer Stichprobe errechnet, diese aus der Gesamtmenge nach einem Zufallsverfahren gewählt und händisch (intellektuell) geprüft. Das Ergebnis dieser Prüfung wird in weiterer Folge als ein gültiges für die Gesamtmenge angesehen.

Dans le cadre de ce projet qui a été réalisé, on a pu constater qu'à certaines conditions lors du chargement des dates descriptives bibliographiques dans le catalogue en-ligne, les méthodes usitées permettant d'identifier les descriptions doubles pourraient être remplacées par un système d'évaluation statistique.

Lors de ce procédé on calcule la taille d'un échantillon pris au hasard en tenant compte du niveau de signification et des limites de tolérance, cet échantillon est choisi parmi l'ensemble des données par une méthode au hasard et contrôlé à la main (intellectuellement). Le résultat de cet examen sera valable pour la totalité des dates par la suite.

---

## 1 Introduction and basic conditions

A few years ago we found ourselves in the well-known situation of having a colleague with outstanding qualifications in a particular area which would be lost to us with his forthcoming retirement.

Before going into retirement, this colleague could be integrated into an extended field of activity.

Since the addressed qualifications – as the description of the project will show – concerned mainly talents<sup>1</sup>, which also were not preservable by the efficient mechanism of knowledge management, the idea was born to use these no longer everyday abilities for a part of his

---

<sup>1</sup> like the fluent reading of the German Kurrentschrift or natural handling of title data in old Greek and Latin language

time remaining in the institution.

A possibility of using these qualifications for the further course of the project was the integration of the coworker into the revision of a hand written inventory register of the library. In more concrete terms it concerned the inventory which provides the call numbers 1-26177.

For the type of the revision it was agreed upon that the results of the project are among other things to represent the production of an electronic form of the raw material; another form (e.g. a purely typewritten) was excluded.<sup>2</sup>

---

<sup>2</sup> in this sense no conventional catalog cards should be made of the entries of the existing catalog in book form. Such a proceeding would be probably not necessarily up-to-date or besides appropriate

A sub-goal of this project lay thus in the retrospective cataloging of specific holdings into the online catalog.

The printed material concerned (registered in the inventory) are primarily German translations of Greek and Latin classical authors from the 18th and 19th century as well as those works in German language which probably rank among the most published from this time. The latest editions among them date anyhow before the year 1925.

However, also a number of „ifs and buts“ were to be considered, which altogether prevented that the project could be completely accomplished or concluded mainly due to a lack of time:

- screen handling, which should be possibly contributed by this colleague, could not be taken into account in his task, because a specific visual impairment and the exemption from work on the computer, resulting from it, made it impossible
- supporting methods of modern cataloging<sup>3</sup>, which are typically carried out by online systems were accordingly not applicable and could be used only outside of his field of activity (but later on in a supplementary way)
- it had to be guaranteed that the training period for gathering bibliographic information with consideration of the set of the cataloging rules „RAK-WB“ on the basis of MAB2 data was as short as possible – otherwise the work performed did not justify under any circumstances the range of the result

## 2 Approach to a solution

The given basic conditions made the development and rather the employment of quite unorthodox methods necessary.

Finally these resulted in the decision that our colleague<sup>4</sup> collects all necessary information „in the most possibly familiar way“ on paper and hands this over to a further person, whose knowledge about the further processing – in the ideal case – could be very small.<sup>5</sup>

<sup>3</sup> this means above all a „formal verification“, the use of electronic „templates“, the representation of title hierarchies supported by automation as a result of the use of the German cataloging rules RAK-WB and according to bibliographic title information for volumes and series with hierarchical storage the linking of titles etc.

<sup>4</sup> who is seen here in his role exclusive as an expert for rule-conformal information retrieval and – under the premises specified in the basic conditions – whom we must offer exclusively already familiar functions

<sup>5</sup> in our case a briefly trained person was employed, whose training period lay within the range of a few hours

The collection (bibliographic and relevant for the holdings information) of the data happened actually in a very conventional way; nevertheless quite a lot of attention was paid to the fact that this is accomplished at least in a very specific and structured form, which consequently could be interpreted by a computer as error free as possible.

Subsequently, the sheets produced in this way were converted by further processing with methods of the OCR (Optical Character Recognition) into computer-readable text files.

