Normierte Archivierung?

Normen und Standards gehören seit alters zu unserem Alltag; im Zusammenhang mit der Informationsgesellschaft und den Informations und Kommunikationstechnologien finden sie wachsende Aufmerksamkeit (Andreas Kellerhals: Das normierte Archiv. Über die Vielfalt der Normen und die normierte Vielfalt, in: Studien und Quellen 26, Bern 2000, S. 273–298). Sie sollten das Zusammenleben und Zusammenarbeiten wenn nicht überhaupt ermöglichen so doch erleichtern. Das gilt auch für die Archivierung, die eine kooperative Arbeit war und bleibt, und seit Langem von unterschiedlichsten Normen und Standards geprägt ist. Gleichzeitig tragen Normen und Standards zur Qualitätssicherung bei.

Dabei ist das Spektrum an unterschiedlichen Normen und Standards breit: Es reicht von ethischen und rechtlichen über professionelle bis zu technischen Normen, von bewusst eingesetzten bis zu implizit «eingebauten» oder internalisierten, von konsensuell bestimmten bis zu einseitig gesetzten. Im Folgenden abstrahieren wir von den rechtlichen und beschränken unsere Überlegungen auf die technisch-professionellen Normen und auch dies speziell im Kontext der digitalen Archivierung².

Das Grundmodell des digitalen Archivs, das Open Archivar Information System (OAIS), ist von den Luft- und Raumfahrtorganisationen unter der Leitung von NASA und mit der Beteiligung der National Archives and Records Administration (NARA) sowie der Research Libraries Group (RLG) 1997 entwickelt und bereits Anfang 2002 als ISO 14721 mit wenigen Änderungen als Standard angenommen worden³. Es ist kein theoretisch modelliertes Konzept, sondern umfasst bereits anerkannte Arbeitsergebnisse der archivischen Praxis. Es verzichtet auf Detailbeschreibungen (im technischen Sinne) von Schnittstellen, Datentypen, Datenformaten etc. Gerade darauf gründet sein Erfolg, denn dieser Standard kann einerseits bei unterschiedlichen archivischen Institutionen implementiert werden, andererseits ist er offen für unterschiedliche technische Systemlösungen und problemlos verträglich mit den sich laufend verändernden Technologien.

Der OAIS-Standard bleibt allerdings reichlich abstrakt und die einzelnen Komponenten wie Ingest, Preservation Planning oder Access sowie die enthaltenen Elemente, wie etwa die verschiedenen «Information packages» sind damit nicht automatisch auch schon standardisiert. Entsprechende Normierungen sind beim Aufbau eines digitalen Archivs entweder noch zu leisten oder die Lücke ist durch den Rückgriff auf bestehende Normen zu schliessen. Dabei kommen unterschiedlichste archivische und technische Normen und Standards in den Blick.

Aus archivischer Perspektive denken wir etwa an ISAD(G), ebenfalls eine Synthese bewährter archivischer Praxis und verwandte Normen für die Verzeichnung oder die ICA-Prinzipien betreffend den Zugang⁴. Diese Normierungsbemühungen schreiben sich in einen allgemeineren Kontext ein, Recherche, Zugang zum Archivgut oder überhaupt zu Informationen wenigstens des öffentlichen Sektors sowie deren weitere Nutzung im Sinne der Open-Data-Bewegung zu regeln; grundsätzlich lassen sich diese Normen auch auf den Zugang zu Informationen des privaten Sektors oder wenigstens der weitgehend staatlich finanzierten wissenschaftlichen Forschung anwenden, wenn denn ein Wille da wäre, solche Daten öffentlich verfügbar zu machen⁵. Insofern in der digitalen Welt künftig die archivischen Metadaten direkt aus den von Anfang an produzierten Metadaten generiert werden – die Erschliessungsarbeit also praktisch entfallen wird – müssen entsprechende Anforderungen im Rahmen der Unterlagen-Produktion implementiert werden. Zu den zu übernehmenden Metadaten gehören jedoch über Informationen zum Recherchieren und zur Ermöglichung des intellektuellen Zugangs im Sinne von ISAD(G) hinaus auch weitere unverzichtbare Informationen wie Hinweise zur Bestimmung der Archivtauglichkeit, zur Umsetzung notwendiger digitaler Erhaltungsmassnahmen oder zusätzliche technische Details, die u.a. einen differenzierten Zugang gewährleisten⁶.

Den Online-Nutzenden von digitalem Archivgut werden die archivierten Unterlagen nicht mehr zwingend in ihrer Originalform angeboten, sondern – je nach Voraussetzungen der Archivierung bzw. nach Anforderungen der Kund/innen – in unterschiedlichen Formaten und Strukturen. Die herkömmlichen Findmittel, aber auch die in unterschiedlichen Archivinstitutionen archivierten Daten werden miteinander verknüpft und können digital zusammenerfasst und computerunterstützt ausgewertet werden. Alle angewandten Normen und Standards müssen also die Lesbarkeit, Verständlichkeit und Nutzbarkeit erhöhen.

