



HTW Chur

Schweizerisches Institut für
Informationswissenschaft

Auswertung der Umfrage

Bedarfsanalyse für ein Angebot „Digitale Langzeitarchivierung“ in den Geisteswissenschaften (data repository)

- Zusammenfassung -

Version: 1.0

Datum: 31.08.2008

Auftraggeber:

Schweizerische Akademie für Geistes- und Sozialwissenschaften SAGW
Vertreten durch Herrn Dr. Markus Züricher
Bern

Projektleitung SAGW: Dr. Beat Immenhauser

Ausgeführt durch:

Hochschule für Technik und Wirtschaft Chur
Schweizerisches Institut für Informationswissenschaft (SII)
Chur

Projektleitung SII: Prof. Dr. Hans-Dieter Zimmermann (Hans-Dieter.Zimmermann@htwchur.ch)
Projektmitarbeit SII: Joachim Pfister, M.A. (Joachim.Pfister@htwchur.ch)

1 Ergebnisse

1.1 Willkommen in Digitalien!

Digitale Daten liegen in fast allen befragten Institutionen vor (89%). In den nachfolgenden Kapiteln wird untersucht, wie mit diesen Daten in der Vergangenheit umgegangen wurde, was aktuell mit ihnen geschieht und was in Zukunft damit aus Sicht der Umfrageteilnehmer passieren könnte.

1.2 Digitale Vergangenheit

Die Problematik, auf früher erstellte digitale Datenobjekte nicht mehr zugreifen zu können, erlebten knapp 20% der Befragten. Der Umfang der verlorenen Daten war recht unterschiedlich und den Befragten fiel es schwer, Schätzungen abzuliefern. Die Angaben lagen zwischen einzelnen Dateien (z.B. ein Dutzend Word-Dokumente) im Megabyte-Bereich, einer Anzahl von Datensätzen (ca. 10000 Stück) oder sogar im Gigabyte-Bereich. Hierbei ist schwierig abzuschätzen, welchen Umfang eines der genannten Dutzend Word-Dokumente hatte. Handelte es sich um kurze Dokumente oder jeweils komplette Bücher? Für eine weiterführende Untersuchung dieser Problematik stellt sich die Frage, ob und wie man den Umfang „normalisieren“ kann, um den eingetretenen Verlust im Vergleich besser einschätzen zu können.

Die Gründe, die zum Verlust der Daten führten, sind vielfältig. Als häufigste Ursache (50% der Befragten¹) wurde angegeben, dass die Software zum Zugriff auf die Daten nicht mehr läuft (z.B. auf Grund von Inkompatibilitäten zum aktuellen Betriebssystem oder da sonstige Software-Voraussetzungen fehlen). An zweiter Stelle mit gleichen Häufigkeiten (35%) liegen die Probleme veralteter Hardware (z.B. Diskettenlaufwerke), die nicht mehr mit aktueller Hardware kompatibel sind und dass die Software zum Zugriff auf die Daten nicht mehr vorhanden ist. An dritter Stelle der Problem-Hitliste liegen mit 15% fehlende Zugangsdaten (Passwort/Nutzername-Kombination). Weniger häufig trat das Problem auf, dass Hersteller zur Sicherung ihrer Softwareprodukte ihre Lizenzserver oder Online-Überprüfungen nicht mehr betreiben, und somit der ursprünglich genutzten Software den Zugriff auf die Daten verweigern (12%). Am wenigsten häufig wurde ein Defekt der Hardware angeführt, die zum Zugriff auf die Daten notwendig gewesen wäre (nur 3.5%).

In den Kommentaren finden sich Highlights, wie z.B. dass durch die Umstellung des Websystems Daten verloren resp. durch Nachlässigkeit gelöscht wurden. Oder aber, dass sämtliche alte Hard- und Software entsorgt wurde.