These again were verified by appropriate programs and prepared to structures of suitable data records, so that they could be loaded online into the central Austrian catalog and afterwards – and in the workflow as noly-break as possible – by replikation into the local online catalog.

While loading these data records bibliographic duplicates were not checked deliberately with a program or in an intellectual way in order to minimize the effort.

In the preliminary stages of the project as well as under operating conditions the estimation of the amount of expected duplicates was verified with statistical samples.

The setting of tasks which should be handled was – under the basic conditions indicated above – finally to show that with a relatively simple statistic procedure useful statements (in the meaning of mathematical proofs) about the estimated error (which can be expected) can be done, if it is aimed not to exceed a certain number of bibliographic duplicates from a given quantity of processed data records.

In *fig.1* the complete workflow is schematically represented.

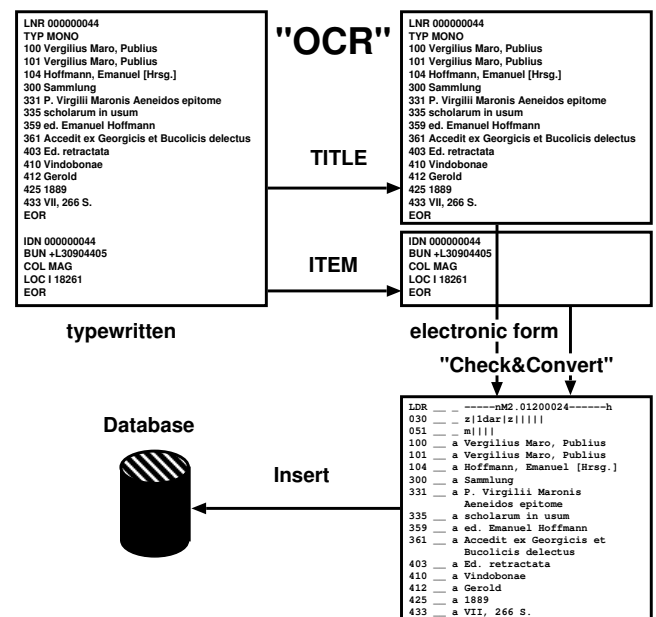


Figure 1: The workflow: schematically illustration

### 3 Building the sample

The developed procedure aimed essentially at processing data records on the basis of „structured texts“ and at loading them into the library system without checking if they are already stored in the system as duplicates or not.

If thus a bibliographic data record to be loaded was already presented in the system (before loading), a duplicate entry for the same title would have been produced. With the premise that such duplicates are to be avoided and respectively may not be produced consciously under any circumstances, it was necessary to decide in the early preliminary stages of gathering the title information whether a „rather large“ or a „rather small“ quantity of already existing titles is to be expected.<sup>6</sup>

This estimation (in the project described here) was substantially affected by the circumstance that bibliographic title-information was not loaded directly into the local library system but rather – according to the usual workflow of cataloging in the Austrian Library Network – in the first step placed in the central library system. Once stored in the central system, these were transferred to the local system by a copying process (replikation).

Therefore it was necessary to examine both systems to finally determine the bibliographic duplicates which could be expected.

In order to sift the inventory numbers (at the beginning) the number of titles for the years of publication from 996 to 1925 – already stored in both systems – was identified.<sup>7</sup>

In 5 year steps the number of existing titles was counted by a database-query and the distribution curve resulting from it was prepared graphically.<sup>8</sup>

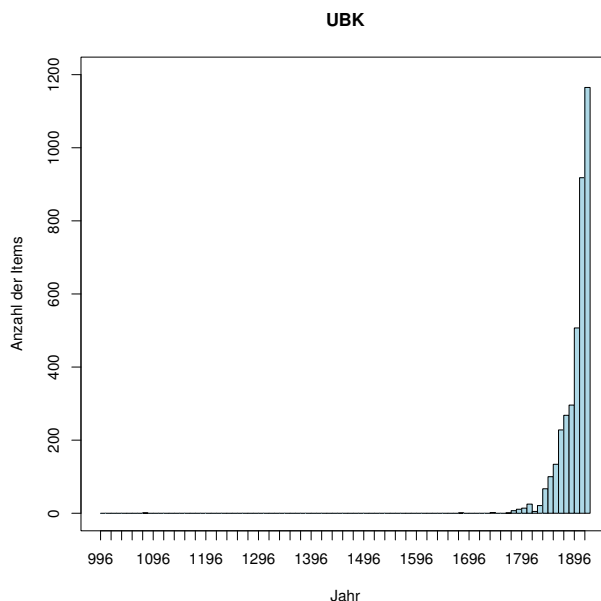
For the chosen statistic procedure the total quantity of the counted titles is not essential.

