
Josef Riegler

Archiv (ohne) Grenzen?

Bemerkungen zu Archivportalen, Open Government, Web 3.0 und der Entwicklung von Archiven¹

Einleitend einige Fragen zum Thema: Wo liegen heute die Grenzen der Archive? Physisch gibt es diese Grenzen nach wie vor, unsere Archivalien sind Unikate. Gibt es diese Grenzen im Nachweis noch oder sind wir in der Entwicklung schon einen Schritt weiter? Sind nicht durch die digitale Bereitstellung die physischen Archivgrenzen längst gefallen? Wie weit werden die Bereichsgrenzen überschritten werden? Wo werden wir als Archive im klassischen Sinne in fünf Jahren, wo werden wir in zehn Jahren stehen?

Werfen wir zunächst einen Blick auf die Entwicklung im Archivbereich seit den späten 1980er-Jahren, jener Zeit also, in der der Paradigmenwechsel in den Archiven einsetzte. Bestandsverzeichnisse und Findbücher waren bis dahin in Einzelfällen im Druck veröffentlicht worden. Verzeichnisse vieler Archivbestände waren nur im Lesesaal einzusehen, durften mitunter auch bei völlig freien Beständen nur in Ausnahmefällen kopiert werden.

Erst mit dem Vordringen des Internets in die Archive ab Mitte der 1990er-Jahre setzte eine bis heute anhaltende Entwicklung zur Öffnung der Archive ein. Archive begannen, zunächst vereinzelt, sich als Informationsdienstleister zu sehen. Das Arkanum des Archivs und seiner Bestände dürfte mit Ausnahme besonders schützenswerter Bereiche endgültig der Vergangenheit angehören. Viele Archive gingen dazu über, ihre Bestandsinformationen mehr oder weniger detailliert online verfügbar zu machen. Das Spektrum reichte und reicht noch immer von Einzelverzeichnissen in Form einer Datei, die über das Internet zur Verfügung gestellt wird, bis hin zu teilweise sehr komfortabel online benutzbaren Datenbanken, die meist mit Archivinformationssystemen verbunden sind.

Der Gedanke, das Service der Archive weiter zu verbessern, traf mit dem Wunsch der Archivbenützer nach einem möglichst vollständigen Nachweis der gesuchten Informationen in unterschiedlichen nationalen und internationalen Archiven zusammen. Damit begann eine rege „Bautätigkeit“ in der digitalen Archivwelt. Der „Portalbau“ nahm seinen Anfang und stetigen Aufschwung. Zunächst wurden in diversen Forschungsprojekten themenbezogene Archivalien in unterschiedlichen Archiven erhoben und im Internet der Zugang zu diesen „virtuellen“ Archiven geschaffen. Für bestimmte Bereiche und Forschungsvorhaben hat das bis heute seine Daseinsberechtigung und kann für den Benutzer sehr komfortabel sein.

Ich erinnere mich an so manches Gespräch im Zusammenhang mit der spartenübergreifenden Nutzung von Informationen, die in Archiven, Biblio-

¹ Text des Vortrages am Österreichischen Archivtag 2011 in Eisenstadt, ergänzt durch Quellenbelege.

theken und Museen vorhanden sind. Nicht immer konnte ich mich des Eindrucks erwehren, dass gewissermaßen ein Wettlauf darum herrschte, wer das interessantere Portal kreieren konnte. Wie gehaltvoll die darin nachweisbaren Informationen sein sollten, stand mitunter eher im Hintergrund. Die Portalprotagonisten begaben sich nur ungern in die Niederungen der Archivarbeit, wenn es darum geht, in die Portale Leben, sprich gute Daten zusammen mit den unerlässlichen Metadaten, zu bringen.

Mit dem Streben nach Portalen und den damit aus der Sicht des potentiellen Nutzers verbundenen Vorteilen verschwand mitunter das Eigentliche aus dem Fokus des Interesses: verlässlicher Inhalt mit hoher Treffergenauigkeit. Das veranlasste mich vor längerer Zeit zu folgender Äußerung: Der Portalbau ist mit dem Bau von Pagoden vergleichbar. Einige Spieler auf dem Feld der historischen Information, manchmal als Gedächtniseinrichtungen bezeichnet, haben eine mehr oder weniger große Menge digitaler Informationen. Darüber wird bildhaft eine kleine Pagode errichtet. Ideenreiche Menschen kennen mehrere solche Pagoden und bauen darüber mit gut klingenden Bezeichnungen eine neue, unter deren Dach die bisher bestehenden leicht abgefragt werden können. Bis der nächste Baumeister die Idee für eine weitere, noch größere Pagode hat. Das lässt sich vor allem topographisch noch einige Ebenen weiterspinnen, bis wir schließlich für den gesamten Globus bei einer riesigen Pagode angelangt sind. Sie denken dabei vielleicht an Google oder im europäischen Bereich, etwas näher an unserer Disziplin, an das Archivportal Apenet bzw. dessen Nachfolger.