1.3 Digitale Gegenwart

1.3.1 Bestandesaufnahme

Von den 149 Befragten gaben 132 (d.h. 89%) an, dass digitale Daten vorhanden sind, die sie in den nächsten 5-7 Jahren verwenden werden (zu 98%) und die gesamthaft als relevant für die Öffentlichkeit eingeschätzt wurden (95%; SAGW-Institutionen: 100%). Mehrheitlich (81%²) handelt es sich dabei um Digitalisate. Fast genauso häufig trifft man auf von Anfang an digital erstellte Inhalte (73%). Mittels OCR entstanden bei 42% der Umfrageteilnehmer digitale Daten. Datenbanken werden von knapp 70% der Befragten eingesetzt. Die digitale Textedition wird weitaus seltener durchgeführt (28%). Bei den SAGW-Institutionen ist anzumerken, dass Datenbanken häufiger als bei den restlichen Befragten eingesetzt werden und dass die Objekte von Anfang an bereits Digital vorliegen.

¹ bezogen auf diejenigen, bei denen ein Verlust aufgetreten ist (Teilmenge der Befragten mit n = 26)

² In diesem Kapitel beziehen sich die Angaben auf diejenigen Befragten, die das Vorhandensein digitaler Daten angaben (n=132).

In den Kommentaren wurde bemängelt, dass nicht klar war, ob sich die Angaben auf einen Forscher oder für ein ganzes Institut beziehen sollen. Die Befragten wurden aufgefordert, die bei Ihnen vorliegenden digitalen Inhalte kurz zu Beschreiben. Die Antworten finden sich in Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.**

Auf die Frage, welche Formate vorliegen, herrschte durchsuchbarer Text (82%) vor, gefolgt von Bildern (63%) sowie Bildern, die Text beinhalten (62%) und die somit per OCR-Verfahren weiter erschlossen werden könnten. Weniger häufig liegen Audio- (23%) und Videodateien (14%) vor. An sonstigen Dateien wurden in den Kommentarfeldern Datenbanken (Access, Filemaker, MySQL), Statistik-Daten (SPSS) oder digitale Zeichnungsdaten für Pläne (AutoCAD, ArchiCAD) oder auch Daten für GIS (Geographisches Informationssystem) genannt.

Die überwiegende Mehrheit (58%) gab an, relationale Datenbanken zu verwenden. Andere als relationale Datenbanken werden von knapp 33% der Umfrageteilnehmer verwendet, wobei die dort angegebenen Softwarelösungen (Bilddatenbanken, Records Management Systeme) oder Anwendungsbereiche (eigene Projektarbeit mit Online-Bibliotheken oder Datenbanknutzung assoziierter in- und ausländischer Projekte) ein recht heterogenes Bild geben. Die Frage, was „Datenbanken“ sind hätte in der Umfrage präzisiert werden müssen, um Mehrdeutigkeiten auszuschliessen (z.B. „Welche Softwareprodukte nutzen Sie? Welches Datenbankmanagementsystem nutzen Sie?“).

Zum Erstellen und Bearbeiten digitaler Daten werden mehrheitlich zu 70% Standardanwendungen (Microsoft Office, Open Office etc.) eingesetzt. Webapplikationen werden nur knapp von einem Viertel (24%) verwendet (bei den SAGW-Institutionen ist dies mit 36% deutlich ausgeprägter). Die Liste mit den ansonsten verwendeten Programmen findet sich in Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.**

Die gespeicherten Volumina betragen nach Schätzungen der Teilnehmer:

- im Megabyte-Bereich 9 Nennungen zwischen 10-325 MB,
- im Terabyte-Bereich 10 Nennungen zwischen 1-15 TB,
- im Gigabyte-Bereich insgesamt 33 Nennungen zwischen 1-1000 GB.

Metadaten werden bei knapp der Hälfte aller Teilnehmer erfasst (49%; bei den SAGW-Institutionen waren dies 64%). Dabei werden anerkannte Metadaten-Standards nur zu einem Drittel (34%) verwendet; unwesentlich mehr werden nicht-erkannte Standards verwendet (35%). Zugleich gaben die Befragten ebenfalls zu knapp einem Drittel an (31%), dass sie es nicht wissen. Dies deutet darauf hin, dass bezüglich der Nutzung von Metadaten eher ein diffuses Bild herrscht. In den Kommentaren spiegelt sich dieser Befund wieder: „J'ignore la fonction des métadonnées et je ne les utilise pas“.

