

OTTO OBERHAUSER UND HEINRICH MISLIK

GRENZEN DES WACHSTUMS - BIBOS UND DIE LAST DER JAHRE

*O weh! O weh, nun ists vorbei!
Die Last bricht mir das Kreuz entzwei.
(Goethe, Faust II, 4. Akt)*

Wenn die im Österreichischen Bibliothekenverbund eingesetzte Software Bibos-2 nicht vor der Jahrtausendwende durch ein neueres Produkt ersetzt wird, dürfte der 1. Januar 2000 ein interessanter Tag werden. Wie man seit geraumer Zeit weiss -- die Medien sind voll davon -- wird das Jahr 2000 für alle Computersysteme, in denen die Jahreszahlen in den Datumsfeldern zweistellig (statt vierstellig) abgespeichert sind, zu einer kritischen und vor allem teuren Wende werden.¹

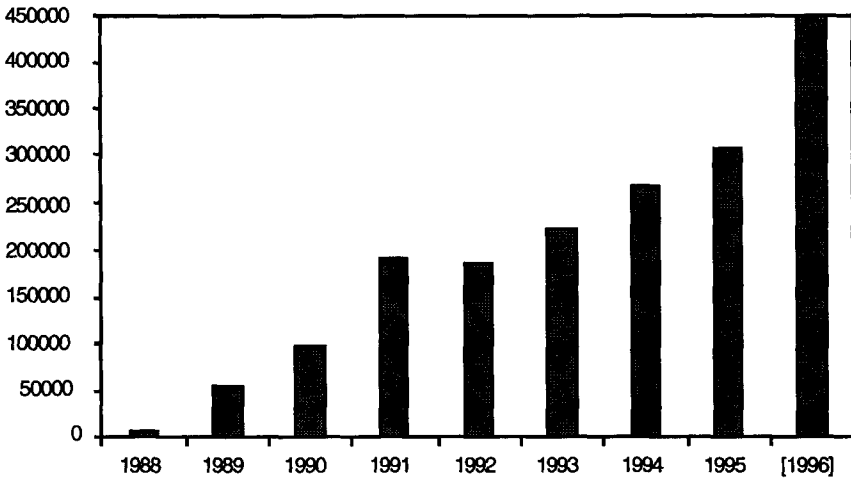
Überall, wo auf der Basis des Datums berechnet (z.B. Lieferzeiten), ausgewählt (z.B. Listendruck) oder sortiert (z.B. Budgetanzeige) wird, wird es zu Problemen kommen, da das System nicht wissen kann, dass z.B. "00" nicht 1900, sondern 2000 bedeuten soll. Bei Systemen, in denen die Verwaltung von Fristen eine bedeutende Rolle spielt (also auch bei bibliothekarischen Entlehnsystemen), wird dieser Effekt aber schon vor oder gar lange vor der Jahrtausendwende eintreten -- so kann etwa für Benutzer mit langen Entlehnfristen bereits etliche Monate vor dem Ende des Jahrtausends kein Buch mehr ausgegeben werden.

Die Datumswende ist allerdings nur eine der zahlreichen Grenzen des Wachstums, denen Bibos-2 unterworfen ist. Der Verbund hat schon heute eine Grössenordnung erreicht, deren Auswirkungen die seinerzeitigen Erfinder des Bibos kaum vorhergesehen haben dürften. Die Abbildung zeigt den Zuwachs der Titelmengen seit Anbeginn der Verbundarbeit (mit einer Extrapolation für 1996); die Datenbank wächst heute rasanter denn je (bedingt durch die Zunahme der Zahl der Verbundbibliotheken sowie der dort auftretenden Datenmengen, nicht zuletzt aber auch aufgrund des Retro-Projektes der ÖNB). Zu den fast 1,5 Millionen Titeln (über 2,7 Millionen Nutzungen bzw. 2,4 Millionen Exemplare) kommen auch noch rund 4,8 Millionen Fremddaten mit einem Wachstum von derzeit etwa 400.000 Datensätzen pro Jahr (ebenfalls zunehmend). Allein auf dem

1 Vgl. z.B.: *Computerwelt*, 18.02.1996, p.10, bzw. <http://www.year2000.com/cgi-bin/clock.cgi/> (die WWW-Seite des *Year 2000 Information Center*).

