

Konzept zur Archivierung von digitale Unterlagen im Bischöflichen Archiv St. Gallen

von Stefan Kemmer, kemmer@bistum-stgallen.ch

Stand 22.9.2008

Problematik

Das Problem der Archivierung digitaler Akten erreicht kleine und mittlere Archive mit einer gewissen Verspätung. Noch geht es nicht um komplett digitalisierte Arbeitsabläufe, sondern um wenige, nicht mehr auf Papier druckbare Materialien.

Doch auch die Archivierung dieses kleinen Teils stellt eine grosse Herausforderung dar, da mit zusätzlichen finanziellen und personellen Ressourcen nicht gerechnet werden kann.

Das vorliegende Konzept zeigt den Umgang mit digitalen Archivalien im Bischöflichen Archiv St.Gallen. Dieser ist bewusst einfach konzipiert und soll mit akzeptablem Aufwand den Weg in die digitale Zukunft offen halten. Die Anforderungen an eine Langzeitarchivierung werden damit zwar noch nicht ganz erfüllt, aber das Archivgut geht wenigstens auf mittlere Frist nicht verloren.

Struktur des digitalen Archivs

Digitales Archivgut

Da der Aufwand zur Pflege eines digitalen Archivs hoch ist, sollte es möglichst klein bleiben. Alles was sich sinnvollerweise ausdrucken lässt, wird weiterhin konventionell archiviert. Dazu gehören neben Textdokumenten aller Art auch E-mails, die Buchhaltung, das Personalverzeichnis und sonstige, mittels Datenbank generierte Listen.

Der Inhalt des digitalen Archivs lässt sich in drei Gruppen unterteilen, für die der Handlungsbedarf verschieden gross ist:

1. Material, das ausschliesslich in digitaler Form vorliegt, z. B. Fotos, Audiodateien von CDs oder aus dem Internet, Webseiten.
2. Material, das auf analogen Datenträgern vorliegt, aber – je nach Beschaffenheit des Trägermaterials – früher oder später digitalisiert werden muss, z.B. Audiokassetten, Videobänder, Schmalfilme, Schallplatten. Die alten Datenträger werden nicht vernichtet, sondern als Sammlungsgut weiter aufbewahrt, um ggf. vom technischen Fortschritt profitieren zu können.
3. Material, das auf analogen Datenträgern vorliegt, dessen Nutzbarkeit in digitaler Form aber grosse Vorteile verspricht, z. B. Findmittel aller Art, sekundäres Informationsmaterial, das den Zugang zu den Beständen erleichtert.

Physikalische Ordnung, Speicherplatz

Beim zur Aufbewahrung nötigen Speicherplatz wirken sich die vergleichsweise geringen Mengen günstig aus: die 30'000 Dateien belegen derzeit mit 20 GB nur den zehnten Teil einer Festplatte. Auch in Zukunft wird daher der Speicherplatz dank technischem Fortschritt wohl kein Problem darstellen.

Diese Daten werden im Rahmen der Kanzleitätigkeit täglich, wöchentlich und quartalsweise auf Band gesichert (gemäss Reglement). Um aber vom Archiv her eine gewisse Kontrolle über die Sicherungen zu behalten, werden zusätzlich wöchentlich bis monatlich Kopien auf eine externe Festplatte gespeichert (je nach Intensität der Arbeiten am Archiv). Eine zweite externe Festplatte (von einem anderen Hersteller) erhält quartalsweise eine Kopie des Archivs und wird ausser Haus aufbewahrt.

Logische Ordnung

Das Archivgut wird in einer Ordnerstruktur abgelegt, die in sich nicht verändert werden darf, aber unabhängig vom konkreten Speicherort (etwa einem Laufwerk) bleibt. Diese unterscheidet auf der obersten Ebene zwischen Archivalien und Findmitteln und wird weiter unten anhand von Beispielen für das jeweilige Archivgut erklärt. Über alle Archivalien werden md5-Prüfsummen erstellt, mit deren Hilfe auch die Sicherungskopien kontrolliert werden.