Die Frage, ob zusätzlich zu den Metadaten aussagekräftige und administrative Beschreibungen vorhanden sind, um die Daten zukünftig sachlich korrekt zu interpretieren, wurde von der Mehrheit mit 47% bejaht (bei den SAGW-Institutionen sogar deutlich mit 73%). Verneint wurde diese Frage von 28% und fast ebenso häufig (25%) wurde mit „Weiss nicht“ beantwortet. Ein einheitliches Vorgehen besteht nicht, was manchmal sogar zu Problemen führt, wie sich aus den Kommentaren herauslesen lässt³.

1.3.2 Aktuelle Aktivitäten zur Erhaltung und Nutzung digitaler Daten

Auf die Frage, wer aktuell für die Archivierung digitaler Daten zuständig ist, stellt sich das Bild der Zuständigkeiten recht heterogen dar. Mehrheitlich handelt es sich um permanent angestellte Mitarbeiter (39%; bei den SAGW-Institutionen sogar 64%). Temporär Beschäftigte sind immerhin zu

³ • „die Datenbank ist selbst programmiert und nur unzulänglich dokumentiert“

• „Je nach Projekt; für "Uesserschwyz" vorhanden, nicht für alle andern.“

• „Jeder Mitarbeiter pflegt hauptsächlich seine persönlichen Daten. Probleme entstehen beim Mitarbeiterwechsel.“

knapp einem Drittel (32%) mit dieser Aufgabe betraut. Als verantwortlicher Akteur wird die Projektleitung (33%) betrachtet. Eine interne Organisationseinheit wie z.B. das Rechenzentrum wird von 26% als verantwortlich erachtet, während externe Institutionen oder Dienstleister nur zu 12% angegeben werden (von den SAGW-Institutionen sogar 3 mal häufiger mit 36%). In den Kommentaren wird deutlich, dass diese Umstände weniger zufriedenstellend sind (siehe Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.**), z.B.: dass dies nur im Rahmen der Projektförderung gemacht werden kann oder dass die Organisations-IT von Fragen zur Langzeitarchivierung keine Ahnung hat. Sicherungskopien der digitalen Objekte werden fast überall erstellt (92%, bei den SAGW-Institutionen zu 90%). Knapp 8 % aller Institutionen geben an, keine Sicherungskopien zu erstellen.

Die digitalisierten Daten werden überwiegend für interne Zwecke z.B. durch die Teammitglieder genutzt (67%; bei den SAGW-Institutionen sogar deutlich mehr mit 81%). Eine vollumfänglich öffentlich zugängliche Publikation wird von 36% der Befragten als aktuelle Nutzung angegeben, während eine ausschnittsweise Publikation von 44% der Teilnehmer im Moment genutzt wird. Die SAGW-Institutionen zeigen sich hinsichtlich der Publikation der Datenobjekte deutlich zurückhaltender als die Nicht-SAGW-Institutionen (nur 27% und 36%).

Nach Projektende geben 41% der Umfrageteilnehmer an (SAGW-Institutionen: 46%), dass die Daten mit Weiterbetreuung bei ihnen bleiben. Ohne Weiterbetreuung verbleiben fast ähnlich viele digitalen Objekte (35%; SAGW-Institutionen deutlich geringer mit 18%). Nur knapp 13% gaben an, dass die Datenobjekte extern gesichert werden (SAGW-Institutionen: 18%). Die Frage der Weiterbetreuung ist bei gut einem Viertel (26%) noch offen (bei den SAGW-Institutionen deutlich mehr bei 36%). In den Kommentaren wird vielfach das hochschuleigene Rechenzentrum bzw. der dortige IT-Dienstleister als möglicher Partner benannt. Die Antworten in den Kommentaren zu dieser Frage sind sehr unterschiedlich und bewegen sich zwischen dem Spektrum von fehlendem Problembewusstsein („sind auf meinem Rechner gespeichert“), der Erkenntnis der Problematik („Es hängt viel von der SAGW-Initiative ab“), teilweise gekoppelt mit Ohnmacht („... nicht eine Frage der Infrastruktur, sondern der Finanzierung von Personal“) bis hin zu konkreten Projekte, die angegangen werden.