Produktions-System² lasten täglich rund 250.000 Transaktionen -- mehr als das Doppelte dessen, was für die verwendete Betriebssystemumgebung eigentlich als Limit gilt.

**Wachstum des Verbundes
(jährlicher Zuwachs in Titeln)**



Im folgenden soll es um die Auswirkungen dieser Grössenordnungen und die damit verbundenen Probleme gehen. Wir wollen einerseits all jene Grenzen systematisch darstellen, die für die Bibos-Bearbeiter und zum Teil auch -Benutzer mehr oder weniger sichtbar sind, andererseits aber auch jene Grössenprobleme des Bibos³ aufzeigen, die nur "hinter den Kulissen", also für die Systembetreuer im EDV-Zentrum, erkennbar sind.

2 Der OPAC läuft ja bekanntlich seit mehreren Jahren in einem davon getrennten System.

3 Damit ist hier und im folgenden immer nur Bibos-2 gemeint.

Limits bei der Suche und Ergebnisanzeige

Die Vorteile der EDV kommen bei grossen Datenmengen nur dann zum Tragen, wenn ein adäquates Instrumentarium für die Suche in diesen Daten zur Verfügung steht. Das im Bibos vorhandene Recherche-Instrumentarium, welches in früheren Jahren als ausreichend erscheinen mochte, vermag heute längst nicht mehr zufriedenzustellen. So führen immer mehr Suchen zu Ergebnismengen, die weit über dem Anzeigelimit von 100 Titeln liegen und selbst durch Hinzufügen weiterer Suchbegriffe nicht entsprechend verringert werden können, ohne den Recall entscheidend zu verschlechtern. Zudem fehlen verfeinerte Einschränkungsmöglichkeiten (z.B. NOT-Operator auf der Standard-Suchmaske, Proximity-Operatoren, Begrenzung auf single-word-descriptors⁴). Der Kommandomodus, d.h. die Eingabe von Suchformulierungen auf Basis der Operatoren AND/OR/NOT,⁵ der umso notwendiger ist, je mehr Daten zu durchsuchen sind, steht im Original-Bibos für die Publikumsuche überhaupt nicht zur Verfügung.

Abgesehen davon, dass die Möglichkeiten der Maskierung im Bibos auf eine simple Trunkierung am Wortende beschränkt sind,⁶ wird hier als weiteres, gravierendes Limit eine Begrenzung auf 150 verschiedene Wortendungen wirksam. Diese Zahl erscheint nicht nur im Vergleich mit anderen Retrievalsystemen,⁷ sondern insbesondere im Hinblick auf die deutsche Sprache (die ja bekanntlich besonders stark zur Kompositabil-dung neigt) als völlig unzureichend. Zwar könnte im Bibos dieses sogenannte mask limit -- ein Parameter des verwendeten Retrievalmoduls -- theoretisch verändert werden, doch haben schon vor einigen Jahren am EDV-Zentrum durchgeführte Tests gezeigt, dass bereits eine geringe Erhöhung dieses Parameters verheerende Auswirkungen auf die Antwortzeiten zur Folge hätte. Letztere sind bei der trunkierten Suche -- die mit zunehmender Grösse der Datenbank immer wahrscheinlicher das mask limit erreicht bzw. übersteigt -- ohnedies nicht sehr gut. Erschwerend kommt hier noch dazu, dass das System so programmiert ist, dass es das Überschreiten des mask limit erst dann erkennt, wenn es bereits eine ODER-Verknüpfung der ersten 150 Varianten durchgeführt und

-
- 4 z.B. "Physik" als eingliedrige SW-Kette (im Gegensatz zu "Physik" als Teil einer mehrgliedrigen SW-Kette oder eines mehrgliedrigen Schlagwortes).
 - 5 Im Bibos *freie Suche* genannt.
 - 6 Gerade bei einer überwiegend deutschsprachigen Datenbank dieser Grösse sind die Optionen Links- und Innenworttrunkierung eigentlich unverzichtbar.
 - 7 Beispielsweise beträgt dieses Limit bei DIALOG 2.000, bei STN sogar 20.000 Varianten.

zwischen gespeichert hat. Dies bedeutet nicht nur eine unnötige Systembelastung, sondern auch überflüssige (und aufgrund des oben Gesagten immer häufiger auftretende) Wartezeiten für die schlussendlich ohnedies durch ein Null-Ergebnis frustrierten Bearbeiter bzw. Benutzer.