The comparison between *fig.2* and *3* shows that for the entire regarded range of years the quantity of the countable titles per year is substantially larger in the central library system than in the local system of the University of Klagenfurt. However, it is substantial for the determination of the sample size (with counted values) for

<sup>6</sup> in the determination of the quantities which can be expected thus also the further decision lay, whether such a way is passable – and arguable in the Austrian Library Network

<sup>7</sup> here it is to be pointed out that the determined number of brought in titles, which refer to publication years before 1496 in many cases ascribe to typing errors. However, this disturbs neither the results nor the selected proceeding in further consequence

<sup>8</sup> attention should be paid to the fact, that the evaluated counting originates from November 2000. On this counting also all derived – or in further consequence computed – values base upon



**Figure 2:** Local library system of the University of Klagenfurt: Population of the counted titles within the publication years of 996–1925. Counted in Nov. 2000

our procedure that the two systems show a very similar (actually: the same) population. In particular the considered period of 1796–1925, which is actually substantial for our project, shows a fitting population also on closer examination.<sup>9</sup>

For this reason the sample size for both systems can be selected identically – and it is further not necessary to draw the samples according to a special strategy.

For the computation and respectively the estimation of the sample size the approach shown e.g. in Sachs (1992, p.444) was selected. Thereby it is essentially assumed that at counted values the minimum extent of a sample does not depend on the analyzed population.<sup>10</sup> Rather it is to be assumed that values for the „significance level“ of the expected results (with us: 90%) as well as values for the tolerance bounds of the accomplished measurement (here:  $10 \pm 5\%$ ) must be presumed.

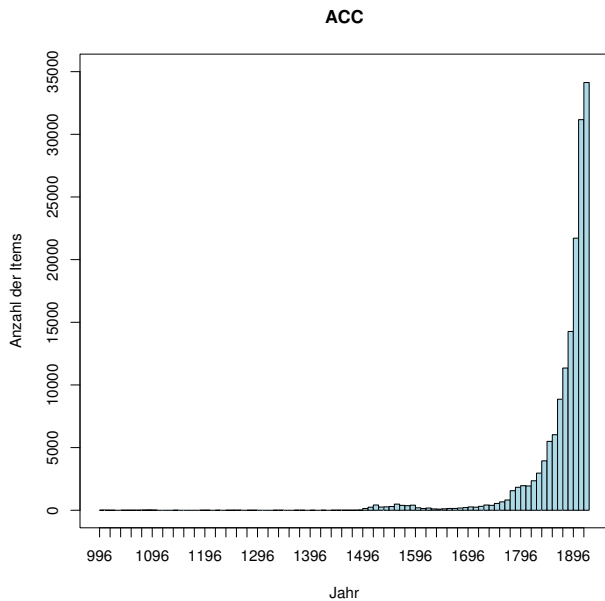
According to these requirements the sample size is calculated as follows:

1. from the significance level of 90% results  
 $\Phi(z) = (\gamma + 1)/2 = 0.95 \rightarrow z = 1.645$
2. with compliance of tolerance bounds of  
 $p = h \pm \varepsilon\% = 10 \pm 5\%$

<sup>9</sup> since in this text all illustrations cannot be shown, these are stored and can be retrieved completely with the following link

[http://www.uni-klu.ac.at/ub/ub-edv/projekte/bock/graphiken/alle\\_abbildungen.pdf](http://www.uni-klu.ac.at/ub/ub-edv/projekte/bock/graphiken/alle_abbildungen.pdf)

<sup>10</sup> this approach is amongst other also referred by Dürr & Mayer (1987, p.136-137)



**Figure 3:** Austrian Central Catalog: Population of the counted titles within the publication years of 996–1925. Counted in Nov. 2000

3. these values are to use on

$$n = \frac{z^2 \cdot h \cdot (1 - h)}{\varepsilon^2} = \frac{1.645^2 \cdot 0.1 \cdot 0.9}{0.05^2} = 97.39 \approx 100$$

The value estimated above states that from the examined call numbers 1–26177 100 must be selected.

