



# 1 **Digital Humanities und die Auswertung digitaler Unterlagen**

## 1.1 **Workshop des Schweizerischen Bundesarchivs (BAR), 12. Juni 2013**

### 1.1.1 **Bericht**

Auch die Archive unterstehen dem digitalen Wandel. Das BAR erarbeitet gegenwärtig ein Nutzungsangebot, das dem entspricht. Dabei stehen quantitative und qualitative Zugänge im Vordergrund: Unter **Auswertung** verstehen wir die Beantwortung einer auf eine bestimmte Datenbasis bezogene Fragestellung, darunter insbesondere: Recherche, Datenkritik, Interpretation und Datentransfer. Den in den *Digital Humanities* diskutierten Nutzungsmethoden wie *data mining*, *data visualization* oder *distant reading* soll also entsprochen werden können.

Moderiert von Guido Koller, BAR, diskutierten Experten und Expertinnen folgende Fragen: Liegen den *Digital Humanities* grundlegend neue Forschungsfragen zugrunde? Welches sind die **Bedürfnisse der scientific community** in Bezug auf Forschungsinfrastrukturen? Welche Anforderungen stellt sie dabei ganz konkret an die Archive?

Andreas Kellerhals, Direktor BAR, wies in seinem Grusswort auf das Archiv als Voraussetzung für *e-open-sciences* hin. Wie muss ein digitales Archiv aussehen? – Vieles wird erst angedacht, konzipiert oder geplant. Deshalb ist der Einbezug der *scientific community* so wichtig.

Peter Fler, BAR, beleuchtete die Wandlung der Archive von Datenlieferanten zu Kooperationspartnern in „rekursiven Informationsanreicherungsprozessen“, kurz: vom **gate keeper zum data broker**. Das Terrain ist noch unsicher: Behördliche digitale Unterlagen werden erst seit Kurzem in Archive abgeliefert. Erste Erkenntnisse: Der Grad der *Datafizierung* von Unterlagen hängt nicht von deren Herkunft ab, des Weiteren sind benutzerfreundliche *User Interfaces* notwendig und entsprechende *Tools* müssen den Austausch gewährleisten können.

Enrico Natale, Geschäftsführer *infoclio*, bot einen kurzen Überblick zu der noch jungen Geschichte der von Susan Schreibman so genannten **Digital Humanities**. Der *umbrella term* meint die Anwendung von „computational methods“ im Forschungsbereich der Geisteswissenschaften. *Digital Humanities* sind vor allem in den englischsprachigen Ländern verbreitet.

Beat Estermann, Berner Fachhochschule, interessiert, wer Daten zur Verfügung stellt und wer sie nutzt. Die Lieferanten sorgen sich um den Aufwand zur Aufbereitung der Daten und um rechtliche Probleme (Urheberrechte, Datenschutz etc.). Einnahmeverluste werden überraschend selten als Problem genannt. Überraschend ist auch: Viele Nutzer gewichten beim Entscheid für ihr Forschungsthema die Tatsache der **Verfügbarkeit von Daten** höher als die Fragestellung. Sie haben hohe Ansprüche an die Integrität der von ihnen benutzten Daten.

Martin Stoll, *SonntagsZeitung*, berichtete aus der noch jungen Praxis des **Datenjournalismus**. Die Hauptfrage hier: Wie generieren Zeitungen Geschichten aus Daten? Der Aufwand zur Beschaffung z.B. statistischer Daten ist enorm. Öffentliche Stellen sollten sie übersichtlich aufbereiten sowie „unkompliziert“, „gratis“ und „möglichst uneingeschränkt“ zur Verfügung stellen.

Armin Grossenbacher, Bundesamt für Statistik, bereitet grosse demographische, wirtschaftliche, politische und kulturelle Datenmengen auf. Die Veröffentlichung misst sich am Interesse

des Publikums und an der „statistischen Ethik“: Neben der Ausrichtung auf **Zahlentabellen** als Grundlage für die Abbildung öffentlichen Handelns kommt der Datenschutz zum Tragen. Die bereitgestellten Daten sollten weiter bearbeitet werden können. PDFs kommen deshalb nur als Kompromiss in Frage. Wichtig sind semantische Beschreibungen (*RDF Triples*, vgl. Präsentation), um die Auffindbarkeit zu gewährleisten.