Das Fehlen einer Phraseninvertierung insbesondere der Autorennamen⁸ macht sich mit zunehmender Grösse der Datenbank bei der Suche immer stärker bemerkbar. Dieses Feature wurde zwar inzwischen in der Softwareversion Bibos-1⁹ realisiert, da man der Meinung war, dass sich dies "mit Zunahme der Daten als immer nützlicher erweisen" würde,¹⁰ nicht aber in Bibos-2. Am Rande sei erwähnt, dass eine solche Phraseninvertierung angesichts der Grösse der Verbund-Datenbank auch in anderen Bereichen (z.B. bei den Feldern der Sacherschliessung) für die Suche vorteilhaft wäre.

Mit wachsender Datenmenge wird das Auftreten von Indexeintragungen (Suchbegriffen) mit hoher Trefferzahl immer wahrscheinlicher. Das im Bibos verwendete Indizierungsmodul bedingt, dass alle Terme, die in einem Subindex für jeweils 400.000 Dokumente auf mehr als 16.384 Titel zeigen, der sogenannten Bitinvertierung (einer ökonomischeren Indizierungsform) unterworfen werden. Diese Bitinvertierung macht vor allem bei der Buchbearbeitung zu schaffen, kann aber auch Auswirkungen auf die Suche haben: Überschreitet nämlich die Trefferzahl für einen bestimmten Begriff die genannte magische Grenze (und wurde dieser Begriff nicht in weit vorausblickender Weise beim jeweils letzten Neuaufbau des Index bitinvertiert), so ist das betreffende Wort plötzlich nicht mehr suchbar. Alterfahrenen Anwendern mag das Beispiel "Wien" noch in schauriger Erinnerung sein.

Seit langem ist bekannt, dass die Hierarchiesuche zu den besonders systembelastenden Features zählt. Vor allem bei der Anzeige aller Titel einer (grossen) Hierarchie kommt es zu extrem langen Antwortzeiten, zumal bei der Hierarchiesuche kein Anzeigelimit besteht und das System gegebenenfalls tausende Titel für die sortierte Anzeige aufbereiten muss. Je grösser die Datenbank wird, desto grösser werden auch bestimmte Hierarchien, und desto grösser wird auch der Bedarf, sie angezeigt zu erhalten. Diese Problematik ist in Bibos vermutlich nicht mehr lösbar.

8 Vgl.: Wagner, H.: PICA versus BIBOS. *Bibliotheksdienst*, 27 (7) 1993, p.1013-1025.

9 Diese wird in einem anderen Bibliothekenverbund für die Verwaltung etwa der Hälfte des oben genannten Titelvolumens eingesetzt.

10 EDV Ges.m.b.H.: *BIBOS Leistungsbericht, 1994: Bibliothekenverbund AKW, BMUKA, LBB*. Wien 1995, p. 4.

Auch bei der Ergebnisanzeige gibt es grössenbedingte Effekte. Je mehr Exemplare zu einem Titel vorhanden sind (z.B. bei Lehrbüchern), desto störender wirkt sich die -- durch die Buchbearbeitung nicht beeinflussbare -- Aufreihung der Exemplare aus, die nicht nach Standorten bzw. Exemplarzählern, sondern im wesentlichen chronologisch (d.h. nach dem "Alter" des Exemplarsegments) erfolgt. Grenzen gibt es aber auch bei besonders langen Datensätzen, bei denen die Suchbarkeit und auch die Anzeige systemseitig eingeschränkt sind.

Limits bei der Buchbearbeitung

Grössen- und Mengenprobleme belasten zunehmend auch die Buchbearbeitung. Es ist z.B. keine Neuigkeit, dass das Ausdrucken von Buchhändlerbestellungen oft sehr lange dauert und -- ab einer gewissen Menge -- zu einer unverhältnismässigen Systembelastung (mit extremen Einbrüchen in der Performance) führt. Zum Teil ist dies darauf zurückzuführen, dass das System beim Generieren der Bestellzettel mit grossem Aufwand (sequentiell) die Datenbank der sogenannten Drucksegmente¹¹ durchsucht. Da für zahlreiche Erwerbsvorgänge (z.B. Rechnungslegung, Stornierung) die Drucksorten nie angefordert werden, löscht das System die entsprechenden Drucksegmente nicht -- wie vorgesehen -- nach sechzig Tagen, sondern belässt sie auf Dauer in der Datenbank. Zwar versucht das EDV-Zentrum, dieser Entwicklung entgegenzuwirken, indem seit Herbst 1995 täglich alle über ein Jahr alten Drucksegmente sowie jene für die Rechnungslegung vorgesehenen gelöscht werden, doch kann dadurch das Wachstum dieser Datenbank nicht wirklich aufgehalten werden.

