

# A FADING HERITAGE - PERMANENT PRESERVATION OF DIGITAL PRIVATE COLLECTIONS

Karl-Ernst LUPPRIAN\*

**LUPPRIAN, Karl-Ernst, A FADING HERITAGE - PERMANENT PRESERVATION OF DIGITAL PRIVATE COLLECTIONS**, *Atlanti*, Vol. 15, No. 1-2/2005, Trieste 2005.

Original in German, abstract in English, summary in English.

The outcome of a researcher's life will be measured by his return in publications. But a researcher's private collections more often contain a rich wealth of further materials. If these exist in a digital format and have reached a certain age since the acquisition of the collections difficulties may be encountered on the attempt to read the files, especially databases. A successful permanent preservation of digital private collections without unreasonable effort and expenses depends on the observation of certain rules followed as early as at the time of creation of the files. Furthermore it will be necessary to classify the institution acquiring the collections as trustworthy regarding permanent digital preservation.

## **Descriptors (ATLANTI)**

*Storage: Computertapes (3.1.6)*

Die Probleme, die bei der langfristigen Aufbewahrung und Zugänglichmachung elektronischer Unterlagen zu lösen sind, werden seit geraumer Zeit ausführlich diskutiert<sup>1</sup>. Der Diskurs beschränkte sich bislang jedoch vorwiegend auf Kulturgutbewahrende Institutionen wie z.B. Archive, Bibliotheken und Museen. Kaum erörtert wurden diese Schwierigkeiten aus der Sicht auf wissenschaftliche Nachlässe in digitaler Form, die erst nach dem Tod des Gelehrten oder noch viel später in eine solche Institution gelangen und dann aus technischen Gründen häufig nicht mehr zugänglich gemacht werden können<sup>2</sup>.

Anhand von zwei fiktiven Szenarien soll dargestellt werden, welche Maßnahmen der Nachlassbildner selbst treffen sollte bzw. muss, damit seine elektronischen Unterlagen überhaupt für eine langfristige Aufbewahrung in benützbarer Form geeignet sind. Danach ist zu fragen, welcher Einrichtung der Nachlass übergeben werden soll, um

---

\* Archivdirektor, Generaldirektion der Staatlichen Archive Bayerns, Schönfeldstr. 5, D-80539 München, Deutschland, Tel. ++49 89 28638-2484, E-Mail: iuk@gda.bayern.de.

<sup>1</sup> Zusammenfassend zuletzt von Uwe M. Borghoff et al., *Langzeitarchivierung*, Heidelberg 2003 und Charles M. Dollar, *Authentic electronic records: strategies for long-term access*, Chicago 2000.

<sup>2</sup> Auf individuelle elektronische Sammlungen geht Bryan Bergeron, *Dark Ages II: when the digital data die*, Upper Saddle River, 2002, ein.

für die Nachwelt erhalten zu werden, denn angesichts der hohen technischen, aber auch organisatorischen und letzten Endes finanziellen Anforderungen steht außer Frage, dass Einzelpersonen dies nicht leisten können.

### **Szenario 1 - der Historiker**

Franz Dickbrettbohrer hat sich über Jahrzehnte mit der Maxhütte in Sulzbach-Rosenberg (Ostbayern), einer bedeutenden, aber inzwischen stillgelegten Stahlrohrfabrik, beschäftigt. Er sammelte sowohl Unterlagen zu den technischen Prozessen (Lage- und Baupläne, Konstruktionszeichnungen) als auch betriebswirtschaftliche und die soziale Entwicklung der Arbeitnehmer betreffende Quellen. Während der ältere Teil der Unterlagen auf Papier vorliegt, besteht die Masse aus Dateien verschiedenster Form. Unter anderem finden sich:

Lagepläne der Fabrikgebäude im Format der digitalen Flurkarte (DFK),  
Baupläne in verschiedenen CAD-Formaten,  
Konstruktionszeichnungen (CAD), eingebettet in eine askSam-Datenbank,  
Betriebswirtschaftliche Kalkulationen im Format Microsoft Excel,  
Jahresberichte und Bilanzen im Format PDF,  
Fotografien in den Formaten CGM und JPEG,  
eine GOLIATH-Datenbank mit Arbeitnehmer-Statistiken,

Auswertungen und Entwürfe zur Publikationen im Format Microsoft Word (DOC) in verschiedenen Versionen.