All diese Portale haben ihren spezifischen Nutzen und Vorteil für die Archivbenützer. Sie dienen der Orientierung, geben Auskunft über Standorte und einen Überblick über die Bestände. Sucht man Archivportale, bei denen die verfügbaren Informationen weiter in die Tiefe gehen und auch orts- und personenbezogene Recherchen erlauben, dünnt das Angebot rasch aus. Noch viel dünner wird das Angebot, wenn man Sachinhalte sucht, die sich nicht im Volltext der Beschreibungsfelder ganzer Bestände, von Bestandteilen oder zu Einzelstücken finden lassen.

Damit sind wir bei einem Kernproblem archivischer Informationserschließung angelangt. Intellektuelle Sacherschließung von Archivgut ist eine anspruchsvolle und zeitintensive Aufgabe. Sie kann von Archiven wegen der stagnierenden oder auch schrumpfenden Ausstattung mit gut qualifiziertem Personal nicht im wünschenswerten Ausmaß geleistet werden.

Das ist auch der Grund, warum die meisten Archive bei der Verzeichnung von Archivgut sich auf die textliche Beschreibung der verzeichneten Inhalte beschränken. Die explizite Erschließung durch Metadaten in den Bereichen Ort – Person – Sache ist nicht sehr oft verfügbar.

Dahinter steht die Tatsache, dass reiner Text keinen weiteren Aufwand an Input für das Archiv nach sich zieht. Volltext kann einfach nach dem gewünschten Wort durchsucht werden. Die nach Beschreibungsstandards wie ISAD oder EAD angelegten elektronischen Verzeichnisse sind die Voraussetzung für das

Einbinden von Informationen einzelner Archive in beliebig große Archivportale. Wie wir wissen, hat die Suche nach Volltext viele Vorteile, aber auch einige entscheidende Nachteile. Gefunden wird der Suchbegriff nur dort, wo er zeichengenau vorkommt. An der Lösung des Synonymproblems wird schon lange gearbeitet und voraussichtlich auch noch länger gearbeitet werden müssen.

Moderne Archivinformationssysteme erlauben die Volltextsuche in bestimmten oder auch in allen Feldern, sowohl innerhalb eines Bestandes als auch global über das ganze System. Das ist bei überschaubarer Verzeichnungsmenge gut zu verwenden, die Trefferzahl ist begrenzt. Aber bei hoher Trefferzahl wird das Ergebnis rasch unüberschaubar – gefüllt auch mit Treffern, die nicht relevant sind.

Mit Archivportalen ist noch ein anderes Problem verbunden. Es gibt Portale, die zentral von den beteiligten Archiven die Metadaten der Bestände erhalten, mitunter auch die Digitalisate von Archivalien. Das ist aus der Sicht der Benutzer sehr praktisch. Aber: Wie wird zitiert? Das die Originale haltende Archiv oder die Website des Portals? Ich habe den Eindruck, dass eher das Letztere der Fall ist. Es ist ja viel einfacher, man braucht sich nicht um die korrekte Bezeichnung des Ausgangsarchivs zu bemühen. Das Ergebnis einer solchen Vorgangsweise ist, dass das die Originale haltende Ausgangsarchiv in den Hintergrund tritt und nicht wahrgenommen wird. Anders verhält es sich, wenn vom Portal nur die Suchanfrage bearbeitet und die Ergebnisse vom Ausgangsarchiv präsentiert werden.

Fassen wir vor dem Hintergrund meines Referatstitels zusammen: Die Überschreitung der Archivgrenzen durch das Einbringen unserer Informationen in Portale ist in unterschiedlicher Tiefe gelöst. Das vom Verein ICARUS betriebene Urkundenportal *Monasterium*² stellt einen Sonderfall dar. Von den Archiven der österreichischen Gebietskörperschaften sind mit Stand September 2011 nur das Niederösterreichische Landesarchiv, das Oberösterreichische Landesarchiv und das Vorarlberger Landesarchiv in einer in der Anfangsphase stehenden Portallösung vertreten, die über den Server von AUGIAS läuft³.

