



ARTIKEL

2.1. Theoretische Grundlagen einer digitalen Quellenkritik

von [Althage, Melanie](#); [Feichtinger, Moritz](#); [Siebold, Anna](#);

Version 1.1 | Veröffentlicht 29. Juli 2024

Die technologischen Entwicklungen seit der Mitte des 20. Jahrhunderts führten zu einem tiefgreifenden Medienwandel. Durch die umfangreiche Digitalisierung hat sich nicht nur die Art und Weise, *wie*, sondern auch das, *womit* bzw. *worüber* wir forschen, also unsere Quellen, grundlegend gewandelt (Hiltmann 2022). In immer größerer Zahl stehen physische Quellen als Digitalisate repräsentiert orts- und zeitunabhängig online zur Verfügung. Mit sogenannten Born Digitals, das heißt genuin digitalen Quellen, wie bspw. Websites, Social-Media-Daten, Spielen und neuerdings prompt-basierten artifiziell generierten Texten, Bildern oder Videos kommen schließlich gänzlich neue Quellentypen mit je spezifischen Eigenschaften hinzu – ein Umstand, der insbesondere die in den Geschichts- und Geisteswissenschaften verankerte klassische Quellenkritik (siehe Infokasten) vor Herausforderungen stellt. Um beispielsweise die Originalität und Glaubwürdigkeit einer Quelle bewerten zu können und damit die Grundlage für die anschließende Analyse und Interpretation zu legen, ist die Prüfung der Authentizität von zentraler Bedeutung. Doch lässt sich bei digitalen Quellen nicht in gleichem Maße auf traditionelle Authentizitätsmarker wie etwa physische Merkmale zurückgreifen (z. B. Beschreibstoff, Beglaubigungsmittel etc.). Stattdessen bzw. ergänzend werden andere Indikatoren benötigt, wie Metadaten, Änderungshistorien oder Signaturen – sofern diese Informationen zugänglich sind. Hinsichtlich der inhaltlichen Merkmale wie des sprachlichen Stils oder Wortschatzes sind jüngst durch die Produkte generativer “Künstlicher Intelligenzen” (z. B. ChatGPT oder Midjourney) neue Herausforderungen hinzugekommen. Diese Anwendungen sind in der Lage, potentiell text- und bildbasierte Outputs zu generieren, die von menschlichen Produktionen nicht oder nur sehr schwer zu unterscheiden sind und mithin auf den ersten Blick als authentisches Material eingestuft werden können. Das kann dann problematisch werden, wenn durch die generierten Produkte Wirklichkeitsbezüge hergestellt werden, die jedweder Grundlage entbehren. Ohne entsprechende technische Kompetenzen, Werkzeuge und eine kritische Medien- und Methodenkompetenz sowie eine Erweiterung der Quellenkritik können wir diese Herausforderungen nicht bewältigen.

Dieses Kapitel bietet daher zunächst einen Überblick über die nach Niels Brügger ausdifferenzierten drei Typen digitaler Objekte, mit denen wir in den Geisteswissenschaften durch die umfassende Digitalisierung konfrontiert sind: Digitalisate, Born Digitals und Reborn Digitals (siehe Brügger 2012). Ausführlich werden sodann die wesentlichen Eigenschaften digitaler Objekte kritisch dargelegt, die sie von physischem Kulturgut unterscheiden. Auf Basis dieser Eigenschaften sind abschließend die Fragen einer digitalen Quellenkritik abzuleiten, die sich zwar an den klassischen Konzepten orientieren, aber entsprechend der technologischen Bedarfe weiterzuentwickeln sind. Damit soll dafür sensibilisiert werden, dass zur Wahrung der Kritikfähigkeit und für einen souveränen, reflektierten Umgang mit digitalisierten und genuin digitalen Quellen einerseits ein tiefgreifendes Verständnis der digitalen Objekte als eigenständige Quellentypen notwendig ist, andererseits aber auch ein Verständnis der zugrundeliegenden Technologien, Datenformate, Kodierungen, Metadaten oder auch von Konzepten wie der digitalen Forensik.

Infokasten: Die klassische Quellenkritik – Ein Kurzüberblick

Die Methode der Quellenkritik selbst ist eine seit mehr als 2.500 Jahren geübte und in vielen Stadien entwickelte Praxis, deren Beginn vermutlich die ersten Homerrezensionen im archaischen und klassischen Griechenland gewesen sind. Mit dem Hellenismus setzte eine reichhaltige Tradition der antiken Textkritik ein, die sich nicht auf Homer beschränkte, sondern kritische Ausgaben Hesiods, Pindars, eventuell auch des Anakreon, zusammenstellte, wobei man mit der Entwicklung einer methodisch begründeten Textkritik versuchte, Zufall und Willkür der Textüberlieferung zu berücksichtigen, vor allem immer im Hinblick auf die Frage, was genau der echte Wortlaut der Texte gewesen ist (Zimmermann und Rengakos 2014). Parallel dazu entwickelte sich aus Vergleich und theoretischer Reflexion insbesondere in der Philosophie und Historiographie eine Tradition von Kritik und Kommentierung, die sich etwa in den zahlreichen antiken Kommentaren zu Werken Platons und Aristoteles ebenso niedergeschlagen hat wie in der antiken Bibelexegese, den Auslegungsstrategien der römischen Rechtsliteratur und der nach dem Fall von Byzanz im westlichen Abendland seit den Jahren um 1500 neu einsetzenden Editionspraxis antiker Quellen.

Die noch heute gültigen Grundlagen zur Prüfung des Erkenntniswerts historischer Quellen wurden von den Vertretern des Historismus im 19. Jahrhundert entwickelt. Unter dem vom Renaissance-Humanismus übernommenen Motto „ad fontes“ stützten sie sich in besonderer Weise auf historische Quellen als erkenntnisbildende Grundlage, wobei sie auch von den Einsichten der sich seit dem 17. und 18. Jahrhundert herausbildenden Historischen Hilfswissenschaften profitieren konnten (Jordan 2018, 33–36, 41–45; Baberowski 2005, 63–79). Insbesondere die von Johann Gustav Droysen (1808–1884)

in seiner „Historik“ ausgearbeitete historische Methode mit dem typischen Dreischritt aus Heuristik, Kritik und Interpretation ist hier zu nennen (Droysen 1868). Zu den Grundlagen der Quellenkritik zählen noch heute die Prüfung der Echtheit respektive Originalität (*discrimen veri ac falsi*) und der Unversehrtheit/Vollständigkeit sowie der Authentizität, ebenso die Untersuchung des Überlieferungszusammenhangs/der Provenienz, wozu auch der zeithistorische Kontext zu zählen ist, die Klärung der Autor:innenschaft und die damit verbundenen Fragen nach dem Wissenshorizont, der Tendenz und Intention, aber auch – soweit erkenn- und nachvollziehbar – die Rezeption der Quelle (Droysen 1868, 16–19; Arnold 2002, 255–257; Emich 2019, 40–42).

Typen digitaler Quellen

Einige wesentliche Aspekte, die im Rahmen einer quellenkritischen Auseinandersetzung mit digitalen Forschungsgegenständen zu berücksichtigen sind, lassen sich anhand der Unterscheidung von drei Typen digitaler Objekte und der Berücksichtigung ihrer jeweiligen spezifischen Eigenschaften aufzeigen: Digitalisate, Born Digitals, und Reborn Digitals (Brügger 2012).

Digitalisate

Die digitale Form eines analogen Objekts, ein Digitalisat, ist das Produkt eines Digitalisierungsvorgangs, auch “Retrodigitalisierung” genannt. Grundsätzlich besteht ein Digitalisat aus einer Repräsentation eines ursprünglich nicht-digitalen Objekts in Form von Binärziffern (“binary digits”) einer bestimmten Codierung folgend, d. h. letztlich einer Kette von Einsen und Nullen. Weit verbreitete Digitalisate sind etwa Scans von Büchern und Dokumenten, welche in ihrer digitalisierten Form entweder lediglich als Bild oder inklusive ihres semantischen Gehalts, also des Textes, wiedergegeben werden können, die digitale Kodierung von Videos oder Audio-Aufnahmen etwa in Formaten wie MPEG und MP3, digitale Abbilder von Zeichnungen, Gemälden und Fotografien in verschiedensten Auflösungen und Formaten, aber auch die dreidimensionale Vermessung und Repräsentation von Artefakten, Bauwerken oder Landschaften und Topographien.