These (see *fig.4*) were calculated by a common random process – and subsequently the appropriate titles, in the central catalog as well as in the local catalog were verified for their presence.

On the basis of the sample list it is shown that with the production of the inventory by hand call numbers were not used continuously: some were missing and/or smaller gaps were to be recognized.

For this reason it was decided in that case to choose the next higher actually used call number for the sample.

The thoroughly verification (by hand) showed that out of the 100 examined titles seven (7) in the central catalog system and none in the local catalog were present. This result means that while loading these 100, the coincidentally selected catalog records would produce 7% of bibliographic duplicates in the central catalog (... and none in the local :-).

However, for the check of duplicates by hand following notices are still mentioned:

From the bibliographic descriptions (the catalog records), which were found in the central library catalog in seven cases, our colleague, although very experienced,

80	3649	8540	13031	20396
678	3814	8642	14361	20764
743	3821	9003	14538	20824
758	4015	9082	14628	20941
773	4046	9567	14954	21439
969	4495	9964	15330	21662
1003	4602	9966	15571	21809
1170	5344	10046	15854	22124
1304	5449	10382	15884	22150
1539	5501	10747	15921	22430
1580	5518	10870	16007	23573
1694	5744	11774	16541	23722
1937	6151	11934	16693	23803
2350	6377	12441	17572	24151
2492	6716	12473	18390	24317
2776	6961	12500	18457	24440
2982	7154	12530	18503	24798
3201	8131	12532	18539	24957
3373	8298	12564	19177	24967
3593	8505	12940	19680	25898

**Figure 4:** The call numbers chosen by a common random process in the calculated extent of 100 from the quantity of 1–26177

could not determine always clearly whether the title provided by the catalog completely corresponds to his work or represents (only) a very similar work. In one case this led to the fact that a decision could not be made at all without ordering the appropriate work from the proving library. In two further cases we could agree to regard the bibliographic proof as duplicates in case of doubt.

The reasons for this lay (anyhow) in the very most cases in the – for this time – very typical design of the title pages: The bibliographic information for a sufficient duplicate check, essential for us, was missed occasionally or was not always clearly interpretable.

This knowledge resulted in the attitude that probably many of the duplicates produced show up as such only in those cases, in which the bibliographic data record with the appropriate work can be compared directly (side by side).<sup>11</sup>

## 4 Implementation

The „data acquisition“ happened in this case – as already suggested above – with a (not superseded yet) ball-shaped-head-typewriter on normal, white A4 papersize.

<sup>11</sup> thus above all only under those circumstances, which are called of librarians affectionately „autopsy“ :-)

And as such title-autopsies by users or librarians (systematically) are usually not assumed, the number of title duplicates subjectively perceptible in this procedure is substantially smaller than the number of actually produced and objectively measurable ones

```

LNR 000000044
TYP MONO
100 Vergilius Maro, Publius
101 Virgilius Maro, Publius
104 Hoffmann, Emanuel [Hrsg.]
300 Sammlung
331 P. Virgilii Maronis Aeneidos epitome
335 scholarum in usum
359 ed. Emanuel Hoffmann
361 Accedit ex Georgicis et Bucolicis delectus
403 Ed. retractata
410 Vindobonae
412 Gerold
425 1889
433 VII, 266 S.
EOR

IDN 000000044
BUN +L30904405
COL MAG
LOC I 18261
EOR

```

**Figure 5:** Typewritten catalog record of a monographical work

On these sheets within the upper area the bibliographic information – aligned in two columns – was written:

- the left column essentially contained the identifiers of the MAB2-tags, -indicators and -subfields (for this reason typing-templates, designed on the basis of empirical values, could be provided in the process of the project (similarly to a screen template), which were filled time-saving).  
In order to simplify the further processing additional identifiers were printed apart from the usual tags, which were again removed while transforming the text-structure into data records or simply ignored. Thus, for example, with the identifier LNR the record number was printed before the character transformation with OCR, which became irrelevant in further consequence (after the successful processing), the identifier TYP (=type) (with the allowed values „MONO“ for catalog records of monographic works „HIER“ for catalog records of hierarchically structured editions (e.g. cataloged series or volumes)) indicated the type of the data record, on the basis of which the completeness and the correctness of the existing categories were examined, as well as the indicator EOR, which marked the end of a single record (cf. *fig.5 and 6*)
- in the right column the corresponding catalog information (=the bibliographic data) was printed in the usual way of typing (this means in the correct notation) as an online cataloging librarian does

At the bottom of the sheet the specific copy data (the item information) were noted. Their identifiers are to be interpreted in a similar way as those of the bibliographic entries; their specific labeling form was selected by us at random (arbitrarily):

```

LNR 000000048
TYP HIER
100 Homerus
104 Pauly, Franz [Hrsg.]
105 Pauly, Franciscus
108 Wotke, Karl [Bearb.]
109 Wotke, Carolus
304 Odyssea
331 Homeri Odysseae epitome
335 in usum scholarum
359 ed. Franciscus Pauly
410 Lipsiae u.a.
412 Freytag u.a.
451 Bibliotheca scriptorum graecorum et
451 romanum : A, Scriptores graeci
501 Nebent.: Odyssea. - Überw. in griech. Schr.
530 2. Odysseae lib. XIII - XXIV. - Ed. 4. corr.
530 quam cur. Carolus Wotke, 1887. - XI, 165 S.
EOR

IDN 000000048
BUN +L30904806
COL MAG
LOC I 18256,A,4,2,2
EOR

```

**Figure 6:** Catalog record of a volume typewritten on paper

- the tag identifier IDN contained the (current) counter value of the associated title and constituted thus the reference between the individual copies as well as the appropriate bibliographic catalog record
- with BUN the bar code – stucked in the book by the librarian – was represented
- the content of the tag COL represented the location of the work
- the identifier LOC provided the proper call number of the specific medium

The following processing took place by program control and (at least in case that no errors in the data records which were processed were found) almost noly-break.

That is, that for the librarian-assistant, trained for this job, the application of a „single-button-procedure“<sup>12</sup> was sufficient, in order to save proper title records in the library system made (transformed) from typewritten bibliographic information.

Each typewritten sheet was scanned in the first program step. Afterwards, the generated image file was converted into a standard text file with usual OCR software. Due to the homogeneity of the printed typewriter fonts practically error free results could already be obtained after ten scan runs. Furthermore no other errors could be recognized by the picture-text-transformation in the whole workflow of the project, so that the usual OCR error rate could be completely neglected in this case.

<sup>12</sup> actually such a „single-button-procedure“ could be implemented effectively (and not only in a general manner): with the prominent front side mounted start-button of the scanner all following program steps could be released by a single push

Each text file was immediately processed, that means, the actual production of the individual data records from the scanned information was pursued not through a batch-processing but ad hoc.

In a further step the content of the current text file was checked on the basis of formal criteria.

On the basis of the programming language Perl a simple, sequentially working parser was realized, with which this (according to our requirements) both the typewritten tag information and the formal correctness of the field information were examined. In addition those defined dependences between the individual tags had to be identified (and in the case of an error indicated), which are given by the combined use of the bibliographic data format MAB2 and the german cataloging rules RAK-WB.<sup>13</sup>

```
Record label      LDR  _ _ _ ----nM2.01200024-----
                  h
Enc. det. rec.   030  _ _ _ z|1dar|z|1|1|1|
Spec. det. n-ser 051  _ _ _ m|1|1|1|
1. Person a.f.   100  _ _ _ a Virgilius Maro, Publius
Cross ref. 100   101  _ _ _ a Virgilius Maro, Publius
2. Person a.f.   104  _ _ _ a Hoffmann, Emanuel
                  [Hrsg.]
Collection       300  _ _ _ a Sammlung
Tit. prop. orig. 331  _ _ _ a P. Virgilio Maronis
                  Aeneidos epitome
Furth. tit. inf. 335  _ _ _ a scholarum in usum
Author           359  _ _ _ a ed. Emanuel Hoffmann
Added works      361  _ _ _ a Accedit ex Georgicis et
                  Bucolicis delectus
Ed. stat. orig.  403  _ _ _ a Ed. retractata
Publ. pl. 1. publ. 410 _ _ _ a Vindobonae
Name 1. publ.    412 _ _ _ a Gerold
Years of publ.   425  _ _ _ a 1889
Pagination       433  _ _ _ a VII, 266 S.
```

