

Bericht über den Workshop

Qualifizierung und Vernetzung Öffentlicher Digitalisate

im Architekturmuseum der TU Berlin am 24./25. Juni 2008

Die Digitalisierung von Kulturgut und die Vernetzung öffentlicher Datenbanken verändern die Wissenschaftslandschaft seit über zehn Jahren rapide. Eines der öffentlich geförderten deutschen Digitalisierungsprojekte ist an der Technischen Universität Berlin beheimatet. Die Deutsche Forschungsgemeinschaft (DFG) finanziert seit Beginn des Jahres 2006 das Projekt DIGIPLAN, das in drei Jahren den gesamten Bestand (über 100.000 Zeichnungen und Fotografien) des Architekturmuseums der TU Berlin scannt und in hoher Auflösung im Internet zur Verfügung stellt.

Der von diesem DFG-Projekt veranstaltete Workshop „Qualifizierung und Vernetzung Öffentlicher Digitalisate“ am 24./25. Juni 2008 diente dem Austausch über aktuelle Fragen der Digitalisierung. Der Workshop war in drei große Themenblöcke gegliedert: Gegenwart und Zukunft der Digitalisierung, Qualifizierung durch Standardisierung und Georeferenzierung sowie Vernetzung mittels Schnittstellen und Portalen.

Das Seminar fand im kleinen Kreis von 24 geladenen Fachleute (knapp die Hälfte davon von außerhalb) statt. Leider hatten sowohl die DFG, als auch die Bayerische Staatsbibliothek und Google Deutschland ihre Teilnahme abgesagt. Aber mit dem Architekturmuseum der TU München, dem Bildarchiv Preussischer Kulturbesitz, dem Bundesarchiv, der Deutschen Fotothek, Foto Marburg, dem Göttinger Digitalisierungszentrum, dem Institut für Museumsforschung, dem Landesarchiv Berlin, dem Prometheus-Bildarchiv, den Staatliche Museen zu Berlin, der Staatsbibliothek zu Berlin, der Stiftung Preußische Schlösser und Gärten und der Wikimedia Deutschland waren zahlreiche wichtige Protagonisten der deutschen Kulturgutdigitalisierung vertreten. Unter den Teilnehmern befanden sich demnach etliche Contentprovider, ein Digitalisierungsdienstleister, Portalbetreiber (BAM, Bildindex, Deutsche Fotothek, Prometheus), Vokabularvereinheitlicher sowie Schnittstellendefinierer (Institut für Museumsforschung, Museumsdat).

Es waren keine ausgearbeiteten Vorträge, sondern lediglich Kurzstatements zu den genannten Themenblöcken erwünscht. Der kleine Kreis und der Verzicht auf Frontalvorträge ermöglichten es, dass man ausführlicher als auf anderen Tagungen miteinander ins Gespräch kommen konnte

Nach der Begrüßung durch den Leiter des Architekturmuseums, Dr. Hans-Dieter Nägelke, und einer kurzen Selbstvorstellung aller Teilnehmer begann die Diskussion über den gegenwärtigen Stand und die Zukunft der Kulturgutdigitalisierung in Deutschland mit der Frage nach der Quantifizierung: Wie viel müssen wir insgesamt digitalisieren? Dabei wurde schnell klar, dass die Menge des überlieferungswürdigen Kulturguts noch gänzlich unübersehbar ist. Auch nach zehn Jahren Massendigitalisierung steht man noch ganz am Anfang der Bemühungen. Augenfällig wurde das im Bereich der Bibliotheken am Beispiel der Anzahl der im Göttinger Digitalisierungszentrum (einem von insgesamt zwei nationalen Digitalisierungszentren) in den letzten zehn Jahren gescannten Bücher: 9.220. Angesichts der vielen Millionen in Deutschland seit dem 15. Jahrhundert gedruckten Bücher

eine verschwindend geringe Zahl (die Geschwindigkeit wird sich allerdings in den nächsten Jahren durch den Einsatz von Scanrobotern ca. verzehnfachen).

Die Frage, wo die Kulturgutdigitalisierung in zehn Jahren stehen wird, kann aufgrund mangelnder Zahlen über die Gesamtmenge des Kulturguts und der bislang erreichten und künftig zu erreichenden Digitalisierungsgeschwindigkeit nicht abgeschätzt werden. Dem Tagungsverlauf vorgreifend, sei an dieser Stelle der Einwand von Stefan Rohde-Enslin vom Institut für Museumsforschung angeführt: Wann ist ein Objekt überhaupt als „digitalisiert“ zu bezeichnen? Reicht es, wenn ein digitales Foto vorliegt? Oder ein Scan beliebiger Qualität? Ist nur ein im Volltext erschlossenes Buch als digitalisiert zu bezeichnen? Reicht das Abbild der Seiten? Die Tendenz ist eindeutig: nur hochwertige Scans direkt vom Original (und bei Büchern zusätzlich deren Volltexterschließung) werden als Digitalisierung im engeren Sinne anerkannt.