Es ist wichtig zu betonen, dass es sich bei Digitalisaten nicht um digitale Kopien bestehender Objekte handelt, sondern um Repräsentationen, die auf ein Primärobjekt verweisen und eine endliche Anzahl von dessen Eigenschaften in einer spezifischen (aber nicht notwendigen) Weise abbilden. Mit anderen Worten: Digitalisierung ist immer mit der Entscheidung verbunden, welche Eigenschaften eines Primärobjekts zu repräsentieren und wie diese zu beschreiben sind. Wie etwas repräsentiert wird, geht mit der Wahl einer Kodierung und eines Repräsentationsschemas einher. Verschiedene Kodierungen entsprechen verschiedenen digitalen Formaten: so lässt sich ein historisches Dokument etwa als durchsuchbarer Text und/oder als Bild des Originals repräsentieren. Das Repräsentationsschema legt dabei fest, welche Eigenschaften jeweils auf welche Weise dargestellt werden; bei einem Abbild kann dies etwa der Farbraum und die Auflösung sein, bei Texten der Standard und die Detailtiefe der jeweiligen Annotation. Dabei spielen der Umfang des Digitalisats (entscheidend für Speicherplatz und Übertragungsgeschwindigkeit) oftmals eine ebenso große Rolle wie die Originaltreue oder die umfassende Beschreibung des Primärobjekts. Entsprechend beinhaltet die Digitalisierung stets auch den Verlust einiger Eigenschaften, wie etwa des Geruchs und der Haptik von Urkunden, der Textur von Artefakten oder der Details von Bildern, Ton- und Filmaufnahmen.

Im Kontrast zu diesem inhärenten Verlust von Informationen geht das Digitalisieren insbesondere im wissenschaftlichen Kontext häufig mit einer Anreicherung des Digitalisats um Zusatzinformationen wie Metadaten, Markup und verschiedensten Annotationen einher. Damit lassen sich den Digitalisaten eine große Anzahl von weiteren beschreibenden Informationen hinzufügen, die zur Charakterisierung der Primärobjekte, ihrer Kontextualisierung und Weiterverwendung nützlich sein können. Metadaten einer Urkunde etwa können Informationen zur Autor:innenschaft, zum Entstehungszeitraum oder -ort, aber auch zu dem Material oder dem Format enthalten, die dem Original nur schwer zu entnehmen wären. Annotationen erlauben es zudem, beispielsweise die Struktur eines Dokuments abzubilden (Abschnitte, Absätze, Kapitel, Überschriften, etc.). Eine Annotation von Digitalisaten durch Markup (“tagging”) erlaubt weitere Verfahren der Analyse, der Verknüpfung und des Vergleichs. Linguistische Annotationen von Texten etwa (“Part of Speech”) lassen stilometrische Aussagen zu oder ermöglichen eine automatisierte Übersetzung in andere Sprachen; die Auszeichnung von signifikanten Personen, Orten, Institutionen, Daten, etc. (“Named Entity Recognition”) eröffnet die Möglichkeit zu automatisierter Inhaltsanalyse und Verknüpfung mit anderen Forschungsdaten (etwa durch “Topic Modeling”); auch Geoinformationen, mit denen historische Bilder oder Fotografien annotiert werden, erlauben weitergehende Untersuchungen und Querverbindungen.

Diese Prozesse gehen weiter über ein bloßes Einscannen der Primärobjekte hinaus, weshalb in diesem Kontext auch von Verdatung gesprochen wird: Verdaten bedeutet, dass Materialien „in einer Weise codiert werden müssen, die grundsätzliche Entscheidungen seitens der Forschenden verlangt, was jeweils zu markieren, zu annotieren ist, wie Informationsgehalte zu hierarchisieren sind in dem Sinne etwa, was als ‚Text‘ und was als ‚Metadaten‘ aufzufassen ist und ähnliches mehr“ (Krämer 2018, 6). Die Verdatung bedeutet somit eine Reihe von editorischen Zusatzarbeiten, in denen die Quellen möglichst umfassend annotiert und beschrieben werden, um einen Großteil ihrer (auch paratextlichen) Eigenschaften zu erfassen. Eine formalisierte Beschreibung und Annotation sowie die Beachtung von

Standards – im Falle digitaler Editionen etwa nach den Leitlinien der Text Encoding Initiative – dienen dazu, einer wiederum formalisierten und automatisierten Auswertung großer und heterogener Bestände Vorschub zu leisten und die Verknüpfung von digitalisierten Quellen und ihrer jeweiligen Metadaten untereinander (im Sinne von Linked Data) zu gewährleisten (Sulé und Lapeyra 2016; Berners-Lee, Hendler und Lassila 2001; W3C 2014; Berners-Lee 2006). Dieser Vorgang kann mit einem hohen Arbeitsaufwand verbunden sein, Fachwissen voraussetzen und mündet in (oftmals überschaubare) Sammlungen von Daten, die Christof Schöch als „smart data“ bezeichnet hat (Schöch 2013). Ein prototypisches Beispiel dafür sind digitale Editionen, die über die vollständige Transkription eines Textes hinaus Metadaten, Markup, weiterführende Informationen und Verlinkungen im Text oder zu externen Inhalten enthalten. Die Retrodigitalisierung von historischen Artefakten führt zudem idealerweise zu einem erleichterten Zugang zu ihnen, da ihre digitalen Kopien meistens über das Internet und die Portale der Gedächtnisinstitutionen der Öffentlichkeit zur Verfügung gestellt werden.

Born Digitals

Von Digitalisaten zu unterscheiden sind Quellen und Artefakte, die originär digital entstehen und sowohl durch digitale Medien dargestellt werden müssen als auch primär für die Weiterverwendung und Nutzung/Bearbeitung mit digitalen Medien vorgesehen sind. Sogenannte Born Digitals stellen eine heterogene und stetig wachsende Klasse möglicher Quellen für die Geschichtswissenschaft dar. Diese Klasse kann verschiedene Objekte umfassen, etwa digitale Fotografien und Bilder, Audio- und Videodateien, sämtliche Formen digitaler Kommunikation (Chats, E-Mails, Textnachrichten), digitale Publikationsformen wie Websites, Blogs, Podcasts etc. sowie Computerspiele, Software, Betriebssysteme, Protokolle usw. Originär digitalen Objekten ist gemeinsam, dass sie – nach einer weit verbreiteten Definition durch Kenneth Thibodeau (Thibodeau 2002) – drei Dimensionen bzw. Eigenschaftsklassen aufweisen: Zum einen sind sie stets an einen konkreten *physikalischen* Speicherort gebunden. Ob es sich dabei um die Ionisierung mikroskopisch kleiner Sektoren eines mechanischen Magnetspeichers wie etwa von Festplatten (Hard Drive Disk, HDD) oder um eine Mikrospannung in einem Transistor auf Flash-Speichern (Solid State Drive, SSD) handelt, digitale Daten haben (mindestens) einen physisch manifesten Ort, dem sie „eingeschrieben“ sind (Kirschenbaum 2012, 10). Darüber hinaus haben Born Digitals *formal-logische* Eigenschaften. Damit wird ihr spezifisches Format bezeichnet und festgelegt, in welcher digitalen Umgebung und mit welchen weiteren digitalen Objekten sie auf welche Weise interagieren können: Um eine Datei eines bestimmten Typs und Formats öffnen zu können, ist eine bestimmte Software notwendig, und um eine bestimmte Software nutzen zu können, müssen wiederum bestimmte Anforderungen an das Betriebssystem erfüllt sein. Die *konzeptionelle-semantische* Dimension schließlich beschreibt ihren eigentlichen, für Menschen erkennbaren, Inhalt und Sinn.