Die Entwicklung neuer Archivierungslösungen wird nie aufhören. Sie wird somit nicht nur bezüglich der Übernahme von neuen komplexen Unteragentypen notwendig bleiben, sondern es werden sich auch bezüglich der Entwicklung des Zugangsangebots zu Archivgut ständig neue Herausforderungen stellen. Die einheitliche Art und Weise, wie man automatisch Informationspakete zwischen verschiedenen Systemen austauscht, die Vernetzung der Daten usw. macht den Einsatz weiterer Standards nötig.

Das Referenzmodell OAIS und der ICA-Standard ISAD(G) legen die Leitlinien fest, wie man die Archivbestände aufbaut und wie die Arbeitsprozesse des digitalen Archivs weiterentwickelt werden. Die Konkretisierung dieser Vorgaben wird mithilfe weiterer immer eindeutigerer (technischer) Standards erreicht, beispielsweise müssen Paket-Formate für das SIP definiert werden oder deskriptive Standards für die Beschreibung der archivischen Metadaten in Archivsystemen. Eine Verfeinerung des jeweiligen Detaillierungsgrads geschieht dann immer nach demselben Prinzip.

Diesbezüglich legt das Schweizerische Bundesarchiv in seiner SIP-Spezifikation⁷ fest, wie man digitale Unterlagen an das BAR abliefern kann. Dabei stützt sich das BAR, bei der Metadatenbeschreibung auf den Dublin Core Standard⁸ (DCMI Kompatibilität) und ISAD(G). Der SIP-Standard, Version 4.0 des BAR ist offen zugänglich und bildet die Grundlage für die Standardisierung der Ablieferungsschnittstelle des eCH. Es existieren auch andere Standardbeschreibungen für «Container-/Paket-Formate», wie etwa METS⁹, die wiederum in einer Kombination mit einem Format wie EAD¹⁰ oder Dublin Core für die Beschreibung von Informationsobjekten benutzt werden können. Alle diese Beschreibungen sind relativ neu, ihre Ausbreitung deshalb unterschiedlich gross, und es besteht noch eine gewisse Konkurrenzsituation zwischen diesen verschiedenen Normen und Standards. Allen ist allerdings gemeinsam, dass sie bald in veränderten neuen Versionen vorliegen werden. Die digitalen Ablieferungen ins Archiv von 2009 werden nicht die gleiche Form haben wie 2012 oder 2016.

Teil der SIP-Spezifikation des BAR ist die Bestimmung der vom BAR akzeptierten archivtauglichen Dateiformate. Das BAR hat einen entsprechenden Katalog publiziert (PDF/A, CSV, TIFF, SIARD usw.)¹¹. Archivtaugliche Dateiformate verweisen ihrerseits wiederum auf weitere technische Normen und Standards, wie ISO Latin-1 (ISO 8859-1) und ISO Latin-9 (ISO 8859-15) oder Unicode (ISO 10646) und SQL:1999 (ISO/IEC 9075:1999).

Solche archivtauglichen Dateiformate müssen offen dokumentiert und genau spezifiziert sein, damit ein Einsatz in generischen Umgebungen möglich wird; sie sollten aus heutiger Sicht eine Lebensperspektive von zehn und mehr Jahren besitzen, wie beispielsweise das TIFF-Format, das in den späten achtziger Jahren des 20. Jahrhunderts entwickelt wurde. Dieser Katalog bedarf aber seinerseits einer ständigen Überprüfung. Gleichzeitig ist die Abstimmung mit den Produktionssystemen der Unterlagen absolut zentral. Die Durchführung der Konvertierungen für die Archivierung könnte man beispielsweise vermeiden, würde man in den Produktionssystemen von Anfang an offene Formate verwenden.

Eine ständige Überwachung der Technologie ist für die Planung und Durchführung von Erhaltungsmassnahmen (Migrationsstrategie) notwendig und beeinflusst unmittelbar den Zugang zum Archivgut. Das BAR prüft deshalb gegenwärtig den offenen Standard für XML-basierte Dateien, Office Open XML (OOXML, ISO 29500) von Microsoft, der bereits in den Büroautomationsanwendungen der Bundesverwaltung verbreitet ist.

Eine dauerhafte technische Lösung für die digitale Archivierung ist nicht zu erwarten. Im besten Sinne des OAIS bleibt ein digitales Archiv ein dynamisches System, dessen Kernaufgaben die Überlieferungsbildung und bestmögliche Nutzung des Archivguts sind, und das immer von der technischen Innovation ausserhalb des archivischen Umfelds stark beeinflusst werden wird. Es entstehen OAIS-kompatible IT-Lösungen für digitale Archive, komplexe Datentypen und Daten aus neuen Produktionssystemen (Geodaten, Twitter usw.) müssen übernommen und gesichert, moderne Kommunikationsplattformen (z.B. BAM, APE12) müssen bedient werden.