1.3.3 Publikations-, Informations- und Wissensverbreitungsverhalten

Die Umfrageteilnehmer äusserten mehrheitlich (67%), dass sie ihre digitalen Datenbestände nur mit Einschränkungen Anderen zur Verfügung stellen wollen. Ohne Einschränkung waren es wesentlich weniger (25%; SAGW-Institutionen zu 36%). Vor allem bei Bildern wurde in den Kommentaren auf die bestehenden Bildrechte als Schwierigkeit hingewiesen. Ebenso waren häufige Nennungen in den Kommentaren Probleme mit dem Datenschutz z.B. bei Interviews, deren Inhalte Persönlichkeitsrechte tangieren könnten und daher den Status vertraulich inne haben.

Gemäss den Selbstauskünften der Umfrageteilnehmer werden bereits vorliegende Forschungsergebnisse zu gut drei Vierteln mit Anderen geteilt (75%). Dabei werden folgende Kommunikationskanäle hauptsächlich genutzt (in absteigender Reihenfolge): E-Mail (50%), portable Medien wie USB-Sticks(40%), Netzlaufwerke (23%; SAGW-Institutionen zu 36%) oder auf dem Postweg (12%; SAGW-Institutionen zu 27%).

Internationale Datenverbände spielen nur bei 9% der SAGW-Institutionen eine Rolle, bei den restlichen Befragten belief sich dies auf 3% (EUROCLIMHIST, Fundmünzen D-A-CH, Memobase, „Vita Regularis“ der Berlin-Brandenburgischen Akademien der Wissenschaften). Nationale Datenverbände nutzen 4% der Umfrageteilnehmer (unter den Nennungen der Dienste waren bspw. www.ortsnamen.ch, www.nebis.ch, EuropeanArt.net, SIKART). 4% der Umfrageteilnehmer gaben an, ein nationales Datenzentrum zu nutzen. Ein Teilnehmer gab an, ein disziplinspezifisches, internationales Datenzentrum in Anspruch zu nehmen („Vita Regularis“ der Berlin-Brandenburgischen Akademien der Wissenschaften). Andere Wege des Übermittels von digitalen Forschungsdaten werden von 14% der Umfrageteilnehmer verwendet (beispielsweise Websites, „das Internet“, Wiki).

1.3.4 Motivation für oder gegen das Teilen von digitalen Objekten

1.3.5 Motivation pro Teilen

Generell wurde folgenden Aussagen zugestimmt (in abnehmender Zustimmung):

- Durch gemeinsames Teilen erfolgt ein Erkenntnisgewinn, der für die gesamte Disziplin wichtig ist. (56%)
- Andere sollen Querbezüge zu meinen/unseren Daten herstellen können, um so zur Vernetzung der Daten und letztlich des Wissens beizutragen. (51%)
- Interdisziplinäre Zusammenarbeit wird dadurch einfacher. (41%)
- Die Daten wurden mit öffentlichen Mitteln erstellt und sollen der (forschenden) Öffentlichkeit daher zugänglich sein. (39%)
- Erhöhte nationale Wahrnehmung der eigenen Tätigkeit. (36%)
- Neue Dateninterpretationen und -analysen (z.B. in Form von Visualisierungen oder Meta-studien) werden dadurch ermöglicht. (weniger häufig, nur 30%)
- Erhöhte internationale Wahrnehmung der eigenen Tätigkeit. (26%)

Generell wurden die folgenden Aussagen eher abgelehnt:

- Ist aus einer externen Verpflichtung erwachsen (z.B. durch Geldgeber etc.) (28%)

Weder positiv, noch negativ (→ deutet auf Unkenntnis resp. ungenaue oder schlecht verständliche Fragestellung hin):

- Eine Übertragung der Wiki-Idee (Web 2.0 als „Mitmach-Web“) mit der Möglichkeit, Daten zu kommentieren, zu bewerten und zu neuen Anwendungen zu kombinieren, ist für eigene Daten, resp. die eigene Disziplin, grundsätzlich denkbar.

1.3.6 Motivation contra Teilen

Die Prozentwerte zu den Hinderungsgründen belaufen sich alle im einstelligen Prozentbereich. Daraus kann geschlossen werden, dass sie für die befragten Institutionen nicht als solche erachtet werden.