Diese drei Dimensionen lassen sich anhand einer digitalen Fotografie veranschaulichen: Wenn wir mit einem Smartphone ein Foto aufnehmen, handelt es sich dabei um eine Datei, die an einem spezifischen physischen Ort abgelegt ist, sei es auf dem integrierten Flash-Speicher des Smartphones oder in der Cloud, also einem entfernten Rechenzentrum, in das die Datei via Internet übertragen wird. Nur selten wird die Datei dabei als Ganzes gespeichert und übertragen, oftmals wird sie in kleinere Elemente zerteilt und entsprechend der Logik ihres Formats und des jeweiligen Betriebssystems abgelegt. Zudem kann das Foto, je nach Format und Aufnahmesoftware, eine größere oder kleinere Auflösung haben, mithilfe von Filtern bereits bearbeitet aufgenommen worden sein und unterschiedlich umfangreiche Metadaten etwa zu Aufnahmezeitpunkt, Ort, Belichtung etc. enthalten. Die formal-logischen Eigenschaften des Fotos – seine (digitale) Ursprungsumgebung und seine Formatierung etwa – bestimmen, wie es dargestellt und weiterverarbeitet werden kann. Nicht jede Software kann jedes Format darstellen, aber manche Programme und Anwendungen wiederum ermöglichen umfangreichere Formen der Manipulation, Analyse, Weiterverarbeitung, Rekombination etc. Von den vorgenannten Ebenen gänzlich unberührt bleibt der semantisch-konzeptionelle Gehalt des Fotos, also was genau darauf abgebildet ist, etwa eine Katze, eine Straßenszene, ein Porträt oder auch das Abbild einer weiteren Fotografie. Diese drei Ebenen sind für eine Quellenkritik von originär digitalen Objekten von großer Bedeutung.

Reborn Digitals

Einen von Digitalisaten und Born Digitals zu unterscheidenden dritten Typus digitaler Quellen stellen Reborn Digitals dar, die verändert oder aufbereitet wurden und außerhalb ihres ursprünglichen Kontextes bzw. ihrer informationellen Umgebung existieren. Niels Brügger, der diese Unterscheidung eingeführt hat, prägte dafür den Begriff „reborn-digital material“ (Brügger 2012, 104). Brügger hat Konzept und Begriff der Reborn Digitals ursprünglich ausgehend von archivierten Websites entwickelt, wie sie vor allem das Internet Archive, aber zunehmend auch staatliche Archive, zur Verfügung stellen. Solche archivierten Websites können als erstarrte oder tote Abbilder ihrer Originale verstanden werden: Sie mögen zwar ihr ursprüngliches Aussehen und ihren semantischen Inhalt bewahrt haben, sind aber nicht mehr online aufzufinden; sie haben ihren physikalischen Ort gewechselt. Vor allem sind sie ihrer ursprünglichen formal-logischen Funktionalität und Umgebung beraubt: Die Hyperlinks funktionieren möglicherweise nicht mehr, eingebettete Inhalte sind zum Teil nicht mehr abrufbar und selbst das Aussehen der Websites kann durch die Darstellung in modernen Browsern verzerrt sein. Reborn Digitals

können aber auch andere konservierte und migrierte digitale Objekte sein, etwa Datensätze, die aus obsoleten Datenbank-Programmen extrahiert wurden, Software und Computerspiele, die nur noch auf emulierten Betriebssystemen lauffähig sind, sowie generell digitale Objekte, die unter Verlust ihrer ursprünglichen Funktionalität und Interaktivität in aktuellere Formate konvertiert wurden, meist, um ihren semantisch-konzeptionellen Inhalt zu bewahren.

Eigenschaften Digitaler Quellen

Digitale Objekte weisen eine Reihe von Eigenschaften auf, die sie von nicht-digitalen Quellen und Artefakten unterscheiden (siehe z. B. Föhr 2019, 54–64, 94–97). Diese Eigenschaften sind in einer digitalen Quellenkritik zu prüfen.

Interaktivität und Manipulierbarkeit

Zu den wesentlichen Merkmalen digitaler Objekte gehören ihre Interaktivität und einfache Manipulierbarkeit (hier in neutralem Sinne). Gerade weil sich mit digitalen Schreibprogrammen verfasste Texte einfach verändern lassen oder digitale Fotos mit entsprechenden Bildbearbeitungsprogrammen leicht nachbearbeitet werden können, erfreuen sich digitale Objekte großer Beliebtheit. Ihre logisch-formalen Eigenschaften erlauben spezifische Interaktionen mit ihnen durch Menschen oder Software, welche wiederum neue digitale Objekte hervorbringen können. Für eine digitale Quellenkritik ist es daher unerlässlich zu fragen, welche Formen der Interaktion in einem digitalen Objekt angelegt sind oder einer typischen Nutzung des Objekts entsprechen. Ein triviales Beispiel dafür sind PDF-Dateien, die ausfüllbare Felder enthalten können oder schreibgeschützt sind. Programme und Applikationen, die zur Bearbeitung von Texten, Bildern, Video- und Audiodateien gedacht sind, interagieren mit speziell für diesen Zweck erstellten Dateiformaten. Für den Austausch oder die Publikation hingegen werden Dateiformate erstellt, die eine Be- und Überarbeitung nicht vorsehen.

Nicht jede Interaktion bedeutet gleichzeitig Manipulation. Das hängt davon ab, ob das Original erhalten bleibt und das Ergebnis der Interaktion als zusätzliche Datei abgelegt wird oder ob die Originaldatei mit ihrer veränderten Version überschrieben wird. Eine wichtige Frage bei der digitalen Quellenkritik ist demnach, ob etwaige Interaktionen und Manipulationen dem Objekt eingeschrieben sind, etwa in einer Versions-Historie, Log-files etc. Gemeinhin werden solche Dokumentationen der Weiterverarbeitung angelegt, wenn eine Datei neu abgespeichert wird oder Versionen explizit benannt werden. Wird jedoch eine Kopie des Originals erstellt, bearbeitet und schließlich die Ausgangsdatei gelöscht, so lassen sich Manipulationen nicht mehr ohne weiteres feststellen. Ein anschauliches Beispiel dafür ist der Bearbeitungsmodus in Texteditoren.

Flüchtigkeit und Volatilität

Digitale Objekte haben einen inhärent flüchtigen, volatilen und ephemeren Charakter. Dieser Charakter ergibt sich aus allen drei Eigenschaftsklassen. Zunächst lassen sich Objekte, die nur als Binärcode unter einer bestimmten Adresse im Speicher eines Computers gespeichert sind, einfach löschen, indem die entsprechende Speicheradresse aufgelöst oder neu überschrieben wird.

Genau wie nicht-digitale Quellen sind auch digital born Objekte nicht mehr lesbar, wenn die Speichermedien verloren gehen, zerstört werden oder mit der Zeit zerfallen und degradieren. Allerdings sind digitale Objekte und Medien darüber hinaus zu ihrer Auslese an Maschinen gebunden. Dateien und Software aus der Frühzeit des digitalen Zeitalters etwa, die auf Magnetbändern oder Floppy Disks gespeichert wurden, erfordern Zugang zu entsprechenden historischen Peripheriegeräten, um ausgelesen werden zu können.

Die bereits geschilderte Gebundenheit digitaler Objekte an logisch-formale sowie physikalisch-materielle Infrastrukturen und Umgebungen und die rasende Weiterentwicklung der Informationstechnologien führen außerdem zu einer beschleunigten technischen Obsoleszenz auf der formal-logischen Ebene. Der frühe Tod des ehemals ubiquitären Flash-Players im Internet ist dafür ein gutes Beispiel, ebenso wie auf obsoleten Trägermedien wie Floppy-Disks und Magnetbändern gespeicherte Inhalte (Rothenberg 1999; 2000). Die Bibliotheks- und Archivwissenschaften stemmen sich mit Versuchen der Etablierung von Standards zur Langzeitarchivierung gegen diese „digitale Amnesie“, die nicht nur durch technische Obsoleszenz verursacht wird, sondern auch durch Fehler und Qualitätsreduktion beim Kopieren, deren ohnehin hohe statistische Häufigkeit durch jeden weiteren Kopiervorgang noch einmal zunimmt (Baucom 2019; Ries und Palkó 2019).