Figure 7: Example of a generated data record for loading a monographic title information

Besides the checking for errors when parsing<sup>14</sup> the texts first steps were undertaken for data preparation. Since

<sup>13</sup> MAB2 = maschinelles Austauschformat für Bibliotheken = machine readable data interchange format for libraries (version 2), RAK-WB = Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken = German cataloging rules for the alphabetic catalogs in scientific libraries.

Reminded is here besides on the specific meaning, of the already mentioned auxiliary tags MONO and HIER which were added temporarily to bibliographic descriptions and holdings (items) information, with which the formal checking of tag dependences could be handled and further calculated which tags must exist at least in a certain type of a single data record. The given dependences are not described in this text furthermore; they are available completely from our program sources

cf. <http://www.uni-klu.ac.at/groups/ub/ub-edv/projekte/bock/code/bock.pl.txt>

<sup>14</sup> its tasks basically were limited to the function converting the typewritten, paper-oriented external text information into a Perl-internal Hash-oriented data structure which fits better the needs of data programming and in further consequence the handling of texts with the computer

```
Record label      LDR  _ _ _ ----nM2.01200024-----h
Enc. det. rec.   030  _ _ _ z|1dar|z|1|1|1|
Spec. det. n-ser 051  _ _ _ m|1|1|1|
1. Person a.f.   100  _ _ _ a Homerus
2. Person a.f.   104  _ _ _ a Pauly, Franz [Hrsg.]
Cross ref. 104   105  _ _ _ a Pauly, Franciscus
3. Person a.f.   108  _ _ _ a Wotke, Karl [Bearb.]
Cross ref. 108   109  _ _ _ a Wotke, Carolus
Uniform title    304  _ _ _ a Odyssea
Tit. prop. orig. 331  _ _ _ a Homeri Odysseae epitome
Furth. tit. inf. 335  _ _ _ a in usum scholarum
Author           359  _ _ _ a ed. Franciscus Pauly
Publ. pl. 1. publ. 410 _ _ _ a Lipsiae u.a.
Name 1. publ.    412 _ _ _ a Freytag u.a.
1. Coll. tit. ori. 451 _ _ _ a Bibliotheca scriptorum graecorum et
                  romanum : A, Scriptores graeci
Footnotes        501  _ _ _ a Nebent.: Odyssea. - Überw. in griech.
                  Schr.
Tit. ref.        530  _ _ _ a 2. Odysseae lib. XIII - XXIV. - Ed. 4.
                  corr. quam cur. Carolus Wotke, 1887. -
                  XI, 165 S.
```

Figure 8: Example of a generated data record for loading bibliographic information of a volume

with modern hardware the expenditure of time for these program steps is subjectively not recognizable and in addition within the range of milliseconds no longer really measurable, it was decided, to perform all steps of the further processing (the loading procedures into the library system excluded) in each case; even in case that errors have already been recognized during the parsing process.<sup>15</sup>

The proper conversion in this step included e.g. the elimination of redundant blanks or the replacement of those special characters, which were missing on the ball of the typewriter and had therefore to be marked by characters to be replaced later on.<sup>16</sup>

Those text files, which due to the checking conditions appeared to be correct and could be loaded in further consequence into the library system became afterwards those „encoded tags“ added, which are defined in the defaults of MAB2 to keep data records in machine-readable procedures better interpretable. These are among other things the additional tags 030 and 051 besides the information about the Leader LDR.<sup>17</sup>

Since title and item information are to be loaded in the existing library system by different routines, these were splitted into different data records and stored separately according to the most practicable load routines.

As a load format the appropriate system template-format came to application, which is used in the client of the library software also to prepare bibliographic templates for cataloging which shorten recurrent activities in the workflow.

<sup>15</sup> thus in many cases better (more extensively and more meaningfully) error logs for the concerned librarian could be provided

<sup>16</sup> e.g. in our case the character „|“ were chosen as a replacement character for a „|“ and pair of the characters which consists of „|“... and „...|“ became afterwards replaced by „|“... and „...|“

<sup>17</sup> for examples see *fig. 7* and *8*

Since these templates are directly accessible to the Client in its program directory, the concerned directories were furnished in such a way that all check- and load-routines have write- and/or read-access there.