Formate

Die Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen KOST bietet eine umfassende und detaillierte Bewertung von Formaten. Die Konsequenzen daraus sind für das Bischöfliche Archiv:

- Das ursprüngliche Format (jpg, mp3) wird bei Fotos und Audiodateien beibehalten.
- Formate von Digitalisierungen orientieren sich am bereits vorhandenen Material.
- Für Findmittel sind nur Textdateien erlaubt (.html, xml; txt).
- Ergänzendes Sammlungsgut („graue Literatur“) wird bereits als pdf übernommen, oder, bei anderen Formaten, in pdf/A konvertiert.

Digitale Fotos

Fotos werden zu einem bestimmten Anlass aufgenommen, lassen sich also gut nach Aufnahmedatum ablegen und nach Aufnahmejahr gruppieren. Dabei ist das Aufnahmedatum vom Erstellungsdatum der Datei zu unterscheiden und muss aus den EXIF-Metadaten ausgelesen und anhand anderer Akten überprüft werden.

Die Fotos erhalten als Signatur eine 5-stellige Nummer (numerus curens), ergänzt durch die laufende Nummer innerhalb der Serie. Dies entspricht dem Umgang mit analogen Fotos.

Muster: . /Archivalien/Fotos/Fotos[Jahr]/[Datum jjjjmmtt, Anlass]/[Signatur]

Beispiel: . /Archivalien/Fotos/Fotos2007/20070403_Chrisammesse/11279_10.jpg

Alle Fotos werden mit Metadaten im IPTC-Bereich versehen. Belegt werden dabei folgende Felder:

- Copyright Notice: der ISIL-Code des Archivs (CH-000245-0) gemäss SNB
- Object Name: der Dateiname, gleichzeitig die Signatur
- Caption/Abstract: Stichwort zum Anlass, bei einzelnen Fotos Namen von Personen

Auf detailliertere Beschriftungen muss aus Zeitmangel verzichtet werden. Da zu den fotografierten Anlässen normalerweise auch Akten anfallen, könnten viele Angaben auch zu einem späteren Anlass noch nachgetragen werden.

Digitale Audiodaten

Ein Grossteil der Audiodateien gelangt im Format mp3 aus dem Internet oder von Daten-CDs ins Archiv. Vereinzelt Audio-CDs werden konvertiert, um einfacher auf sie zugreifen zu können. Digitalisiert werden darüber hinaus alle Kassettenbänder. Diese sind meist von bescheidener Qualität, oft benutzt und bereits jetzt nicht mehr von allen Geräten abspielbar.

Die Digitalisierung von Schallplatten ist dagegen nicht ganz so dringlich. Dank ihrer vergleichsweise robusten Qualität und der schlichten Ansprüche an ein Lesegerät sind sie vermutlich länger zugänglich, als Kassetten.

Beim Inhalt handelt es sich nicht um Musikaufnahmen, sondern um Interviews, Reportagen oder andere Radiobeiträge von Mitarbeitern des Bischöflichen Ordinariates. Ist die Verstehbarkeit gewährleistet, können daher Abstriche bei der Aufnahmequalität in Kauf genommen werden.

Audiodateien werden nach Sendedatum (bei Radiobeiträgen) oder Erstellungsdatum abgelegt und erhalten eine Dossiersignatur:

Muster: ./Archivalien/Audio/Audio[Jahr]/[Datum jjjjmmtt]/[Signatur]

Beispiel: ./Archivalien/Audio/Audio1998/19980614/Y_8_5_g.mp3

Folgende Metadaten werden in den Dateien hinterlegt:

- Interpret: der ISIL-Code des Archivs
- Titel: der Dateiname, gleichzeitig die Signatur
- Kommentar: Stichwort zum Inhalt

Darüber hinaus werden für Metadaten separate Dateien akzeptiert, die ebenfalls als Dateinamen die Signatur führen und im gleichen Verzeichnis abgelegt sind.

Ein umfangreiches Beispiel für eine digitalisierte Audiokassette sähe etwa so aus:

- Y_8_5_r_SeiteA.mp3 Inhalt der Kassette, Seite A
- Y_8_5_r_SeiteB.mp3 Inhalt der Kassette, Seite B
- Y_8_5_r.jpg Eingescanntes Cover
- Y_8_5_r.txt sonstige, nicht weiter strukturierte Metadaten
- Y 8,5 r die Audiokassette selbst

Digitale Videodaten

Videos gelangen nur selten ins Archiv und werden in MPEG-4 (H.264) konvertiert. Dabei werden von Videobändern (zur Zeit nur VHS) zudem ausser Haus DVDs erstellt.