Open Government Data

Dieses Schlagwort geistert derzeit durch Medien, Fachzeitschriften und so manche Onlinequelle. Einige österreichische Gebietskörperschaften, wie z. B. die Städte Wien⁴ oder Linz⁵, gehören zu den öffentlichen Einrichtungen, die bereits auf den Zug aufgestiegen sind. Im Rahmen der Zukunftsgespräche wurde heuer beim Europäischen Forum Alpbach unter dem Titel „Umgang mit Daten

2 www.monasterium.net

3 <http://www.archivnet.at>

4 <http://data.wien.gv.at>

5 <http://www.linz.at/leben/55497.asp>

des öffentlichen Sektors“ ein Weißbuch zu Open Government publiziert⁶. Das Thema wird z. B. in den USA (2008), in Australien und in Großbritannien schon seit einigen Jahren viel diskutiert.

In diesem Zusammenhang ist daran zu erinnern, dass die Richtlinie PSI (Public Sector Information, 2003)⁷ der EU sowie das entsprechende österreichische Informationsweiterverwendungsgesetz (2005)⁸ die Weiterverwendung von Verwaltungsdaten regeln. Bis dato sind dabei die Daten von Archiven und Museen ausgenommen.

Nach den Open Government Data-Prinzipien sollen alle öffentlichen Daten verfügbar gemacht werden, d. h. Daten, die nicht durch besondere Bestimmungen (Datenschutz, Sicherheitserfordernisse oder andere rechtliche Zugangsbeschränkungen) geschützt sind. Eines der Kriterien ist, dass die Daten zur automatisierten Weiterverarbeitung strukturiert zur Verfügung gestellt werden. Ein weiteres Kriterium betrifft die Forderung, dass Einschränkungen zur Nutzung der Daten nicht gemacht werden sollen. Die Nutzung der Daten kann mit Kosten belegt werden, diese dürfen aber nicht diskriminierend sein. Diese Vorgaben sind in der PSI-Richtlinie bereits enthalten.

Zusammenfassend: „Offene Verwaltungsdaten sind Datenbestände, die von der öffentlichen Verwaltung im Bereich der Hoheitsverwaltung und privatwirtschaftlichen Verwaltung im Interesse der Allgemeinheit erhoben werden und in einem anerkannt offenen, maschinenlesbaren Format zur beliebigen, digitalen Weiterverarbeitung zur Verfügung gestellt werden.“⁹

Derzeit stehen im Fokus Daten aus den Bereichen Geoinformation, Budget, Verkehrsdaten etc. Einige Anbieter stellen im Rahmen von OGD-Services auch den Wortlaut der Gemeinderatssitzungsprotokolle online. Nicht unerwähnt bleiben soll der für Wien verfügbare digitale „Bedürfnisanstaltenführer“, der die Standorte öffentlicher Toiletten ausweist¹⁰.

Im Kulturbereich liegt der Fokus derzeit auf den Standorten und Öffnungszeiten von Museen, Bibliotheken und Sammlungen. Archive sind derzeit in den OGD-Portalen Österreichs noch nicht ausgewiesen. Die Berücksichtigung von Daten aus Archiven im Rahmen von OGD-Initiativen ist noch nicht wirklich angelaufen. Ein erster Schritt in diese Richtung ist das Einbringen von Kulturdaten und damit auch von historischen Daten in die GIS-Systeme der Gebietskörperschaften. Daran wird gearbeitet.

Fazit: Wir sind gut beraten, diese Entwicklung aufmerksam zu verfolgen und unser Angebot ins Spiel zu bringen. Wie die Diskussion derzeit steht, ist für

6 Martin Kaltenbeck und Thomas Thurner (Hg.), Open Government Data – Weißbuch. Edition Donau-Universität Krems, 2011. Digital verfügbar unter: http://issuu.com/semwebcomp/docs/ogd_weissbuch_2011_web (zuletzt abgerufen am 8. 10. 2011).

7 <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:DE:PDF>

8 <http://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20004375>

9 Kaltenbeck-Thurner, Weißbuch, S. 18.

10 <http://data.wien.gv.at/apps/wc.html>

Österreich ein einheitliches Portal für den Zugang zu den im Rahmen von OGD angebotenen Daten vorgesehen. Erforderlich ist auch ein proaktives Vorgehen, unsere Datenangebote müssen leicht gefunden werden können. Vor diesem Hintergrund sind auch die Archivportale zu sehen. Wenn wir mit unseren digitalen Angeboten noch stärker wahrgenommen werden wollen, dürfen wir uns nicht nur auf die Suchmaschinen verlassen, sondern müssen weitere Schritte in die Öffnung und Verfügbarkeit unserer Datenbestände setzen. Wir sind also aufgefordert, wieder einmal eine „Grenze“ zu überschreiten.