All diese Dateien sind auf Disketten gespeichert, einige davon im Format 5,25 Zoll, die Mehrzahl im Format 3,5 Zoll. Die Fotografien sind auf einer Laserdisk abgelegt.

Einige Jahre nach dem Tod von Franz Dickbrettbohrer wird der Nachlass von seinen Erben dem Bayerischen Wirtschaftsarchiv übergeben. Seinen Rechner haben die Erben "entsorgt".

Im Archiv muss zunächst ermittelt werden, welche Daten sich auf den einzelnen Datenträgern befinden, denn der Verstorbene hat hierzu kein Verzeichnis angelegt. Dann versuchen die Archivare, auf die Daten zuzugreifen. Glücklicherweise verfügen sie noch über ein Laufwerk für Disketten im Format 5,25 Zoll, doch für die Laserdisk gibt es keines mehr. Nach etlichen Monaten liegen die Ergebnisse der "datenarchäologischen" Bemühungen vor:

Lagepläne der Fabrikgebäude: da das DFK-Format dokumentiert ist, konnten diese Pläne mit Unterstützung des Landesvermessungsamts in das offene Format Scalable Vector Graphics (SVG) konvertiert und so gerettet werden.

Baupläne: von den verschiedenen, proprietären CAD-Formaten ließ sich nur eines lesen. Von mehr als 200 Plänen konnten nur 15 konvertiert und gerettet werden.

Konstruktionszeichnungen: Während die askSam-Datenbank ausgelesen und in eine MySQL-Datenbank gewandelt werden konnte, waren sämtliche Zeichnungen unlesbar.

Betriebswirtschaftliche Kalkulationen: Mit einer aktuellen Excel-Version konnten diese Dateien geladen werden. Aufgrund von Inkompatibilitäten mit der alten Excel-Version stimmten jedoch die Berechnungsergebnisse nicht mehr und mussten manuell nachgearbeitet werden.

Jahresberichte und Bilanzen: Diese Dateien lagen im Format PDF Version 1.4 vor und konnten in PDF/A<sup>3</sup> umgespeichert werden. Da die Fonts jedoch nicht innerhalb der Quelldateien abgelegt waren, musste auf Ersatzfonts zurückgegriffen werden, wodurch die Tabellenstruktur z.T. zerstört wurde.

Fotografien: Da für die Laserdisk kein Laufwerk zu finden war, musste auf die Rettung verzichtet werden, ungeachtet der Standard-Dateiformate, die keine Probleme verursacht hätten.

Eine GOLIATH-Datenbank mit Arbeitnehmer-Statistiken: Das Datenbanksystem GOLIATH hatte keine Exportmöglichkeit, und das interne Datenformat war nicht offengelegt. Die Daten sind also verloren.

Auswertungen und Entwürfe zur Publikationen: Etwa die Hälfte der DOC-Dateien konnte mit einer aktuellen Word-Version layoutgetreu auf PDF/A umgespeichert werden. Von der anderen Hälfte ließ sich wenigstens der Text - wenn auch mit z.T. fehlenden oder falsch dargestellten Sonderzeichen und veränderten Seitenumbrüchen - retten.