Scans eines Originals werden mitunter „Direktdigitalisat“ genannt. Je nach Art der Vorlage ist diese Qualitätsstufe verschieden definiert. Gegenstände, die auf einen Scanner gelegt werden können, können dadurch „direktdigitalisiert“ werden. 3D-Objekte, insbesondere Architektur, können mit einer Digitalkamera direktdigitalisiert werden; in aller Regel aber werden fotografische Reproduktionen von 3D-Objekten digitalisiert. Eine neue Möglichkeit, die zunehmend an Verbreitung gewinnt, ist der 3D-Scan von räumlichen Objekten. Dabei liest eine an einem Metallarm geführte Digitalkamera rund um das zu digitalisierende Objekt die Entfernung zum Objekt und die Bildinformationen ein und setzt die gewonnenen Pixelwolken zu einem virtuellen Objekt zusammen.

Beim Themenfeld Qualifizierung der Digitalisate durch Standardisierung und Georeferenzierung interessierte die Teilnahme besonders das *enrichement* der Datenbanken durch Geoinformationen. Wolfgang Crom von der Kartenabteilung der Staatsbibliothek monierte eingangs, dass der Terminus „Georeferenzierung“ für die hochpräzise Einmessung eines Objektes durch Geodäten reserviert sei. Für den Sprachgebrauch im Bereich der Kulturgutdigitalisierung empfahl er die Begriffe „Lokalisierung“, „Geo-Metadaten“ oder auch allgemein „Verknüpfung mit einem Raumbezug“. In der Tat ist die relative Unschärfe und notwendige Nichtpräzision ein Charakteristikum der Kulturgutlokalisierung.

Nachdem Google und Flickr schon seit einigen Jahren raumbezogene Informationen visualisieren, war die Deutsche Fotothek im Jahr 2007 zumindest für den deutschsprachigen Raum Avantgarde, als sie die Einbindung vieler ihrer Online-Digitalisate in GoogleMaps freischaltete. Das Architekturmuseum der TU Berlin konnte auf dem Workshop den Dialog mit den Nutzern des eigenen Online-Angebotes vorführen. In einer sogenannten Schnitzeljagd können angemeldete Nutzer online an der Lokalisierung der ca. 100.000 Scans mitarbeiten. Der Dialog fragt nach der Validität der vorgeschlagenen Lokalisierung, der Nutzer kann die Lokalisierung gegebenenfalls ändern und speichern. Für eine gewisse Anzahl von Hinweisen erhält der Nutzer einen Bonus in Form eines Faksimileprints aus den Beständen des Museums. Dieses *outsourcen* von Aufgaben an die web-community trägt den englischen Neologismus *crowdsourcing*.

Es sind drei Qualitätsstufen der historischen Georeferenzierung zu differenzieren:

- 1) Historische Ortsnamen werden mit Geodaten verknüpft (zum Beispiel die Lokalisierung der Wikipedia-Einträge).

- 2) Historische Karten werden entzerrt und georeferenziert und können so als Grundlage für aktuelle und historische Geodaten fungieren.
- 3) Vision: Historische Adressen (Straßennamen und Hausnummern) werden mit Geodaten versehen.

Die ex-post-Lokalisierung historischer Adressen ist sehr zeitintensiv und bislang nur höchst rudimentär erfolgt. Immerhin wird die Speicherung der veralteten gegenwärtig im Netz vorhandenen Geoinformationen zu einem nach und nach immer dichter werdenden Raum-Zeit-Archiv der Jahre ab ca. 2005 führen.

Die Suche im *WorldWideWeb* erfolgt bislang fast ausschließlich mit Hilfe der Volltextsuche. Die beiden Haupterkennnisprinzipien Raum und Zeit werden kaum benutzt, da diese Daten im Netz bislang unzureichend dokumentiert sind. Es gibt Projekte, die Suchergebnisse zumindest in einer Zeitachse zu visualisieren. Auch der Raumbezug lässt sich bei Flickr, Panoramio u.a. anderen *websites* mittlerweile schon sinnvoll suchen. Die Kombination beider ergäbe eine Raum-Zeit-Matrix, die völlig neue Nutzungen und Erkenntnisse ermöglichte, insbesondere unter Einsatz von mobilen GPS-tauglichen Laptops.¹ Angeblich stehen große internationale Projekte kurz vor der Präsentation, so Ralf Stockmann (GDZ, Göttingen).

Prof. Dr. Hubert Locher (Foto Marburg) erklärte das konstatierte große Bedürfnis nach Verortung im Internet mit dem weitverbreiteten Unbehagen an der prinzipiellen Ortlosigkeit der virtuellen Welt.