Selbst dort, wo willkürlich gesetzte Systemparameter noch verändert werden können, zeigen sich die definitiven Grenzen des Wachstums. Erst kürzlich stellte sich heraus, dass die Suche nach solchen Bestellnummern, die bei mehr als 600 Exemplaren vergeben worden waren, unvollständige Ergebnisse zur Folge hatte. Die Hinterfragung dieses Problems führte zur Erhöhung des betreffenden Limits (durch die Herstellerfirma) auf 4.000 Exemplare, was das systemseitige Maximum bedeutet. Es ist aber bereits abzusehen, dass in der Realität höhere Fallzahlen auftreten werden. In einem anderen Fall -- bei der Kurzanzeige Lieferant/Erwerb für alle Bestellungen -- führte schon vor längerer Zeit das Fehlen eines systemseitig gesetzten Limits zu verbundweiten Abstürzen, sobald eine kritische Grenze überschritten wurde. Inzwischen hat die Herstellerfirma ein absolutes Systemlimit eingeführt (8.000 Bestellungen).

11 Daten, die für die Erstellung der "Drucksorten für den Erwerb" benötigt werden.

Eine andere Funktionalität, die Performanceprobleme bewirkt, ist die online aufgerufene Kurzanzeige Bearbeiter. Auch hier wird eine Datenbank sequentiell durchsucht, um eine sortierte Anzeige zu erstellen -- ein Vorgang, der in der Frühzeit des Bibos unproblematisch gewesen sein mag, beim heutigen Stand von mehr als 1.000 vergebenen Bearbeiterparaphen jedoch verbundweite Wartezeiten von erheblichen Ausmassen bewirkt.¹²

Sogar bei der Verwaltung kleiner Datenbanken zählt die Möglichkeit, globale Änderungen vorzunehmen, zu den nützlichen Features. Ab einer gewissen Grösse muss diese Funktionalität jedoch als unverzichtbar gelten; wo sie -- wie im Bibos -- fehlt, kann die bearbeiterseitig erforderliche Datenpflege nur mehr mit extrem hohem Aufwand (d.h. Datensatz für Datensatz) und wohl auch nicht mehr zuverlässig durchgeführt werden.¹³ Das Fehlen einer Verwaltungskomponente für Normdateien im Bibos trägt zudem dazu bei, dass viel mehr einzelne Änderungen als eigentlich nötig wären durchgeführt werden müssen.

Mit zunehmender Grösse des Verbundes bzw. mit wachsender Datenmenge in den (grossen) Verbundbibliotheken erweist sich der Umstand, dass Bibos-2 nicht in der Lage ist, eine Bibliotheksstruktur (zentrale und dezentrale Bibliotheken) abzubilden, als immer gravierender. Die Funktionalitäten, bei denen dies notwendig wäre, sind zu zahlreich, um hier im Detail abgehandelt zu werden (z.B. Lieferantenverwaltung, Drucksorten für den Erwerb, OPAC-Anzeige, Sacherschliessung).

Limits für Betrieb und Wartung des Systems

Eine wesentliche Ursache für die Grenzen des Wachstums, von denen hier die Rede ist, liegt im Betriebssystem (VSE) und den damit verbundenen sonstigen Systemkomponenten (DL/1 als Datenbank- bzw. CICS als Transaktionssystem), die 1987 als Basis für den Bibos-Betrieb im zukünftigen Bibliothekenverbund ausgewählt worden sind.¹⁴ In ihrer Kom-

12 Als Ausweg mag die vom EDV-Zentrum monatlich erstellte und (ausserhalb des Bibos) online abrufbare Bearbeiterliste gelten, doch kann dadurch nicht verhindert werden, dass die *Kurzanzeige Bearbeiter* (etwa durch unbeabsichtigte Eingabe einer falschen Paraphe) trotzdem fallweise ausgelöst wird.