Ohne einen entsprechenden Aufwand in der Pflege, regelmäßige Aktualisierung und letztlich ihre Archivierung verschwinden Webinhalte beispielsweise unwiederbringlich in dem Augenblick, in dem Server vom Netz gehen.^[1] Ein instruktives Beispiel dafür ist der Webhosting-Service GeoCities, der nach 25-jähriger Existenz wegen mangelnder Lukrativität eingestellt wurde. Damit löste sich ein wichtiges Element der Frühgeschichte des Webs von einem Tag auf den anderen einfach auf und die Bewahrung seiner Überreste wird nur von privaten Initiativen gewährleistet.^[2]

Kopierbarkeit

Ein weiteres Merkmal digitaler Objekte ist ihre einfache und nahezu beliebig wiederholbare Kopierbarkeit. Digitale Datenverarbeitung erlaubt es, durch nur wenige und standardisierte Befehlskombinationen eine beliebige Anzahl scheinbar identischer Klone von Dateien anzulegen. Bei diesem Vorgang bleiben sowohl die formal-logischen Eigenschaften als auch der semantische Inhalt digitaler Objekte erhalten, es ändert sich lediglich ihr physikalischer Speicherort – einerseits für die Nutzenden durch die Zuweisung eines neuen Dateinamens oder Verzeichnisses, andererseits für das Betriebssystem durch die Vergabe einer neuen Speicheradresse. Kopien digitaler Objekte werden nicht nur bewusst durch menschliche Nutzung, sondern noch weitaus häufiger durch Programme und Betriebssysteme automatisiert erstellt. Durch mehrfache Redundanz und in regelmäßigen Zeitabständen erstellte temporäre Sicherungskopien gewähren digitalen Datenverarbeitungssystemen einerseits die Kohärenz und Integrität von Dateien angesichts der Gefahr eines potentiellen Verlusts bei ihrer Verarbeitung und Übertragung, andererseits aber auch die mögliche Rückgängigmachung bereits getätigter Manipulationen und die Wiederherstellung des vorhergehenden Zustands.

Multidimensionalität

Digitale Objekte sind stets multidimensional in dem Sinne, dass ihr (semantischer) Inhalt in anderer Form (Codierung), teilweise an einem anderen (Speicher-)Ort und mit anderem Gehalt und Umfang existiert als diejenigen Objekte, die auf der formal-logischen Ebene interagieren. Ein Beitrag auf einem Weblog zum Beispiel existiert als Inhalt (Text), als visuelles Erscheinungsbild (Textgestaltung, Struktur und Abbildungen), wie es von verschiedenen Browsern (unter Umständen in unterschiedlicher Weise) dargestellt wird, und als zugrundeliegender Quellcode mitsamt anderer Dateien, die vom Content Management System verwaltet werden. Es ist evident, dass diese verschiedenen Dimensionen jeweils unterschiedliche, potentiell quellenkritisch relevante, Informationen enthalten können. Die digitale Quellenkritik sollte die Bezüge und Unterschiede der verschiedenen Dimensionen stets mit analysieren, bzw. die Unzugänglichkeit oder das Fehlen von einzelnen Dimensionen deutlich ausweisen und die interpretatorischen Konsequenzen reflektieren und erörtern.

Konkret ist also stets zu fragen, auf welchem Speichermedium sich ein digitales Objekt ursprünglich befand und ob diese materielle Dimension Größe, Formatierung und Aufbau des Objekts beeinflusste. So sind etwa auf Magnetbändern gespeicherte Dateien nur sequentiell abrufbar, über begrenzte Bandbreiten zu übermittelnde Objekte oft in Pakete unterteilt und komprimiert etc. Desgleichen ist zu überlegen, wie die ursprüngliche Umgebung einer originär digitalen Quelle beschaffen war und wie sie sich auf ihre Beschaffenheit und Bearbeitbarkeit auswirkt. Bei historischen Videospielen ist etwa zu bedenken, welche Leistung die CPU erbringen musste, um die Bewegtbilder und Interaktionen flüssig darzustellen, ob einzelne Elemente sukzessive in den Speicher geladen werden mussten, welche Peripheriegeräte (wie Bildschirm und Controller) verfügbar und notwendig waren. Schließlich ist nach der Integrität und Authentizität des Objekts zu fragen, also ob sich Manipulationen und Verzweigungen feststellen lassen oder nicht. In der Praxis werden sich nur selten in Bezug auf alle drei Dimensionen Aussagen treffen lassen. Gerade dann aber ist das Fehlen solcher Kontext- und Metainformationen festzuhalten; eventuelle Implikationen für den Aussage- und Quellenwert müssen reflektiert und markiert werden.

Medialität/Materialität

Originär digitale Objekte haben, wie herkömmliche Quellen und ihre Digitalisate auch, eine bestimmte Medialität. Im Unterschied zu nicht-digitalen Quellen können sie allerdings leichter ihren Speicherort und damit ihre Materialität ändern: Sie lassen sich kopieren und verschieben und können damit auf verschiedenen Speichermedien und Materialien existieren, wie etwa auf Magnetspeichern (etwa Festplatten, Disketten, Magnetbändern etc.), auf optischen Speichern (CD-ROMs, DVDs) sowie auf Halbleiterspeichern in ihren verschiedensten Formen. Auf all diesen Speichermedien können sie wiederum vollständig oder teilweise, dauerhaft oder temporär abgelegt sein. Diese leichte Verschiebbarkeit verdankt sich letztlich dem Umstand, dass digital kodierte Informationen, wie oben ausgeführt, mühelos kopierbar sind. Daher ist auch ihr konzeptuell-semantischer Gehalt tendenziell leichter veränderbar, als dies bei nicht-digitalen Quellen der Fall ist, und solche Veränderungen lassen sich schwerer nachvollziehen. In der Regel werden bei modernen Computersystemen zum Beispiel der Inhalt einer Textdatei (der Text selbst), die Metadaten (Erstellungsdatum, Zeitpunkt der letzten Überarbeitung, Sprache, Umfang etc.) und die Formatierung (Fett, Kursiv, Blocksatz, Schriftart und -größe etc.) getrennt gespeichert. Eine Feststellung der Originalität oder Bearbeitung lässt sich daher nur dann mit Sicherheit treffen, wenn sämtliche dieser Teildateien zugänglich sind, das Objekt also vollständig vorliegt, d. h. dessen Integrität gewährleistet ist.

Gleichwohl kann die materielle Dimension digitaler Objekte die Chance zur kritischen Prüfung von Authentizität, Vollständigkeit und Provenienz bieten, denn die Materialität physikalischer Speichermedien beeinflusst die formal-logischen Eigenschaften digitaler Objekte. Ob Daten sequentiell abgelegt wurden oder in willkürlich durch ein System adressierten Sektoren, liefert Anhaltspunkte in Bezug auf den ursprünglichen Speicherort von Dateien. Die zu Beginn des Internetzeitalters stets limitierte Bandbreite hinterließ spezifische Spuren in

Form der Komprimierung und Reduktion von Datensätzen. Die Möglichkeit der Einbettung von Medieninhalten und die dynamische Darstellbarkeit von Webinhalten auf verschiedenen Endgeräten wiederum erlaubt Mutmaßungen zum Entstehungszeitraum von digitalen Objekten.

Quellenkritische Fragen

Da die Verdatung respektive Datafizierung von Forschungsgegenständen im zuvor behandelten Sinne mit Eingriffen und Entscheidungen verbunden ist, die sich wiederum auf deren digitale Gestalt auswirken, ergeben sich für die historische Quellenkritik eine Reihe von Fragen. Diese zielen zum einen auf den Prozess der Digitalisierung selbst ab und beziehen sich auf die mediale Ebene der Überlieferung. So ist danach zu fragen, mit welchem Programm ein Objekt gescannt oder fotografiert wurde, ob dabei Bildfehler, z. B. durch fehlerhafte Kalibrierung, entstanden sein könnten und ob das Objekt in einem Format vorliegt, das heute noch maschinenlesbar oder möglicherweise bereits veraltet ist. Zum anderen ist in Bezug auf den Entstehungskontext die inhaltliche Ebene der Überlieferung zu berücksichtigen, etwa welches Interesse und welcher Kontext sich hinter der Digitalisierung und Verdatung verbergen, welche Metadaten hinzugefügt oder bewusst weggelassen wurden und welche in der Quelle implizit enthaltenen Informationen möglicherweise nicht expliziert wurden (und somit nicht maschinenlesbar und auffindbar sind). Im Folgenden sollen diese angerissenen quellenkritischen Fragen genauer in den Blick genommen und mögliche Lösungsansätze diskutiert werden. Allgemein gilt dabei wie für eine nicht-digitale Quellenkritik auch, dass das Nachdenken über Gehalt, Herkunft und Glaubwürdigkeit einer Quelle zu Erkenntnissen führen kann, auch wenn einzelne Fragen offenbleiben – oder gerade dann. Darüber hinaus ist für den Fall der digitalen Quellenkritik zu konstatieren, dass dieselben Eigenschaften, die ihre quellenkritische Überprüfung erschweren, gleichzeitig Anhaltspunkte für ihre Kritik und Kontextualisierung bieten können.