Damit kommen wir zu einem u.E. kritischen Punkt: Normen und Standards sollen die Zusammenarbeit erleichtern und zur Qualitätssicherung beitragen, Effizienz und Verlässlichkeit stehen im Zentrum. Das hat verschiedene Konsequenzen:

1. Es muss in der Vielfalt der Normen eine Auswahl getroffen werden.
2. Die Normierung eines Sachverhalts muss schnell genug erfolgen, damit sie genutzt werden kann, bevor rückwirkend eine reale Vielfalt auf ein standardisiertes Mass zurückgestutzt werden muss – was meist unmöglich ist oder unverhältnismässigen Ressourceneinsatz verlangte.
3. Beim Rückgriff auf Normen oder bei der Normierung selber dürfen wir das Hauptziel unserer Tätigkeiten nicht aus dem Blick verlieren: Es geht nicht um Perfektion, sondern um Praktikabilität in guter Qualität, um Einschränkung von Vielfalt zur Effizienzsteigerung, um Minimierung des permanenten Anpassungsdrucks und um Steigerung der Verlässlichkeit. Mit anderen Worten: Der immer wieder feststellbaren Überreglementierung ist entgegenzutreten – das hilft wiederum, durchaus nötige Normen sinnvoll ein- und durchzusetzen.

Das heisst auch: Wir sind im Rahmen eines Life Cycle Managements von Akten, Daten, Informationen sehr an Normierungen interessiert. Diese sollten in erster Linie sicherstellen, dass die Inhalte, welche uns zentral interessieren und welche wir zuverlässig überliefern können müssen, systemunabhängig und möglichst einfach verwaltet, gepflegt, erhalten und genutzt werden können. Weder bezüglich der archivischen Erschliessung noch betreffend der Aktenführung/Geschäftsverwaltung halten wir alle real existierenden Normen aus praktischer Sicht für sinnvoll. Wir sind auch noch nicht überzeugt, dass die entsprechenden Normierungsprozesse rechtzeitig zu Ergebnissen führen; das Schlimmste, was uns passieren könnte, wäre eine Verabschiedung von Normen, die überholte Lösungen durchsetzen und uns dadurch mehr blockieren, als durchaus erwünschte und zukunftssträchtige Entwicklungen unterstützen. Wir plädieren deshalb auch tendenziell für minimalistische Lösungen. Unsere eigene Erfahrung hat uns gelehrt, dass selbst im Kern sinnvolle Normen wie ISAD(G) bei der Umsetzung in konkrete (hausinterne) Regeln zu einer Komplexitätssteigerung führen können, die nicht mehr handhabbar ist¹³.

Zur Frage des zeitgerechten Normierens ist zudem anzumerken, dass die Standardisierungen – von der ISO bis zu eCH.ch – aus guten Gründen eine gewisse Zeit beanspruchen. Während dieser Zeit setzt sich aber in der Realität oft bereits eine Art (de facto) Standard durch, welcher für die Archivierung letztlich von grösserer Bedeutung ist, als eine unbeachtete De-jure-Norm. Wir müssen die Probleme von heute mit den Mitteln von heute lösen und nicht die beste aller Welten erfinden. Dabei müssen wir aber immer im Blick haben, dass die Informationsproduktion und die – zeitlich oft deutlich distanzierte – Informationsnutzung in enger aber nicht gegenseitiger Abhängigkeit stehen. Das zeigt sich in vielerlei Hinsicht. Marktgetriebene Vereinheitlichungen oder nachträgliche Normierungen können uns dabei entgegenkommen – vgl. etwa das Beispiel der HTML-Vereinheitlichung, welche die Web-Archivierung erleichtern wird. Ein weiteres Beispiel hierfür ist die oben erwähnte offene Standardisierung bislang proprietärer Dateiformate¹⁴, die neue Handlungsspielräume eröffnen und damit lange als unverzichtbar verstandene Arbeitsschritte wie die Konvertierung von Dateien obsolet werden lassen. Gleichzeitig gilt es zu beachten, dass archivierte Informationen immer in einer bestimmten Form vorliegen werden und – möglicherweise nur mit enormem Aufwand – technisch für neue Nutzungsmöglichkeiten erschlossen werden können. Die geltenden Transparenznormen – das rechtliche Öffentlichkeitsprinzip etwa – wird möglicherweise einen Einfluss darauf haben, ob der Motor der Entwicklung weiter primär auf der Seite der Datenproduktion und Anwendungsentwicklung liegt oder ob sekundäre Nutzungsansprüche schon früh auf diese Informationsproduktion einen Einfluss ausüben wird. Wie auch immer: Wir werden Entwicklungen sowohl folgen können, als auch deren Potenzial besser nutzen lernen müssen.

Damit kommen wir zum Schluss zur Frage der Auswahl von Normen: Für welche SQL-Sprache soll SIARD funktionieren? Soll der GEVER-Einsatz der ISO-Norm 15489 folgen oder MoReq 2010¹⁵ oder ICAREq?

Wir müssen hier Entscheidungen treffen und dann mit den Konsequenzen leben lernen. Es ist deshalb jenseits der klassischen Normierungen und Standardisierungen wichtig, «Comunities of practice» zu bilden. Diese können durch gemeinsame Entscheide im Sinne von Best-Practice-Lösungen entwickeln und betreiben, die allen Beteiligten einen Nutzen bringen und Verlässlichkeit, Qualität und Wirtschaftlichkeit gewährleisten.