1.3.7 Kompetenzen zur Digitalisierung und Langzeitarchivierung

Das Vorhandensein von Kompetenzen zur Digitalisierung innerhalb der Institution wurde von 49% aller Befragten bestätigt (SAGW-Institutionen sogar 63%). Im Gegensatz dazu steht das Vorhandensein von Kompetenzen zur Langzeitarchivierung, das nur zu 38% vorhanden ist. Dies spiegelt sich auch in der erhöhten Bereitschaft der Umfrageteilnehmer wieder, diese Dienste von extern zu beziehen. Für die Digitalisierung ist die Häufigkeit geringer (10%; SAGW-Institutionen 20%), als für die Langzeitarchivierung (20%, SAGW-Institutionen mit deutlich höheren 45%). Dies geht mit der Beobachtung einher, dass in 44% aller Institutionen generell keine Kompetenzen zur Digitalisierung oder Langzeitarchivierung innerhalb der Organisation vorhanden sind. Bei den SAGW-Institutionen war dieser Anteil kleiner (36%). Ein geringer Anteil (13%) gab an, es nicht zu wissen.

Die Kommentare zu dieser Frage liefern ebenfalls ein ganzes Spektrum von Beobachtungen zwischen Hilflosigkeit („Die Archivarin war zwar schon am NESTOR Seminar, fühlt sich aber überfordert von dieser Arbeit“), Bewusstsein um diese Problematik („Längerfristig ja, die Frage ist aber noch nicht aktuell.“) bis hin zum Hinweis, dass Digitalisierungs- und Archivierungsprojekte teuer sind und eine finanzielle Unterstützung ideal wäre.

Bei den Befragten stossen potentielle Weiterbildungsangebote bei der Digitalisierung und Langzeitarchivierung auf grösseren Anklang als Ausbildungsangebote in diesem Bereich (Ausbildungsangebote immer so um 10% geringer; bei den SAGW-Institutionen sogar um 20%). Dennoch überwiegt mehrheitlich bei den Umfrageteilnehmern die Haltung, dies noch nicht beurteilen zu können. Eine Ausnahme bildeten die SAGW-Institutionen, die diese zögerliche Haltung nicht teilen (nur zu 9%, d.h. jeweils ein Befragter).

Das Interesse zur Ausbildung an resp. mit einem digitalen Forschungswerkzeug war wesentlich geringer, als bei der Digitalisierung und Langzeitarchivierung (20%; SAGW-Institutionen jedoch überwiegend positiv dazu eingestellt mit 55%). Die mehrheitliche Haltung mit „Weiss nicht“ aller Befragten (45%; bei den SAGW-Institutionen wiederum geringer ausgeprägt mit nur 18%) ist daher nicht verwunderlich.

1.4 Digitale Zukunft

1.4.1 Geplante Aktivitäten

Die geplante Digitalisierung von Text- und Bilddaten steht im Zeithorizont der nächsten 1-2 Jahren hoch im Kurs (Textdokumente 53%, gefolgt von Bilddokumenten mit 42%; die SAGW-Einrichtungen erachten die Bilddokumente gleich wichtig, wie die Textdokumente). Audio und Videodaten stellen weniger begehrte Digitalisierungsobjekte dar, denn zu 45% (44% für Video) werden geplante Aktivitäten in diesem Segment verneint. Als weitere Daten wurden digitale Kartendaten benannt sowie Noten.

Die Mehrheit der Nicht-SAGW-Institutionen gibt an, keine Langzeitarchivierungsprojekte in den nächsten 1-5 Jahren zu starten. Wenn Langzeitarchivierungsprojekte dennoch in einem Zeithorizont der nächsten 1-2 Jahre projiziert sind, widmen sie sich vornehmlich Textdaten (30%), Datenbanken (28%) und Bildern (23%). Bei den SAGW-Institutionen werden dagegen häufiger Langzeitarchivierungsprojekte in den nächsten 1-2 Jahren angegangen, wobei die Medien Bilder, Textdokumente und Datenbanken allesamt gleichauf liegen (36%).

Generell herrscht die Bereitschaft vor, zukünftige digitale Datenbestände mit anderen zu teilen, jedoch nur mit Einschränkungen (83%; SAGW-Institutionen 91%). Ohne Einschränkungen wurde dies von 19% (SAGW-Institutionen: 9%) bejaht. Abgelehnt resp. nicht gewusst wurde dies von verschwindend geringen 5% (keine Angabe von SAGW-Institutionen).