Autor:innenschaft

Eine zentrale Frage der Quellenkritik ist die nach der Autor:innenschaft; also: wer hat die Quelle erstellt? Die Klärung der Autor:innenschaft einer digitalen Quelle wird durch mehrere Eigenschaften digitaler Objekte sowie durch die spezifischen Verfahren der Wissensproduktion im digitalen Zeitalter erschwert. Viele digitale Texte, speziell im Web publizierte, werden kollaborativ von mehreren Autor:innen verfasst (wie dieses Handbuch) oder nicht namentlich gekennzeichnet. Die Einbeziehung von Künstlicher Intelligenz bei der Produktion von Inhalten oder zum Beispiel die Nutzung automatisierter Übersetzung stellen eine weitere Problematisierung der Autor:innenschaft im herkömmlichen Sinne dar. Während sich bei einigen Texten und Bildern vielleicht noch Ansprüche auf Urheberschaft durch identifizierbare Personen finden, stellen andere, etwa solche, die mit generativer KI erstellt wurden, die Frage der Urheberschaft vor völlig neue Probleme. Software wie etwa Computerspiele und Anwendungsprogramme oder gar Betriebssysteme werden wiederum gemeinhin von großen Teams und kontinuierlich entwickelt, eine Urheberschaft einzelner Elemente lässt sich unmöglich einzelnen Personen zuweisen.^[3] Die vorgenannte Eigenschaft der Interaktivität digitaler Objekte hat zudem zur Konsequenz, dass so manche digitale Quelle erst durch Interaktion entsteht (gewissermaßen “emergiert”), ohne von einem Subjekt im herkömmlichen Sinne verfasst worden zu sein.

Abfrageergebnisse von Datenbank-Anwendungen etwa können erst durch die Kombination mehrerer verschränkter Suchanfragen und Filter entstanden sein. Deutlichstes Beispiel hierfür ist eine Websuche: Die erste Ergebnisseite einer Google-Suche ist nur schwerlich einer konkreten Person zuzuweisen, weder die Suchenden noch der Google-Algorithmus können hier sinnvoll als Verfasser:in beschrieben werden. Vielmehr handelt es sich bei der ersten Ergebnisseite um das Resultat eines komplexen Zusammenspiels aus Suchbegriffen, Standort des ausführenden Browsers, verfügbaren Webinhalten zum Suchzeitpunkt und deren Sortierung durch den (geheimen) Google-Algorithmus. Eine damit zusammenhängende “Emergenz” neuer digitaler Objekte sind Inhalte, die durch Bots, also automatisiert ablaufende Mikroprogramme, generiert werden, wie etwa Wettervorhersagen, automatisierte Übersetzungen und Textrekombinationen, die einen nicht unerheblichen Teil der Inhalte des World Wide Web ausmachen.

Einen Sonderfall bildet eine falsch zugeschriebene Autor:innenschaft, die sich unter digitalen Vorzeichen ebenfalls leichter umsetzen und schwerer erkennen lässt, wie etwa im Fall von Plagiaten und Fake News. Wird im ersteren Fall die Urheberschaft und die zugrundeliegende kreative oder analytische Leistung von Personen beansprucht, welche die eigentliche Quelle gar nicht oder nur unzureichend ausweisen, so werden im Fall von Fake News Aussagen und Informationen bewusst Personen und Institutionen “in den Mund gelegt”, entweder um diese zu diskreditieren oder um die Glaubwürdigkeit der Fehlinformationen durch die falsche Angabe einer “seriösen” Quelle zu erhöhen.

Perspektivität, Standpunkt und Intention

Nicht anders als bei nicht-digitalen Quellen ist auch bei digital born Quellen und Artefakten nach der Perspektivität und der Intention ihrer Produktion und Publikation zu fragen. Dies lässt sich am Beispiel eines im Internet verbreiteten Fotos illustrieren, das etwa eine bekannte Person der Zeitgeschichte bei einer bestimmten Handlung zeigt. Aus der spezifischen Perspektive (hier ganz wörtlich zu verstehen im Sinne von Kamera- und Aufnahmewinkel) lässt sich ermitteln, von wo aus das Foto aufgenommen wurde, wie also das räumliche Verhältnis zwischen Bildobjekt und Kamera gewesen sein muss. Weitaus schwieriger zu überprüfen ist der Bildausschnitt. Nur anhand der oben ausgeführten Quellenkritik lässt sich abschätzen (oder auch nicht), ob das Bild nachträglich zugeschnitten wurde, um einen bestimmten Kontext oder ebenfalls im ursprünglichen Bild vorhandene Kontexte zu betonen oder zu verbergen. Die für die historische Einordnung wichtigste Frage bleibt aber auch im Fall des digital aufgenommenen, bearbeiteten und publizierten Bildes die Intention der an diesen Schritten jeweils beteiligten Personen. Warum erschien dieser Moment bildwürdig und handelte es sich um ein Einzelbild oder eine Bildserie, aus der ein ganz bestimmtes Bild ausgewählt wurde? Wer bestimmte den Ausschnitt und mit welcher Absicht und nach welchen Kriterien? Wer verantwortete die Publikation und was sollte damit erreicht werden? Sollte die abgebildete Person diskreditiert werden, in ein besonders positives Licht gerückt werden oder hatten das Bild und seine Veröffentlichung eher dokumentarischen Charakter? Gerade im Zeitalter der einfachen Manipulation und Verbreitung von Bildern im World Wide Web sind diese quellenkritischen Fragen, die zu den grundlegenden Werkzeugen der historischen Quellenkritik gehören, unverzichtbare Elemente bei der Einordnung und Bewertung von Elementen im öffentlichen Diskurs geworden.

Authentizität

Authentizität von digitalen Objekten im Sinne der Vollständigkeit und der Ursprünglichkeit (Integrität und Konsistenz) ihres Inhalts wird durch die einfache Kopierbarkeit schwerer zu überprüfen, schließlich lassen sich Manipulationen und Interpolationen nur über Metadaten prüfen, die das Objekt damit auch enthalten muss (Fickers 2021). Gerade beim Wechsel der logisch-informationellen Eigenschaften, also etwa der Konversion eines digitalen Bildes von RAW zu PNG, können Metadaten verloren gehen (in diesem Fall die EXIF-Daten). Nicht selten wird beim Kopieren auch komprimiert, das heißt, die Qualität und/oder Metadaten des ursprünglichen Objekts werden reduziert, um den Speicherbedarf für die Kopie(n) zu verringern. Gleiches gilt bei der Verschiebung auf ein anderes Speichermedium (also des Wechsels der logischen Eigenschaften und des physisch-materiellen Speichermediums), bei der Hinweise auf Provenienz und Authentizität oftmals verloren gehen, etwa indem Systemdateien wie Logfiles, Timestamps und File-Maintenance-Statements nicht mit übertragen werden oder in der neuen logischen Umgebung nicht mehr gelesen werden können. Auch indem etwaige physische Spuren auf dem eigentlichen Speichermaterial nicht kopiert werden, können digital-forensische Zugänge erschwert oder verunmöglicht werden. Die einfache und oftmals automatisiert ablaufende Vervielfältigung digitaler Objekte in zahllose Kopien kann jedoch gleichzeitig die Authentizitätsprüfung erleichtern, indem sich die Quelle mit ihren Klonen abgleichen lässt.