## **Szenario 2 - der Genealoge**

Hubert Grabschaufler hat in drei Jahrzehnten die Geschichte seiner weitverzweigten Familie in Mittelfranken erforscht. Er begann als nur an Stammbäumen interessierter Familienforscher, eignete sich jedoch im Lauf der Zeit die notwendigen Kenntnisse an, um seine Arbeit auf wissenschaftliches Niveau zu heben. Hierbei unterstützte ihn die Gesellschaft für Familienforschung in Franken e.V. (GFF)<sup>4</sup>, deren Mitglied er bis zu seinem Ableben war. In seinem Nachlass, der zum größten Teil in elektronischer Form vorliegt, finden sich:

Gescannte Mikrofilme aus Kirchenbüchern im Format TIFF (binär),  
Abschriften von diesen Mikrofilmen im Nur-Text-Format, kodiert nach ISO-Latin-1,  
Gezeichnete Stammbäume im Format Scalable Vector Graphics (SVG),  
Fotografien von Familienangehörigen im Format JPEG,  
Biografische Ausarbeitungen im Format Microsoft Word 2000,  
eine Datenbank mit Personendaten im Format GEDCOM.

---

<sup>3</sup> Vgl. ISO/CD 19005-1 "Document management - Electronic document file format for long-term preservation - Part 1: Use of PDF (PDF/A)", [http://www.aiim.org/documents/standards/ISO\\_19005-1\\_\(E\).doc](http://www.aiim.org/documents/standards/ISO_19005-1_(E).doc). Eine erste praxisnahe Umsetzung sind die "Empfehlungen zum Erzeugen archivierbarer Dateien im Format PDF" der Österreichischen Nationalbibliothek vom 7.7.2005, herunterladbar unter dem URL [http://www.onb.ac.at/about/lza/ONB\\_PDF-Empfehlungen\\_1-3](http://www.onb.ac.at/about/lza/ONB_PDF-Empfehlungen_1-3).

<sup>4</sup> Siehe <http://www.gf-franken.de>.

Zu diesem Material existiert ein gedrucktes Verzeichnis aller Dateien mit Format- und Inhaltsangaben. Die Dateien selbst sind durchwegs auf CD-ROM hoher Qualität abgelegt; von jeder CD ist eine Kopie vorhanden.

Noch zu Lebzeiten hat Hubert Grabschaufler testamentarisch verfügt, dass die Unterlagen nach seinem Tod der GFF zur künftigen Verwahrung zu übergeben sind.

Die Gesellschaft betreibt ein eigenes Archiv, das im Staatsarchiv Nürnberg untergebracht ist. Sämtliche elektronischen Unterlagen sind auf CD-ROM hoher Qualität gespeichert, und zwar dreifach an drei verschiedenen Lagerorten. Die Inhalte werden in einer zentralen Datenbank nachgewiesen. Zusätzlich werden alle Nachlässe vollständig - sowohl die papierenen als auch die digitalen Teile - auf Mikrofilm ausbelichtet. Der Zugang zum Archiv steht jedem ernsthaft Interessierten gegen eine geringe Benützungsgebühr offen.

Zwei Informatiker, die ehrenamtlich für die GFF tätig sind, sichteten den Datenbestand. Von den TIFF-Dateien wurden für eine einfachere Benützung Kopien im Format PDF/A gezogen. Auch die Worddateien wurden in dieses Format konvertiert. Die Inhalte der GEDCOM-Datenbank wurde mit einem selbstgeschriebenen kleinen Programm in ein Flatfile in dem an XML angelehnten, offenen Format GedML (Genealogical Data Markup Language)<sup>5</sup> konvertiert. Das Flatfile wurde danach in eine MySQL-Datenbank eingelesen. Diese Arbeiten konnten nach drei Wochen abgeschlossen werden.

### ***Schlussfolgerungen aus den Szenarien***

Franz Dickbrettbohrer hat bei der Ablage seiner elektronischen Unterlagen offenbar nicht an eine langfristige Nutzungsmöglichkeit gedacht. Er speicherte seine Dateien in den Formaten, die ihm die Anwendungsprogramme anboten, und auf Datenträgern, die zu seiner Zeit gerade gängig waren. Er unterließ es auch, ein Verzeichnis der Datenträger und ihrer Inhalte anzulegen, was angesichts des Verlustes seines Rechners besonders nachteilig war.