13 Die Verwendung von Makros (bearbeiterseitig) oder Programmen (EDV-Zentrum) auf Basis der PC/3270-Software erleichtert zwar manche globalen Änderungen, kann aber nicht als Ersatz für ein entsprechendes Feature im System selbst gesehen werden.

14 Vermutlich fiel die Wahl auf VSE nicht nur aus Kostengründen, sondern da die einzige Alternative (MVS) auf den damals geplanten lokalen Bibos-Rechnern nicht lauffähig gewesen wäre.

bination -- und zudem mit der verwendeten Programmiersprache PL/I -- vermögen diese Komponenten das Potential moderner Grossrechner nicht ausreichend zu nutzen. Unter VSE muss sich eine Bibos-Installation mit 16 Megabyte Hauptspeicher begnügen -- eine Grössenordnung, die heute sogar im PC-Bereich, wo man die Ressourcen ausserdem nicht mit hunderten Programmanwendern teilen muss, keine Grenze mehr bedeutet.¹⁵ Dass die Transaktionsraten, für welche diese Betriebssystemumgebung als geeignet erachtet wird, längst überschritten worden sind, wurde bereits oben erwähnt. Aufgrund von CICS-internen Limits ist es ferner nicht möglich, mit dem im Bibos eingesetzten Indexsystem wirklich grosse Indizes zu verwalten,¹⁶ weswegen anstelle eines Gesamtindex mehrere Sub-Indizes für jeweils 400.000 Dokumente verwendet werden. Dies führt nicht nur zu einem Mehrverbrauch von Plattenplatz, sondern vor allem auch zu umso höheren Antwortzeiten, je mehr solche Sub-Indizes nötig sind. Als möglicher Ausweg zur Überwindung dieser Grenzen wurde manchmal der Umstieg auf das mächtigere Betriebssystem MVS genannt, der jedoch neben enormen finanziellen und personellen Ressourcen auch eine mehrjährige Vorbereitungszeit erfordert hätte und somit nie realistisch war.

Aufgrund all dieser Gegebenheiten sowie aufgrund des nichtlinearen Wachstums der Datenmenge nähern sich auch die zahlreichen Offline-Verarbeitungen, die mit dem Bibos-Betrieb verbunden sind, zunehmend den Grenzen des Machbaren, sowohl was die zeitlichen, als auch die sonstigen Ressourcen betrifft. Als Beispiele hierfür seien angeführt:

- Der für die Produktion der ÖVK-WB CD-ROM erforderliche Export der gesamten Bibos-Daten dauerte zuletzt bereits über 40 Stunden (im Vorjahr lag dieser Wert noch bei etwa 30 Stunden);
- Eine vollständige Neuinvertierung der Produktions-Datenbank würde heute bereits eine Stilllegung des Systems für drei volle Tage voraussetzen. Da dies dem Produktionsbetrieb nicht zumutbar ist, andererseits aber wegen der Dynamik der Verbunddaten eine regelmässige Neuinvertierung immer dringender notwendig wird,¹⁷ wurde am

15 Mit erheblichem Aufwand (finanziell und organisatorisch) wäre es heute möglich, dieses Limit zu überwinden. Mit einer Realisierung ist in absehbarer Zukunft jedoch kaum zu rechnen.

16 Dies wurde übrigens erst erkannt, nachdem 1988 der Fremddaten-Grundbestand in das System geladen worden war und daraufhin CICS-Abstürze auftraten.

17 Aufgrund der *Bitinvertierung* bestimmter Wörter bzw. einiger Datenfelder (Standort, Signatur etc.) können manuell vorgenommene Datenänderungen dazu führen, dass Begriffe nicht mehr suchbar sind, obwohl sie in der Datenbank aufscheinen.

EDV-Zentrum ein Verfahren entwickelt, mit Hilfe dessen ein einzelner Sub-Index invertiert und so die Gesamtinvertierung auf mehrere Sonntage verteilt werden kann.