Mit der Frage der Authentizität eng verknüpft ist die Feststellung der “Originalität” eines digitalen Objekts. Wie Peter Haber festgestellt hat, kann „Singularität“ unter der Bedingung aufwandsloser und mitunter automatisiert ablaufender Kopien-Bildung digitaler Quellen kein Kriterium mehr für deren Originalität sein (Haber 2011, 108). Dennoch kann es durchaus Argumente dafür geben, eine digitale Quelle – auch wenn es sich um eine Kopie handelt – möglichst nah am Originalzustand untersuchen zu wollen, beziehungsweise es kritisch zu reflektieren, wenn sich dieser nicht mehr rekonstruieren lässt. Dies betrifft zunächst einmal die Medialität: Wie Zeitgenossen Informationen aufnehmen konnten und wie sie sich ihnen darstellten, kann durchaus von Bedeutung für die Rezeptionsgeschichte sein. Eine historische Website etwa, die sich nur nach dem Aufsuchen eines Internet-Cafés und nach verhältnismäßig langer Ladezeit in einem Browser der ersten Generation darstellen ließ, kann nicht ohne weiteres mit ihrer Kopie (oder ihrem Simulacrum) gleichgesetzt werden, die durch Portale wie etwa die Wayback-Machine angeboten wird. Über die Rekonstruktion oder zumindest die Reflexion des historischen “look and feel” hinaus geht die Prüfung der formal-logischen Eigenschaften. Die historischen Möglichkeiten und Begrenzungen der Weiterverarbeitung und Manipulation werden durch diese Eigenschaften und die jeweilige informationstechnische Umgebung (etwa in Form der jeweiligen Peripherie wie Interfaces, Drucker, Ausgabegeräte) determiniert und müssen in die quellenkritische Reflexion mit einfließen.

Vollständigkeit und Integrität

Eine weitere wichtige Frage der Quellenkritik betrifft die Vollständigkeit bzw. “Unversehrtheit” von historischen Überresten: Haben wir es mit Fragmenten oder einem vollständigen Artefakt zu tun, fehlen wichtige Elemente? Die leichte Manipulierbarkeit digitaler Objekte, einschließlich der leichten Löschbarkeit von Inhalten, kann die Überprüfung der Vollständigkeit digitaler Quellen erschweren. Die Frage, ob es sich um ein vollständiges Bild oder um einen Ausschnitt, um den vollständigen Text oder nur um einen Auszug handelt, lässt sich in der Regel nur durch einen Vergleich mit anderen Kopien klären, die ebenfalls auf dasselbe Original verweisen. Auch Metadaten können bei der Klärung dieser Frage hilfreich sein: Weichen Erstellungs- und letztes Bearbeitungsdatum deutlich voneinander ab, ist eine

Interpolation und damit eine mögliche Unvollständigkeit wahrscheinlicher. Für die Analyse und Kritik von digitalen Quellen bieten sich einige Methoden und Werkzeuge aus dem Arsenal der digitalen Forensik an (Ries 2022). Dieser noch sehr junge Zweig der Kriminalistik dient vor allem Ermittlungsbehörden dazu, Beweismittel auf digitalen Medien und Spuren von Verbrechen im digitalen Raum (von Kinderpornografie über Hacking und Erpressung bis hin zu Schadsoftware) zu sichern und auszuwerten. Einige solcher Methoden werden auch zur Rettung von Daten auf beschädigten Datenträgern und Systemen verwendet (Sammons 2014; Gogolin 2021). Sind neben einer einzelnen Datei weitere, auf sie verweisende, logisch-formale Informationen vorhanden, so lässt sich der Lebenszyklus des digitalen Objekts gegebenenfalls anhand von Logfiles nachvollziehen und somit lassen sich Löschungen und Kürzungen erkennen. Auch im Dateiverzeichnis automatisierte Hash-Werte (d. h. Prüfsummen) können herangezogen werden, um die Authentizität und Integrität von Dateien zu überprüfen. Ist die gesamte Umgebung des Primärobjekts – also das Betriebssystem und der Speicher im Original oder in Form eines Bitstream-Images – verfügbar, dann lassen sich eventuelle Löschungen forensisch feststellen, wenn das System die Speicheradressen der gelöschten Inhalte bereits freigegeben, aber noch nicht überschrieben hat und sich damit Fragmente der ungekürzten Datei finden lassen.

Forensische Zugänge eignen sich auch, um festzustellen, was eventuell gelöscht wurde oder welche Dateitypen ursprünglich auf einem Speichermedium vorhanden waren, denn jede Datei enthält einen Header (ein hexadezimal ausgedrückter Wert am Ende jeder Datei), der ihren Typ angibt. Ein MS-Office-Textdokument lässt sich zum Beispiel am Header D0CF11E0 erkennen, eine Datei im Rich-Text-Format am Header 7B5C727466 und ein PDF hat stets die Standard-Endung 255044462D312E (Willer 2012, 121). Obwohl computer-forensische Zugänge erhebliches Potential für die digitale Quellenkritik aufweisen, bedarf es gleichzeitig gewisser Voraussetzungen, die in der Alltagspraxis historischen Forschens selten gegeben sind. So muss zum Beispiel bestenfalls ein Abbild des gesamten Systems oder Speichers vorhanden sein, um sinnvolle forensische Aussagen über einzelne Dateien treffen zu können. Mehr noch, dieses System-Abbild (d. h. Bitstream-Image) muss mithilfe einer speziellen Software, eines Write-Blockers, erstellt worden sein, die es erlaubt, das System oder ein Speichermedium zu klonen, ohne dass diese Interaktion ihrerseits Spuren auf dem Original oder dem Klon (etwa in System-Logdateien) hinterlässt. Nicht zuletzt handelt es sich bei der Computer-Forensik um eine hochspezialisierte Disziplin, die auf umfassenden Kenntnissen von Computerarchitekturen und den Besonderheiten verschiedener Programme und Dateiformate beruht und weder leicht zugänglich noch leicht erlernbar ist.

Die Vollständigkeit eines digitalen Objekts bedeutet somit nicht nur die möglichst komplette Wiedergabe seines semantischen Gehalts, sondern bestenfalls eine möglichst umfassende Dokumentation seiner formal-logischen und materiellen Eigenschaften in Form eines Bitstream-Abbilds des Originals, seiner originären formal-logischen Eigenschaften und Umgebung sowie eine möglichst präzise und umfassende Dokumentation in Form von Metadaten. Fragen nach Vollständigkeit und Integrität müssen sich allerdings nicht nur auf die Beschaffenheit und mögliche Versehrtheit der vorgefundenen Quellen beziehen, sondern können auch in Bezug auf den Umgang mit Quellen in der eigenen Forschungsarbeit von Historiker:innen relevant werden. Welche Quellen oder Teile von Quellen konnten digital gesichert und somit überhaupt erst in die Arbeit eingeschlossen werden? Wie können Entscheidungen transparent gemacht werden, die zu einer bestimmten Form von Verdattung und Selektion führten? Und wie und wo können die digitalisierten bzw. originär digitalen Quellen gespeichert werden, damit sie auch für zukünftige Historiker:innen zugänglich sind? Vor diesem Hintergrund hat Pascal Föhr angeregt, den historischen Arbeitsprozess um einen weiteren Schritt, nämlich den der Quellensicherung, zu ergänzen. Ziel der Quellensicherung sei es, „ein digitales Objekt unter größtmöglicher Wahrung von dessen Integrität und Authentizität so zu speichern, dass es möglichst dauerhaft für die wissenschaftliche Forschung nachvollziehbar (wieder-)verwendet werden kann“ (Föhr 2019, 188). Diese Notwendigkeit spiegelt sich auch in der Forderung nach und Entwicklung von standardisierten Vorgaben, Verfahren und Infrastrukturen für das Forschungsdatenmanagement wider. Insbesondere die Wiederverwendbarkeit von digitalen Objekten stellt Forschende vor Herausforderungen, denn im Gegensatz zu physischen Forschungsausgaben, d. h. Büchern, Monographien und Artikeln, gelangen aufwendig verdatete Forschungsgegenstände oder Sammlungen originär digitaler Objekte nicht in Bibliotheken, sondern liegen verstreut als “distributed materiality” auf Servern und in veralteten Softwaresystemen (Drucker 2013; Blanchette 2011).

Die Frage, wie digitale Objekte beschaffen sind – unversehrt oder womöglich manipuliert, vollständig oder fragmentiert – stellt sich somit sowohl in der Konfrontation mit den jeweiligen vorgefundenen Quellen als auch in Bezug auf die eigene Arbeit mit ihnen und somit dahingehend, wie sie im Rahmen einer zukünftigen Quellenkritik bewertet werden können.