Hubert Grabschaufler hingegen hat sich - vielleicht aufgrund von Ratschlägen der GFF - rechtzeitig Gedanken über die Nutzung seiner Unterlagen nach seinem Tode gemacht. Er speicherte seine Dateien zum größten Teil in Standardformaten und auf Datenträgern eines Typs und führte ein Verzeichnis darüber. Die testamentarische Verfügung über seinen Nachlass bereits zu Lebzeiten sicherte dessen Fortbestand auch rechtlich ab. Mit der Gesellschaft für Familienforschung in Franken entschied sich der Nachlassgeber für eine Institution, die ihr Archivgut - sei es auf Papier oder in elektronischer Form - professionell sichert und durch die zusätzliche Mikroverfilmung auch für sehr lange Zeit der Nachwelt erhält. Damit hat Hubert Grabschaufler sich für ein "vertrauenswürdige Archiv" (trusted repository) entschieden.

---

<sup>5</sup> Vgl. <http://users.breathe.com/mhkay/gedml/rationale.html>. Ein Vergleich zwischen einem GEDCOM- und einem GedML-Datensatz (<http://users.breathe.com/mhkay/gedml/example.html>) zeigt die relativ einfache Umsetzbarkeit in GedML. Die Arbeiten an der Document Type Definition für GedML sind noch nicht abgeschlossen. Doch schon GEDCOM-Datensätze lassen sich als Nur-Text-Dateien ausgeben, sind also prinzipiell bereits archivierungsfähig.

Die beiden Szenarien belegen, dass ein Wissenschaftler sich beizeiten bemühen muss, die Voraussetzungen zu schaffen, um seine elektronischen Unterlagen der Nachwelt zu erhalten. Schon bei der Beschaffung von Anwendungssoftware sollte er prüfen, ob diese in der Lage ist, die erzeugten Daten in offengelegten und lizenzfreien Formaten zu speichern. Hierfür können einige Empfehlungen gegeben werden:

**Texte:** Wenn es nur um Inhalte, nicht aber um das Layout geht, bieten sich Nur-Text-Dateien (plain text) mit ASCII- bzw. ISO-Latin-1-Kodierung an. Hierbei müssen bei diakritischen Zeichen Einschränkungen hingenommen werden. Eine Notlösung ist die Speicherung in HTML (besser: XHTML), die das Layout in Grenzen erhält und die Maskierung diakritischer und anderer Sonderzeichen erlaubt. Ideal ist die Verwendung von XML für die logische Struktur und den Inhalt und eines Stylesheets für das Layout.<sup>6</sup>

Alle diese Dateien können jederzeit weiter bearbeitet werden. Ist dies im Interesse einer authentischen Überlieferung unerwünscht, so sollte man sich für PDF/A entscheiden.

**Bilder:** Für Rasterbilder ist TIFF das "klassische" Format für die Archivierung. Nachteilig sind die Dateigrößen (insbesondere bei Farbbildern), die einen zügigen Umgang mit den Dateien erschweren oder gar unmöglich machen können (z.B. den Versand mittels E-Mail). Das insbesondere für Fotografien beliebte Format JPEG erzeugt stark komprimierte, jedoch verlustbehaftete Dateien. Den letztgenannten Nachteil vermeidet das allerdings noch wenig verbreitete JPEG2000.

Für Vektorgrafiken (z.B. Konstruktionszeichnungen) empfiehlt sich das zur XML-Familie gehörende SVG.

Noch keine langzeitfähige Lösung ist für Audio- und Videodaten in Sicht.

**Datenbanken:** Ein Datenbanksystem sollte in der Lage sein, die Datensätze mit den Feldbezeichnungen in ein Nur-Text-Flatfile zu exportieren, das mit XML strukturiert ist. Unbedingt wichtig ist die Aufbewahrung der Dokumentation zum Datenbanksystem, insbesondere zu seiner internen Struktur. Dadurch sollte es später möglich sein, die Datensätze in ein aktuelles Datenbanksystem zu importieren.

Problematisch wird es jedoch, wenn ein Datenbanksystem Feldinhalte intern als Binary Large Objects (BLOBs) speichert. Der Inhalt dieser BLOBs muss dann eigens exportiert und bearbeitet werden.