- Eine vollständige Neuinvertierung der Fremddaten ist mit der systemseitig bereitgestellten Software vermutlich gar nicht mehr möglich und auch mit Spezialroutinen des EDV-Zentrums nur unter grossem Aufwand durchführbar. Aus Gründen der Datenintegrität (Bereinigung der Effekte früherer Fehler im Ladeprogramm) wäre diese Massnahme jedoch wünschenswert.
- Das wöchentliche Kopieren der Bibos-Daten für das OPAC-System, das bei Stillstand des Produktions-Systems erfolgen muss, verbraucht einen immer grösseren Teil (bereits über fünf Stunden) der Zeit, die während der Nacht für derartige Arbeiten zur Verfügung steht.
- Auch die Batchlisten-Verarbeitung, die gleichfalls während der Nacht erfolgt, ist zunehmend von Grössenproblemen betroffen, da die Laufzeit bestimmter Listenjobs (vor allem solcher, bei denen eine grössere Datenmenge durchforstet werden muss, z.B. bei Inventar- und Standortverzeichnissen bzw. Budgetlisten) mitunter bereits im Stundenbereich liegt. Daher ist es nicht immer möglich, alle angeforderten Listen in der jeweils folgenden Nacht zu erstellen.
- Ähnliches gilt für die Datensicherung, die täglich, wöchentlich und monatlich durchgeführt wird (ebenfalls während des Bibos-Stillstands) und die aufgrund des Gesamtwachstums immer länger dauert bzw. auch immer grössere Speicherressourcen benötigt.
- Schliesslich sind von den wachsenden Datenmengen auch sämtliche zur Datenbankpflege zu zählenden Arbeitsvorgänge (Datenbankreorganisation, Formatanpassung an neue Softwareversionen) betroffen. Diese Massnahmen benötigen nicht nur immer mehr Zeit, sondern auch immer mehr Plattenplatz (der so gut wie nie im Überfluss zur Verfügung steht).

Verborgene Fussangeln

Wie jedes komplexere Softwareprodukt, so enthält auch Bibos eine ganze Reihe von Fehlern, die nur im Echtbetrieb und oft nach Jahren zutage treten. Etliche der bisher bekannten Fehler zeigten sich jeweils erst, nachdem eine bestimmte Grössenordnung erreicht worden war. Typische Beispiele dieser Art sind:

- Lieferte eine Hierarchiesuche genau 100 Treffer, so kam es auf der letzten Seite der Kurzanzeige zur fehlerhaften Einblendung von Titeln aus früheren Suchen.

- Als die Zahl der Fremddaten die Millionengrenze überschritt, stellte sich heraus, dass die Trefferanzeige nur für sechsstelligen Zahlen vorgesehen war.
- In der Online-Budgetstatistik waren die Exemplarzähler für mehrere Bestellarten nur vierstellig, also für Zahlen bis 9.999, vorgesehen.
- Führt die Suche nach einer ISBN zu mehr als 100 Treffern, so führt dies für den betreffenden Bearbeiter zu einem Absturz.
- Als erstmals ein Titel im Verbund von mehr als 20 Bibliotheken genutzt wurde, so kam es bei dem Versuch, diesen Titel zu bearbeiten, zu Antwortzeiten im Stundenbereich sowie anschliessend zu einer fehlerhaften Ergebnisanzeige.
- Wiederholt kam es zu Fehlern bei grossen und/oder komplexen Hierarchien (Absturz von Batch-Listen, Funktionsstörungen bei der Buchbearbeitung, extreme Performanceeinbrüche bei der Online-Anzeige).

Es ist zu befürchten, dass Fehler dieser Art mit dem weiteren Wachstum der Datenbank immer wieder auftreten werden. Man muss sich in diesem Zusammenhang bewusst sein, dass die Bibos-Software nie zuvor für grössere Datenmengen eingesetzt worden ist, als dies im Österreichischen Bibliothekenverbund der Fall ist. Das bedeutet gleichzeitig auch, dass Vorhersagen jeglicher Art über das zukünftige Verhalten und die Tragfähigkeit dieser Software unter der ständigen zunehmenden Last ein unkalkulierbares Wagnis darstellen.

Abschliessend sei darauf hingewiesen, dass es zu einem erheblichen Teil der im Laufe der Jahre immer wieder erneuerten und auf den letzten Stand der Technik gebrachten Hardware (Rechner, Plattenspeicher, Netzwerk) zu verdanken ist, dass angesichts all des oben Gesagten der Betrieb des Verbundes mit der Bibos-Software heute noch möglich ist. Auf die Technik allein wird man sich aber nicht auf Dauer verlassen können.