Provenienz

Ein möglichst lückenloser Nachvollzug der Überlieferungsgeschichte ist ein zentrales Element jeder Quellenkritik. So ist grundsätzlich zu klären, in welcher Form und über welche Stationen historische Quellen überliefert sind, ob sie bewusst oder zufällig hinterlassen wurden und wo sie aufbewahrt werden. Angesichts der leichten Kopierbarkeit und Manipulierbarkeit digitaler Objekte stellt die Frage der Provenienz für die digitale Quellenkritik eine besondere Herausforderung dar. Zum einen sind Forschende im Umgang mit digitalen

Quellen – wie auch oft bei nicht-digitalen Quellen der Fall – mit Parallelüberlieferungen konfrontiert. Wird die “Singularität” als maßgebliches Kriterium von “Originalität” verabschiedet, wie von Peter Haber vorgeschlagen (siehe Haber 2011, 106–108), so erlauben Parallelüberlieferungen das Schließen von Lücken im Provenienz-Zyklus und lassen sich somit vom Problem zum Vorteil wenden. Bis zu einem gewissen Grad lässt sich Provenienz auch prüfen (oder eher: anzweifeln), indem die Integrität und Konsistenz digitaler Objekte im Hinblick auf ihre formalen und materiellen Eigenschaften überprüft werden. Da die formalen Eigenschaften eines digitalen Objekts seine weitere Verarbeitbarkeit und auch Manipulierbarkeit bestimmen und darüber hinaus die Systematik und Materialität seiner Speicherung ebenfalls Größe und Formatierung beeinflussen, lassen diese beiden Eigenschaftsklassen Rückschlüsse auf die Wahrscheinlichkeit bestimmter Überlieferungsketten zu. Ähnlich wie bei den vorgenannten Instanzen der digitalen Quellenkritik kann es auch bei der Prüfung der Provenienz digitaler Objekte kaum darum gehen, unumstößlich und lückenlos eine Überlieferung zu verifizieren, sondern im Gegenteil Lücken im Überlieferungsnachweis klar zu benennen und die entsprechende Wahrscheinlichkeit von Interpolationen mit zu reflektieren.

Verfügbarkeit

Die Frage danach, welche digitalen Quellen überhaupt zur Verfügung stehen und nach welchen Kriterien sie digitalisiert wurden, gehört heute zu einer kritisch-reflexiven digitalen Quellenkritik. Die Kopierbarkeit und gleichzeitige Gebundenheit an informationelle Umgebungen bedingen auch die paradoxe Gleichzeitigkeit der Wahrnehmungen von einem Überfluss und einem Mangel an digitalen Quellen, die Fragen ihrer Provenienz und Aussagekraft berühren. Ohne Zweifel haben die vielfachen Digitalisierungsprojekte historischer Quellenbestände sowie die geradezu exponentielle Vermehrung der Kommunikationswege und Zugänge zu Informationen durch die Verbreitung des Internets zu einem „Zeitalter des Überflusses“ potentieller Quellen für die Geschichtswissenschaften geführt (Rosenzweig 2003; Milligan 2019). Damit stellt sich natürlich auch die Frage der Selektion und Relevanz mit größerer Dringlichkeit. Dem Eindruck eines scheinbaren „Überflusses“ digitaler Quellen und Informationen widerspricht der Befund eines krassen Machtgefälles in der Sphäre des Digitalen, des so genannten „digital divide“ (Ragnedda und Muschert 2013; Putnam 2016; Rogerson 2020). Soziale Gruppen, die nicht über die Fähigkeiten oder die Möglichkeiten verfügen, ihre Perspektiven in die dominierenden und durch Diskursverstärker, wie Medien, wahrgenommenen Diskurssphären einzuspeisen, sind von Ausgrenzung und Nichtbeachtung bedroht. Diese Ausgrenzung ist, wie im Fall der Lese-, Schreib- und Informationskompetenz allgemein, intersektional, das heißt, das Problem betrifft Menschen entlang miteinander verschränkter Zugehörigkeiten wie Alter, Geschlecht, sozialer Status, Kultur, Herkunft etc. Sowohl der Zugang zu digitalen Informationen über das Internet und die Fähigkeit zu ihrer Einordnung, als auch die Möglichkeit, Quellen und Narrative, Debatten oder auch Diskurse digital bewahren und zugänglich machen zu können, wird in hohem Maße durch den ökonomischen Status und die Existenz entsprechender (oft staatlich organisierter) Infrastrukturen bedingt. Historische Fakten, Fachdebatten oder gesellschaftliche Kontroversen, die nicht-westlichen Gesellschaften und Sprachräumen entstammen, sind schon auf der Ebene von Wikipedia weniger präsent und in weitaus höherem Maße im Word Wide Web insgesamt von einer Verdrängung durch von westlichen Perspektiven und Sprachen dominierten Inhalten bedroht. Der Digital Divide betrifft auch die Fähigkeit von Gesellschaften, Gruppen, Institutionen oder Individuen, ihre Narrative und Quellen ihrer Geschichte dauerhaft und global zugänglich zu machen. Es ist evident, dass sich die wirtschaftliche und politische Übermacht der westlichen Industriestaaten gegenüber den Gesellschaften des globalen Südens auch auf dieser Ebene manifestiert (Putnam 2016). Doch selbst die digitale Repräsentation hochindustrialisierter Nationen kann der digitalen Amnesie anheimfallen, wenn sich die staatlichen Strukturen auflösen, wie Anat Ben-David anhand der Top-Level-Domain der ehemaligen Republik Jugoslawien, .yu, gezeigt hat (Ben-David 2016).

Auch in Europa sind bisher lediglich rund vier Prozent der Sammlungen von Gedächtnisinstitutionen digitalisiert worden, worüber die großen Sammlungen von online verfügbaren Digitalisaten hinwegtäuschen können (vgl. König 2020). Hinzu kommt, dass es sich hierbei vor allem um textbasierte Quellen aus dem globalen Norden mit einem Fokus auf die Nationalgeschichte handelt, was Mareike König zufolge einen nachgewiesenen Einfluss auf die Themensetzung in der Geschichtsforschung hat (König 2020).

Vergessen und Suchen

Die Verfügbarkeit bestimmt auch die Möglichkeit und den nötigen Aufwand, der aufgebracht werden muss, um potentiell relevante Quellen aufzuspüren. Die oben dargestellte Volatilität digitaler Quellen durch ihre einfache Löscharkeit, die mögliche technische und materielle Obsoleszenz sowie der entsprechend hohe Aufwand ihrer Bewahrung (durch persistente Identifikation, umfassende Dokumentation, redundante Speicherung und ggf. konstante Migration) führen zu der bereits angesprochenen hohen Verlustrate digitaler Zeitzeugnisse. Was sich für die historische Forschung als bedauerlich darstellen mag, wurde bzw. wird von Zeitgenossen unter Umständen anders bewertet: Das Vergessenwerden, die Auslöschung digitaler Spuren eines früheren, abgelegten Selbst, kann von den Menschen des digitalen Informationszeitalters angesichts des Zwangs zur permanenten Selbstdarstellung im Arbeitsmarkt und der Bedrohung durch staatliche und private Überwachung als segensreich empfunden werden. Die historischen Wissenschaften haben seit ihrem Bestehen mit

lückenhaften Überlieferungen der Vergangenheit gearbeitet und diese Lücken zu einem festen Bestandteil ihrer heuristischen Methode gemacht. Auch eine digitale Quellenkritik muss daher mit reflektieren, was nicht überliefert sein könnte oder ihr zumindest nicht (mehr) zugänglich ist, und entsprechend die Aussagekraft, Generalisierbarkeit und Repräsentativität ihrer Befunde und Argumente bemessen.

Aus der Volatilität digitaler Objekte, dem gleichzeitigen exponentiellen Anwachsen potentiell relevanter digitaler Zeitzeugnisse und den geschilderten (Macht-)Bedingungen ihrer Zugänglichkeit und Auffindbarkeit ergeben sich auch bestimmte Anforderungen an die Vorgehensweise bei der Suche und Sammlung digitaler Quellen. Verkürzt lässt sich dies anhand der beiden gängigen Methoden der Informationsakquise unter digitalen Bedingungen darstellen: Suchen und Browsen.