**E-Mail:** Mails bestehen aus Kopf (Header), Inhalt (Body) und Anhängen (Attachments). Kopf und Inhalt lassen sich als Nur-Text-Dateien, gegebenenfalls in einer XML-Struktur speichern. Die Anhänge müssen je nach Dokumenttyp eigens konvertiert werden.

---

<sup>6</sup> Im Bereich der Naturwissenschaften ist TEX bzw. LATEX weit verbreitet. Dabei entstehen Nur-Text-Dateien mit 7-Bit-ASCII-Kodierung und eingebetteten Layout-Anweisungen. Aus TEX-Dateien lassen sich Druckausgaben hoher Qualität, aber auch entsprechende PDF-Dateien erzeugen.

## **Wohin mit dem Nachlass? Vertrauenswürdige Archive**

Wie oben erwähnt, hat Hubert Grabschaufler seinen Nachlass einem "vertrauenswürdigen Archiv" übergeben. "Archiv" ist hier ganz allgemein als eine speichernde Institution zu verstehen, die besondere Vorkehrungen für die Langzeitsicherung und -verfügbarkeit getroffen hat. Neben technischen Maßnahmen sind hier besonders organisatorische und finanzielle Absicherungen zu bedenken, die das Bestehen der Institution für lange Zeiträume erwarten lassen. In jüngster Zeit wird diskutiert, ob und mit welchem Aufwand sich solche Einrichtungen als vertrauenswürdig zertifizieren lassen<sup>7</sup>. Sicher ist nur, dass eine zuverlässige Langzeitarchivierung - in welcher Form auch immer - viel Geld kosten wird.

Damit dürfte auch jedermann bewusst geworden sein, dass der einzelne Wissenschaftler die langfristige Erhaltung seiner elektronisch vorliegenden Unterlagen nicht selbst durchführen, sondern lediglich bestimmte Voraussetzungen hierfür schaffen kann. Das wird ihm am besten gelingen, wenn er rechtzeitig die Hilfe des vertrauenswürdigen Archivs in Anspruch nimmt, dem er seine Unterlagen einmal übergeben will<sup>8</sup>.

Für uns Archivare öffnet sich hier vielleicht ein neues Tätigkeitsfeld. Neben der Beratung der schriftgutabgebenden Verwaltungen, für die wir zuständig sind, werden wir in Zukunft Nachlässe nicht nur einwerben, sondern auch bei der Entstehung und Organisation von deren elektronischen Teilen unser - hoffentlich ausreichendes - Fachwissen zur Verfügung stellen.

## **Summary**

The difficulties preserving digital records over the long time are well known. Recommendations and proposals solving these problems have been widely discussed in the framework of institutions like archives, libraries and museums, but not for private collections built up by single researchers working outside of such frameworks. Using two fictitious examples - collections from a historian and from a genealogist - it will be shown which measures the researcher has to considerate to make his collection suitable for preservation, and that the collection has to be given to a trustworthy archival institution capable to guarantee preservation and access in the future.

---

<sup>7</sup> In Deutschland versucht das Projekt *nestor*, eine Plattform für die Zusammenfassung von Kompetenzen und Fachdiskussionen zum Thema Langzeitarchivierung aufzubauen, vgl. <http://www.langzeitarchivierung.de>. Eine eigene nestor-Arbeitsgruppe beschäftigt sich mit der Zertifizierungsfrage.

<sup>8</sup> Vgl. hierzu aus der Sicht einer Universitätsbibliothek bzw. eines Universitätsarchivs: Nancy Deromedi, "When you collect the treasures of learning and learned men, you cannot fail of a university": documenting faculty in the University Archives at the University of Michigan, und Thomas Hyry - Christine Weideman, American appraisal theories and practices - a view from a private archive. Vorträge auf dem 15. Internationalen Archivkongress in Wien 2004, erreichbar über <http://www.wien2004.ica.org>. Für diesen Hinweis danke ich meinem Kollegen Lothar Saupe.