WEB 3.0

Wir alle kennen Anwendungen im so genannten WEB 2.0, das gerne auch als soziales Web bezeichnet wird. Es ist das Web zum Mitmachen, zum Twittern, zum Kommentieren und Setzen der Anmerkung „Gefällt mir“ ebenso wie die Welt der vielen Blogs und Wikis. Auf Facebook sind zwar einige österreichische Archive zu finden – allerdings wurden von Facebook die entsprechenden Artikel aus Wikipedia übernommen. Wirklich auf Facebook aktive Archive gibt es z. B. einige wenige in Deutschland. Der personelle Aufwand, eine Präsenz im Facebook professionell zu pflegen, ist für die meisten Archive zu groß.

Sollen wir noch in diese Entwicklungsstufe des Internets investieren oder uns gleich auf die nächste vorbereiten – das WEB 3.0, oder – etwas präziser ausgedrückt – auf das semantische Web? Gleich vorweg eine grundsätzliche Anmerkung: Das neue WEB wird die bisherigen Web-Versionen nicht ersetzen, sondern um eine wesentliche Eigenschaft erweitern.

Die Prinzipien sind längst vom W3-Consortium definiert worden¹¹. Ziel des semantischen Webs ist im Kern, dass die im Internet verfügbaren Informationen vom Computer interpretiert und weiter verarbeitet werden können. Es geht im Wesentlichen darum, Informationen über Personen, Orte und Sachen miteinander in Beziehung zu setzen. Das soll über maschinenlesbare Texte aus unterschiedlichsten Quellen möglich sein. Voraussetzung dafür ist die Annotation der Texte durch Metadaten, die aus einem kontrollierten Vokabular stammen, die Anwendung von Ontologien, die die logischen Beziehungen enthalten, und die Beachtung von Standards der Beschreibungssprachen. Für den Benutzer ergibt sich die faszinierende Aussicht, eine Anfrage in natürlicher Sprache zu stellen und eine relativ genaue Antwort zu erhalten.

Bei den drei Schlüsselbegriffen Person – Ort – Sache für den Informationsnachweis aus unseren Archiven werden wir hellhörig: Es sind genau die Bereiche, die auch in den Archivsystemen den gezielten Zugang zu den Inhalten des Archivgutes gestatten. Haben wir also gleichsam als schönen Nebennutzen

11 Weiterführende Informationen und Links auf der offiziellen Website des W3C unter: <http://www.w3.org/2001/sw/>

für die zukünftige Entwicklung des Webs diese Grenze schon überschritten? Ich meine, teilweise ja, da wir eine wesentliche Voraussetzung für den Einsatz semantischer Systeme sukzessive schaffen. Aus klug strukturierten Datenbanksystemen, egal ob relationale Datenbank oder auf XML-Basis, lassen sich Regeln für die Herstellung von Beziehungen gut ableiten.

Das Ganze steckt noch in den Kinderschuhen. Es existieren einige formale Sprachen zur Beschreibung von Ontologien¹². Sie sind in der Anwendung ziemlich anspruchsvoll und erst in einigen, vor allem naturwissenschaftlichen Bereichen im Einsatz. Ansätze gibt es auch in der Genealogie. Da ist es relativ leicht, eine biologische Beziehung zu definieren. Mit den Definitionen „Ist Sohn...“, „Ist Ehefrau...“ kommt man schon recht weit. Eine millionenfach verwendete Anwendung liefert uns Facebook mit „Is friend of“.

Für unser großes Arbeitsfeld in den Archiven wird sich der Einsatz dieser Weiterentwicklung des Webs voraussichtlich eher zaghafst einstellen. Es wird noch dauern, bis aus unseren maschinenlesbaren Daten wiederum durch Maschinen neues Wissen generiert werden kann. Wann diese Grenze tatsächlich erfolgreich überschritten werden kann, lässt sich derzeit noch nicht absehen.

Zur Entwicklung der Archive

Wir haben in den letzten drei Jahrzehnten einen tiefgreifenden Paradigmenwechsel in den Archiven erlebt. Das Arkanum ist der Offenheit gewichen. „Archive dienen der Information, nicht der Geheimhaltung“, war der Titel eines Vortrages auf dem diesjährigen deutschen Archivtag.