1.4.2 Wünsche

Innerhalb der Umfrage sollte der Wunsch der Institutionen ermittelt werden, ein quelloffenes, standardisiertes Forschungswerkzeug (Bilddatenbank, Editionswerkzeug) einzusetzen. Diese Frage wurde deutlich bejaht (62%; SAGW-Institutionen 72%). Enthaltungen waren mit einem Drittel (33%; SAGW-Institutionen 27%) zu verzeichnen, wohingegen nur 13% diese Idee klar ablehnten (SAGW-Institutionen niemand). In den Kommentaren kam der Zweifel zum Ausdruck, ob so ein generelles Werkzeug den individuellen Bedürfnissen Rechnung tragen kann. Ausserdem wurde die Frage als zu unspezifisch erachtet.

Generell war für die Umfrageteilnehmer denkbar, die digitalen Daten kollaborativ zu er- oder bearbeiten (61%). Die Fraktion der Unentschlossenen beläuft sich auf 29%. Dieser Idee ablehnend gegenüber stehen 18% der Befragten (SAGW-Institutionen nur 9%). In den Kommentaren wurde mehrfach angemerkt, dass die Frage unverständlich war – die kollaborative Idee des Er- und Bearbeitens von Inhalten hätte evtl. anhand eines konkreten Beispiels verständlich gemacht werden sollen. Dies könnte auch ein Indiz dafür sein, dass bei manchen Befragten in deren Disziplin dies (noch) nicht möglich resp. üblich ist.

Auf die Frage, ob eine Unterstützung bei der Planung und/oder Durchführung von Langzeitarchivierungs-Aktivitäten gewünscht ist, antworteten 30% mit nein (SAGW-Institutionen niemand). Vor allem ist technische Unterstützung für die tatsächliche Durchführung gefragt (58%; SAGW-Institutionen: 100%). Eine organisatorische Unterstützung wird von 52% begrüsst (SAGW-Institutionen: 90%). Die Hälfte (50%; SAGW-Institutionen sogar zu 73%) haben Interesse an Informationen zu Metadaten-Standards. Breiten Zuspruch fand der Wunsch, finanzielle Unterstützung zu erhalten (58%; SAGW-Institutionen: 82%).

1.4.3 Externe Dienstleister

Die Bereitschaft, einen externen Dienstleister für die digitale (Langzeit-)Archivierung einzusetzen wurde von der Mehrheit nur unter Einschränkungen begrüsst (57%; SAGW-Institutionen 72%). Ohne Einschränkungen war dies nur ein minimaler Teil (14%; SAGW-Institutionen 9%). Klar abgelehnt wurde diese Idee von 19% (niemand aus den SAGW-Institutionen). Unentschlossen waren 23%. Für das Angebot eines externen Dienstleisters (Speicherung, Erhaltung der Zugänglichkeit) eine Gebühr zu entrichten, wurde ohne Einschränkung von den wenigsten gut geheissen (7%; SAGW-Institutionen: 18%). Die Mehrheit gab an, dies nur bei einer separaten Förderung, die nicht zu Lasten des Projekt-Budgets geht, zu akzeptieren (46%; SAGW-Institutionen: 73%). Eine ablehnende Haltung zur Entrichtung von Gebühren äusserten fast ähnlich viele Umfrageteilnehmer (40%; niemand aus den SAGW-Institutionen). Unentschlossen zeigten sich in dieser Frage nur 17%. Die Kommentare zu dieser Frage gehen in die Richtung, dass der Unterhalt der Daten eine öffentliche Aufgabe ist und nicht von privaten Dienstleistern erbracht werden sollten („Im Grunde wäre das ein typisches Akademie-Projekt“).

1.4.4 Schaffung eines Dienstleistungsangebots

Auf diese Frage („Würden Sie die Schaffung eines Dienstleistungsangebots im Bereich Langzeitarchivierung digitaler Daten in den Geisteswissenschaften begrüssen?“) antwortete die überwältigende Mehrheit mit ja (75%; SAGW-Institutionen: 91%). Die Schaffung nur mit Einschränkungen befürworteten 26% (SAGW-Institutionen: 9%). Ablehnend standen nur 3% der Befragten diesem Ansinnen gegenüber (niemand aus den SAGW-Institutionen). 6% konnten es nicht sagen.