Die Überarbeitung der DFG-Praxisregeln im Förderprogramm Kulturelle Überlieferung und die Fraunhofer Studie „Auf dem Weg zur Deutschen Digitalen Bibliothek (DDB)“ sowie der Komplex Open Source Software und die Erörterung der (DFG-)Förderungsmöglichkeiten kamen entgegen anderslautender Ankündigung aus Zeitgründen nicht zur Sprache.

Am zweiten Tag des Workshops standen Fragen der Vernetzung im Mittelpunkt: Gruppieren sich zusammengehörnde Informationen bevorzugt mit Hilfe von intelligenten Suchalgorithmen oder mittels thematisch zusammengestellter Portale? Beide Lösungen setzen vereinheitlichte Metadaten (Personennamen, Ortsnamen, Sachverschlagnwortung) voraus; die Portalisierung verlangt zudem eine Erleichterung des Datenharvesting durch vereinheitlichte Austauschschnittstellen wie MUSEUMSDAT. Um das Ergebnis vorwegzunehmen: Den anwesenden Vertretern der Portale BAM/EUBAM, DDB, EUROPEANA, POIKILE) gelang es nicht vollständig, die Zweifel an der Sinnhaftigkeit der Spiegelung von Online-Datenbanken in mehreren Portalen auszuräumen – es sei denn, die Portale bieten die Daten exklusiv an, wie es im Bildindex von Foto Marburg in vielen Fällen der Fall ist. Oder aber, wie im Fall von PROMETHEUS, das Portal bietet einen hohen funktionalen Mehrwert (Leuchtpult, Einschränkung der Suche auf relevante Bildbestände, Umwandlung in eine Powerpoint-Präsentation, Online-Präsentation mit Doppelbildprojektion, Klarheit der Bildrechte).

Portale leisten im Idealfall eine Verknüpfung von Informationen im Sinne einer Suchabfrage, die in voller Kenntnis des Materials die richtige Frage stellt. Davon sind die Portale aber zumeist weit entfernt. Die Suchmaschinen hingegen sind bekanntlich bei der mit-

¹ Eine topographiezentrierte Such- und Ergebnismaske strebt die kalifornische Electronic Cultural Atlas Initiative an, s. <http://ecai.org>.

helfenden Analyse der Anfragen und der Bewertung von Suchergebnissen recht erfolgreich.

Über die großen Massen der mittlerweile erzeugten Scans müsste eigentlich eine maschinelle semantische Inhaltsanalyse laufen – denn das Problem ist nicht mehr länger die Bildqualität, sondern vielmehr die Unzulänglichkeit der Metadaten. Was wäre eine textbezogene Suche im Netz wert, die sich nicht auf die Volltexte, sondern nur auf die (zumeist nicht vorhandenen) Metadaten stützen würde? In diesem Stadium befindet sich gegenwärtig – und noch auf viele Jahre – die Suche nach Bildinformationen.

Der eigentliche Schrecken der (Buch-)Massendigitalisierung ist die Doppelt-Digitalisierung, nicht der Datenverlust. Keinem der Teilnehmer war auf Nachfrage ein Beispiel für einen gravierenden dauerhaften Datenverlust oder ein gravierendes Kompatibilitätsproblem eines Digitalisierungsprojektes bekannt. Aus diesem Grund war die Langzeitarchivierung auch kein Thema des Workshops.

Rege genutzt wurde die Möglichkeit, dem Auditorium im Internet die eigene oder andere vorbildliche Websites vorzuführen. In Anlehnung an ein Begleitphänomen der zeitgleich stattfindenden Fußball-Europameisterschaft wurde dafür der Terminus *public surfing* geprägt – eine neue, zukunftssträchtige Art wissenschaftlicher Zusammenkünfte.

Aus Göttingen stammt folgende Zukunftsmusik: Man könnte vielleicht einmal die Regale der Bibliotheken röntgen und die Druckerschwärze der Buchstaben wie die Zellen des menschlichen Körpers in einem Computertomographen erkennen und auslesen – ohne die Bücher auslagern und aufklappen zu müssen.

Folgende offene Fragen verblieben als Fazit des Workshops:

- Wann ist ein Objekt überhaupt als „digitalisiert“ zu bezeichnen?
- Wie kommt man in kurzer Zeit und ohne großen finanziellen Aufwand zu einigermaßen umfangreichen und einheitlichen Metadaten?
- Wird die Ausweitung ortsbezogener Informationen (sowohl die Lokalisierung des jeweiligen Internetnutzers, wie auch der vorhandenen historischen und aktuellen Informationen) das Internet und seine Nutzung vollständig verändern?

Zum Schluss wurde die Hoffnung geäußert, dass ein derartiger kollegialer Austausch sich in naher Zukunft wiederholen wird.

Berlin, 24. Juli 2008
Dr. Benedikt Goebel

Architekturmuseum der TU Berlin
Straße des 17. Juni 152, 10623 Berlin
Tel. +49 (0)30 314-25010
Fax +49 (0)30 314-79492
architekturmuseum@ub.tu-berlin.de