Welche Entwicklungsschritte lassen sich identifizieren:

- Die Öffnung der Archive ist erfolgt. Die Sperrfristen wurden durch Archivgesetze meist auf 30 Jahre verkürzt. Nur für besonders schützenswerte Daten gelten andere Kriterien.
- Findmittel wurden und werden in großer Zahl online gestellt und können von den Suchmaschinen indiziert werden.
- Online zugängliche Datenbanken gestatten das Durchsuchen einzelner Archive auf einfache Weise.
- Viel Archivgut wurde digitalisiert und ist über das Web benutzbar. Diese Arbeiten sind im Laufen und werden noch lange laufen müssen, um beim analogen Archivgut die von den Benützern erhoffte, z. T. auch geforderte digitale Benützung in viel größerem Ausmaß möglich zu machen.

¹² Z. B. RDF (Resource Description Framework), siehe <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222> oder OWL (Web Ontology Language), siehe <http://www.w3.org/TR/owl2-overview>

- Noch mehr Archivgut steht in den digitalen Lesesälen der Archive zur Nutzung bereit.
- Themenbezogene Archivportale führen verstreutes Archivgut virtuell zusammen.
- Die Ausstattung österreichischer Archive mit Informationssystemen ist schon weit gediehen.
- Für die digitale Langzeitarchivierung wurden und werden Lösungen gebaut und eingesetzt. Einige von uns haben die Vorführungen des ÖStA für seine Lösung miterlebt. Bis sie wirklich voll einsatzfähig ist und die entdeckten Probleme gelöst sind, dürfte es noch ein wenig dauern.

Was steht uns noch bevor? Das werden in nächster Zeit die Informationsfreiheitsgesetze sein. Die USA und Großbritannien haben bereits „Freedom of Information Acts“ erlassen¹³. Was darin festgelegt ist, geht weit über das in der OGD-Schiene Festgelegte hinaus. Diese Gesetze betreffen zunächst die laufende Verwaltung und in der Folge die Archive des öffentlichen Sektors. Freedom of Information steht oft in einem direkten Zusammenhang mit dem rein digital erzeugten Schriftgut.

Die Dimension dessen, worauf Bürger elektronisch zugreifen können sollen, geht weit über das hinaus, was man auf den ersten Blick erwarten würde. Alle Daten der Verwaltung sollen grundsätzlich offen zugänglich sein. Die Sperre dieser Daten aus Datenschutz-, Sicherheits- oder Urheberrechtsgründen ist die Ausnahme. Gefordert wird die digitale Einschaumöglichkeit in alle Verwaltungsunterlagen; auch in die internen Besprechungsprotokolle, externen und internen Mails und alle Vermerke.

Bedenkt man, dass nach der für die öffentliche Verwaltung geltenden Generalvorgabe der Stand einer Sache jederzeit aus dem Akt ersichtlich sein muss, ist über jedes Telefonat, jede SMS oder sonstige auf den Aktgegenstand bezogene Nachricht ein schriftlicher Vermerk anzulegen.

Wenn es in Österreich soweit ist, werden in einem ersten Schritt die elektronischen Aktensysteme mit einer entsprechenden Funktion erweitert werden müssen. In einem zweiten Schritt kommen die Archive der Gebietskörperschaften ins Spiel. Auch sie werden all das aus ihren digitalen Archiven digital bereitstellen müssen.

Um das tun zu können, brauchen wir nicht nur digitale Archive, sondern etwas, was zeitlich viel näher liegt: Wir müssen – soweit ich das für Österreich abschätzen kann – bald und viel stärker als bisher in das Recordsmanagement einsteigen. Das bedeutet, die übliche Grenze der archivischen Vorfeldarbeit

¹³ Erste Ansätze in den USA bereits 1966 mit zahlreichen nachfolgenden Änderungen. Die seit 1979 vorgenommenen Novellierungen sind unter <<http://www.justice.gov/oip/foi-upd.htm>> ausgewiesen. Das Gesetz für Großbritannien wurde 2000 erlassen: Freedom of Information Act 2000. <<http://www.legislation.gov.uk/ukpga/2000/36/contents>>.

weit zu überschreiten und so die Voraussetzungen zu schaffen, dass wir auch zukünftige Aufgaben bewältigen können.

Fazit: Wir können den kommenden Anforderungen als Archivarinnen und Archivare nur genügen, wenn wir unsere fachlichen Kompetenzen ständig weiterentwickeln. Im Sinne des Referatstitels werden wir die bisherigen Grenzen unserer Arbeit immer wieder überschreiten müssen, um in der Zukunft erfolgreich bestehen zu können.