Videodaten werden nach Aufnahmejahr abgelegt und erhalten eine Dossiersignatur:

Muster: ./Archivalien/Video/Video[Jahr]/[Signatur]

Beispiel: ./Archivalien/Video/Video2005/Y_8_7_a.mp4

Webseiten

Was der Benutzer beim Aufruf von Webseiten zu sehen bekommt, entsteht im Zusammenspiel von Server, Browser und eingesetzter Hardware. Im Unterschied zu einem herkömmlichen Textdokument gibt es keine eindeutige Originalansicht. Zudem kann sich auch die innere Struktur einer Webseite ändern, etwa beim Einsatz von CMS-Systemen. Die Kopie einer Webseite ist notgedrungen nur eine Annäherung an das ursprüngliche Objekt und unter Umständen nicht vollständig. Hinzu kommt, dass viele der archivierten Webseiten bereits Fehler aufweisen, da sie nicht professionell erstellt wurden.

Gesammelt werden Webseiten des Bistums und seiner Institutionen (Arbeitsstellen), indem je nach Wichtigkeit ein oder mehrmals jährlich eine komplette Kopie erstellt wird.

Diese wird nach Kopiedatum abgelegt und erhält eine Dossiersignatur:

Muster: ./Archivalien/Homepages/[domain]/[Kopiedatum]/[Signatur]

Beispiel: ./Archivalien/Homepages/www_bistum-stgallen_ch/20070101/Y_1_2_a/index.html

Ebenfalls archiviert wird das Logfile, aus dem technische Daten und die beim Kopiervorgang aufgetretenen Probleme hervorgehen.

Findmittel

Für die Erschliessung und Präsentation des digitalen Archivs bieten die im Internet bereits bewährten Techniken grosse Vorteile:

- Findmittel können untereinander und mit den Archivalien unabhängig vom physikalischen Speicherort miteinander verlinkt werden.
- Da html-Dateien nur Text ohne Formatierungen enthalten, bleiben die Findmittel klein und

- bereits mit einfachen Editoren für menschliche Leser gut zugänglich.
- Das Layout kann an zentraler Stelle in einem Stylesheet definiert werden, so dass man sich auf wenige HTML-Befehle beschränken kann.
- Normierungen und Validatoren garantieren eine gewisse Qualität bei der Kodierung.

Alle Dateien, die im weitesten Sinn zur Erschließung des digitalen und des konventionellen Archivs dienen, können im Ordner Findmittel abgelegt werden, z.B:

Übersicht Fotos: ./Findmittel/Fotos/Fotos.html
Übersicht Audio: ./Findmittel/Audio/audio.html
Notizen zur SBK: ./Findmittel/institutionen/SBK.html

Abgelegt wird bei den Findmitteln auch Material, das nicht in das digitale Archiv gehört, aber viel Arbeit spart. So kann man eine als pdf-Datei gespeicherte Druckschrift schnell einsehen und mit der korrekten Signatur zitieren, ohne das Original aus dem Schrank suchen zu müssen.

Technische Hilfsmittel, Software

Auf den Einsatz einer Archivierungssoftware wird wegen der zu erwartenden Versionsprobleme bewusst verzichtet. Für die Bearbeitung des Digitalen Archivs wird nur freie, kostenlose Software verwendet, die aber jederzeit durch andere Produkte ersetzt werden kann.

Bilder:

- irfanview (Bildbearbeitung, Beschriftung)
- md5summer (Prüfsummen erstellen und kontrollieren)

Audiodatei

- No.23 Recorder (Aufnahme in .mp3)
- mp3tag (Beschriftung in der .mp3-Datei)

Videodateien

- Handbrake (Konvertieren von DVD in MPEG-4)

Homepage

- HTTrack Website Copier (Homepage speichern)

Findmittel

- Programmers Notepad (Codierung von html)
- A-PDF INFO Changer (Beschriftung im PDF ändern)
- pdfCreator (pdf/A aus allen Anwendungen erstellen)
- Easy Thumbnails (generiert thumbnails von Fotos)