„*Software, Hard Facts and the Quality of Interpretation – Big Data in Historical Research*“ lautete der Beitrag von Patrick Kammerer, Universität Zürich. Der Präsident des Vereins *Geschichte und Informatik* beschrieb und erläuterte sein Konzept der **Quellen- respektive Datenkritik** anhand der an der Landesausstellung 1939 präsentierten Geburtenstatistik. Er zeigt mit seinem Beispiel, wie zentral der Entstehungskontext solcher Quellen ist für die historische Auswertung und Interpretation.

Christiane Sibille präsentierte Lösungsansätze der **Digital Diplomatic Documents**. Der Kern bleibt die Auswahl relevanter Dokumente. Die Indexierung schafft einen wissenschaftlichen und generellen Mehrwert für ein immer breiteres Publikum. Vor allem Personen, Körperschaften und Orte lassen sich virtuell sehr gut vernetzt darstellen. Damit werden Recherchen optimiert. Neue Selektionsmöglichkeiten und Technologien zur Auswertung von Quellen werden ausprobiert.

Frédéric Kaplan, EPFL, sprach über die Herausforderungen für die Definition der **Schnittstelle zwischen Digital Humanities und Archiven**. Notwendig für sein *Venice-Time-Machine* Projekt ist ein System, in dem Informationen gut organisiert werden können, da die Quellen so reichhaltig sind: 80'000 Laufmeter aus 1'000 Jahren. Es geht also um den Aufbau eines dynamischen Informationssystems mit „Milliarden von Informationen“ – zum Beispiel: Wer hat wann wo gewohnt? Nach der Digitalisierung der Quellen werden Sprache, Schrift und Rechtschreibung bereinigt. Kaplan hofft auf eine zumindest teilweise automatisierbare Transkription. Die Daten werden schliesslich in einem geohistorischen System mit Metadaten (RDF) semantisch dargestellt. Zur aufwendigen technischen Gestaltung der dokumentarischen und virtuellen Räume (*DublinCore, Pivot ISO CIDOC-CRM* etc.) vgl. die Präsentation zum Projekt.

Die Schlussdiskussion drehte sich um die Frage nach der **Rolle des Archivs**. Für Kaplan kommt diese primär bei der Strukturierung der Information zum Tragen. *User* möchten bei der Formulierung von Anforderungen zur Bereitstellung und Darstellung von Daten einbezogen werden. Die Frage der Unterscheidung von Meta- und Primärdaten ist wichtig auch für eine historisierte Darstellung als Voraussetzung für die Interpretation von Quellen. Es geht also in Archiven darum, Algorithmen für Recherchen zu entwickeln und zu „tunen“.

Zentral bleibt die **Quellen-, respektive Datenkritik**. Wichtig ist die Frage nach den Subjekt-, Rollen(träger)- und Objektbeziehungen in dem zu untersuchenden Dokumenten-Korpus: Wie wird die klassische archivische Frage nach Tradition versus Überrest in digitalen Geschäftsverwaltungs-, Überlieferungs- und Analyse-Systemen inskünftig beantwortet?

Die Teilnehmenden sind sich einig, dass zurzeit neue Muster für Fragestellungen entwickelt werden. Dabei werden individuelle Ansätze vermehrt von kollektiven abgelöst (Stichwort: *Crowds and Clouds*). Die Produktion von Information wird „flüssiger“. Institutionen können helfen, solche Formen der Zusammenarbeit zu orchestrieren. Archive werden sich dabei auf die Datenbasis konzentrieren (Recherche, Datenkritik, Interpretation und Transfer). In diesem Sinne bleiben Archive das **Rückgrat für die Forschung**, wie Andreas Kellerhals versicherte: Sie gewährleisten weiterhin die Authentizität der Informationen und können daneben „Labore“ sein, helfen, neue Forschungsfragen zu testen. Das BAR freut sich auf gemeinsame Projekte!