After the check- and convert-process of a text file into the appropriate format the individual data records were stored in this directory up to the beginning of the load-process.<sup>18</sup>

The actual load scripts were realized with a simple script language, which offers comfortable access to the most important basic functions of the operating system. In our case this was implemented with the program packet „MacroExpress“.<sup>19</sup>

The sequence of saving bibliographic data records corresponded to the usual cataloging workflow in the Austrian library network:

1. saving title information in the central library system
2. copying the concerned data record via „push function“ into the connected local library system and saving it locally
3. linking of possibly existing, locally stored holding records with the locally stored title data records (besides this, those local holdings data records are replicated into the central system, which are necessary for a consistent, network-wide holdings display in the central system (e.g. the existence of local journal issues from MAB2 tag 200))
4. linking of item information to the locally stored title records (while this step rudimentary item data into the central system are copied by replication)

The use of the here described method of data loading had the further advantage that the usual data indexing specially did not have to follow the loading and saving processes.

For the system data were loaded in a way as if particularly busy cataloging librarians were at work – and according to the system processes (=the „system behavior“) that through the librarians human interfaces (the routines of the software client) newly stored data always have to be indexed by the system, we did not pay any attention to the fact of separate indexing.

<sup>18</sup> after loading them into the library system they were actually not deleted, but they were shifted for reasons of smaller error rates into a defined target directory

<sup>19</sup> cf. for this <http://www.macroexpress.com>  
A solution by means of „WSH“ (=Windows-Scripting-Host) or „VBA“ (=Visual Basic for Applications) could not be considered, since the client of the library software does not offer a programmable interface (API, Application Programmable Interface) to both

## 5 Summary

The method, to measure and estimate a possible „error“ (in our case: producing duplicates of bibliographic title information) with statistic procedures while loading bibliographic data, proved as useful.

The alternative approach to use an automatically or purely intellectually supported checking method is to be regarded in contrast to this as substantially expensive and much more time consuming – and (as mentioned in the text) also by machine routines very difficult to implement, when accepting thereby a „not too high“ error rate (cf. Jele (2001, p.65-66)).

This can be seen as a substantial and finally also practicable result, of the here introduced and rather unorthodox project, to „bring“ a hand written library inventory over the intermediate step of newly prepared (on paper), typewritten bibliographic catalog information into the data pool of the Austrian central library catalog. Finally the largest expenditure of time was to be reserved for the scan procedure, since each page had to become completely scanned. That means, the purchase of a fast working scanner with the support of an automatic feeder and lower resolution possibly leads to the fact of minimizing the consume of time during the electronic processing. The further program steps like the OCR, the checking- and converting-routines as well as the data loading into the systems represent a rather insignificant time factor.

## 6 Bibliography

### 6.1 Printed sources

- Dürr, Walter & Mayer, Horst (1987): Wahrscheinlichkeitsrechnung und schließende Statistik. 2., vollst. durchges. und verb. Auflage. Hanser, München
- Jele, Harald (2001): Informationstechnologien in Bibliotheken. Oldenbourg, München
- Sachs, Lothar (1992): Angewandte Statistik. Anwendung statistischer Methoden. Siebente, völlig neu bearbeitete Auflage. Springer, Berlin

### 6.2 Online sources

- <http://www.macroexpress.com> : used program for scripting the software client of the library system
- <http://www.uni-klu.ac.at/ub/ub-edv/projekte/bock/code/bock.pl.txt> : our data parsing routines written in Perl
- [http://www.uni-klu.ac.at/ub/ub-edv/projekte/bock/graphiken/alle\\_abbildungen.pdf](http://www.uni-klu.ac.at/ub/ub-edv/projekte/bock/graphiken/alle_abbildungen.pdf) : the completed table of figures

---



**Dr. Harald Jele** head of the EDP department of the University Library of Klagenfurt

Address:

University of Klagenfurt

Universitaetsstrasse 65-67

9020 Klagenfurt, Austria, Europe

Fax: 0043-463-2700-9599

E-Mail:harald.jele@uni-klu.ac.at