Bei der Suche handelt es sich im eigentlichen Sinne um eine Datenbankabfrage: ein oder mehrere Suchbegriff/e werden einer Datenbank übergeben und optional mit Filtern eingegrenzt oder mit Booleschen Operatoren verknüpft oder trunkiert (Duguid 2009; Mayer 2009). Übereinstimmungen mit Inhalten der Datenbank werden dann als Ergebnisliste angezeigt. Dies setzt voraus, dass die Suchbegriffe in den Titeln der Datensätze enthalten sind (bzw. in den Datensätzen, falls eine Volltextsuche möglich ist) oder vorgängig als Schlagworte vergeben wurden. Es besteht also theoretisch stets die Möglichkeit, dass ein thematisch relevanter Eintrag durch die Suche nicht gefunden werden kann, weil der Suchbegriff nicht darin enthalten oder nicht als Schlagwort vergeben worden ist. Der Erfolg oder Misserfolg der Datenbankabfrage beruht also ebenso sehr auf der Wahl der Suchstrategie und -begriffe, wie auf dem Grad und der Qualität der Annotation und Aufbereitung der Datensätze.

Ein Spezialfall der Datenbankabfrage ist die Websuche mithilfe von Suchmaschinen. Technisch gesehen entsprechen sie der Befragung eines Graphen oder einer Datenbank, da Suchmaschinenanbieter Webinhalte laufend indizieren und zur Verfügung stellen. Bei der Websuche ergeben sich zwei wichtige Probleme: Erstens wird jeglicher Suchbegriff in einer weit verbreiteten Sprache potentiell eher zu viele als zu wenige und noch dazu größtenteils irrelevante Ergebnisse hervorbringen. Zweitens setzen Suchmaschinenanbieter vorgängig Filter ein und sortieren ihre Ergebnisse mithilfe spezifischer Algorithmen; beide Prozesse beeinflussen die Ergebnisliste und werden den Suchenden nicht offengelegt. Das Gegenmodell zur Begriffssuche ist das Browsen, also der Nachvollzug einer systematisch und meist hierarchisch angelegten Ablageordnung. Theoretisch besteht der Vorteil bei dieser Vorgehensweise darin, dass sich so auch Informationen auffinden lassen, die thematisch zum gesuchten Feld passen, jedoch gewisse Suchbegriffe nicht enthalten. Diese Form der Suche ist allerdings an die Ablageordnung und die daraus entstehenden Pfade gebunden, das heißt, der Erfolg dieser Strategie basiert auf der Qualität, Kohärenz und Vollständigkeit der vorgängigen Erschließung und Sortierung der Informationen. Sowohl Begriffs-Suche als auch Browsen beruhen auf oftmals intransparenten Vorentscheidungen von Institutionen, die Suchmasken oder Ablageordnungen zur Verfügung stellen. Beide Suchstrategien laufen damit Gefahr, mögliche Vorannahmen, Vorurteile oder Kenntnislücken solcher Institutionen zu replizieren. Teilweise lassen sich solche Formen von kulturellem oder institutionellem Bias erkennen. Eine Google-Websuche etwa bringt deutlich andere Ergebnisse (und eine andere Gewichtung) hervor, wenn sie (etwa mittels VPN-Tunnel) von einem anderen geografischen Standort aus getätigt wird. In gleicher Weise empfiehlt es sich stets, Wikipedia-Einträge auch noch in anderen Sprachen zur Kenntnis zu nehmen und auf etwaige Unterschiede in Inhalt, Quellenangaben und Schwerpunktsetzung zu prüfen.

//

Literatur

Arnold, Klaus. 2002. Quellenkritik. In: *Lexikon Geschichtswissenschaft. Hundert Grundbegriffe*, hg. von Stefan Jordan, 255–257. Nachdruck v. 2013. Stuttgart: Reclam.

Baberowski, Jörg. 2005. *Der Sinn der Geschichte: Geschichtstheorien von Hegel bis Foucault*. München: C.H. Beck.

Baucom, Erin. 2019. A Brief History of Digital Preservation. In: *Digital Preservation in Libraries: Preparing for a Sustainable Future*, hg. von Jeremy Myntti und Jessalyn Zoom, 3–19. Chicago: American Library Association. Online: https://scholarworks.umat.edu/ml_pubs/31 (zugegriffen: 17. April 2024).

Ben-David, Anat. 2016. What does the Web remember of its deleted past? An archival reconstruction of the former Yugoslav top-level domain. *New Media & Society* 18, Nr. 7: 1103–1119. <https://doi.org/10.1177/1461444816643790>.

Berners-Lee, Tim. 2006. Linked Data – Design Issues. W3C. <https://www.w3.org/DesignIssues/LinkedData.html> (zugegriffen: 17. April 2024).

Berners-Lee, Tim, James Hendler und Ora Lassila. 2001. The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities. *Scientific American*. https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf (zugegriffen: 17. April 2024).

Blanchette, Jean-François. 2011. A material history of bits. *Journal of the American Society for Information Science and Technology* 62, Nr. 6: 1042–1057. <https://doi.org/10.1002/asi.21542>.

Brügger, Niels. 2012. When the Present Web is Later the Past: Web Historiography, Digital History, and Internet Studies. *Historical Social Research / Historische Sozialforschung* 37, Nr. 4: 102–117.

Droysen, Johann Gustav. 1868. *Grundriss der Historik*. Leipzig: Veit. https://www.deutschestextarchiv.de/droysen_historik_1868 (zugegriffen: 17. April 2024).

Drucker, Johanna. 2013. Performative Materiality and Theoretical Approaches to Interface. *Digital Humanities Quarterly* 7, Nr. 1. <https://digitalhumanities.org/dhq/vol/7/1/000143/000143.html> (zugegriffen: 17. April 2024).

Duguid, Paul. 2009. Die Suche vor grep: Eine Entwicklung von Geschlossenheit zu Offenheit? In: *Deep Search: Politik des Suchens jenseits von Google*, hg. von Konrad Becker und Felix Stalder, 15–36. Innsbruck, Wien, Bozen: StudienVerlag.

Emich, Birgit. 2019. *Geschichte der Frühen Neuzeit (1500–1800) studieren*. 2. vollst. überarb. Aufl. München: UVK Verlag. <https://www.utb.de/doi/book/10.36198/9783825247683> (zugegriffen: 04. Juli 2024).

Fickers, Andreas. 2021. Authenticity: Historical Data Integrity and the Layered Materiality of Digital Objects. In: *Digital Roots: Historicizing Media and Communication Concepts of the Digital Age*, hg. von Gabriele Balbi, Nelson Ribeiro, Valérie Schafer und Christian Schwarzenegger, 299–312. Berlin, Boston: De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110740202-017>.

Föhr, Pascal. 2019. *Historische Quellenkritik im Digitalen Zeitalter*. Glückstadt: VWH Verlag Werner Hülsbusch.

Gogolin, Greg, Hrsg. 2021. *Digital Forensics Explained*. 2. Aufl. Boca Raton: CRC Press. <https://www.routledge.com/Digital-Forensics-Explained/Gogolin/p/book/9780367503437> (zugegriffen: 28. Juni 2023).

Haber, Peter. 2011. *Digital past: Geschichtswissenschaft im digitalen Zeitalter*. München: Oldenbourg Verlag. <https://doi.org/10.1515/9783486712339>.

Hiltmann, Torsten. 2022. Vom Medienwandel zum Methodenwandel: Die fortschreitende Digitalisierung und ihre Konsequenzen für die Geschichtswissenschaften in historischer Perspektive. In: *Digital History: Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft*, hg. von Karoline Dominika Döring, Stefan Haas, Mareike König und Jörg Wettlaufer, 13–44. Berlin, Boston: De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110757101-002>.

Jordan, Stefan. 2018. *Theorien und Methoden der Geschichtswissenschaft*. 4. Aufl. Paderborn: Schöningh.

Kirschenbaum, Matthew G. 2012. *Mechanisms: New Media and the Forensic Imagination*. Cambridge, Massachusetts: MIT Press. <https://mitpress.mit.edu/9780262517409/mechanisms/> (zugegriffen: 17. April 2024).

König, Mareike. 2020. Geschichte digital: Zehn Herausforderungen. In: *Geschichtswissenschaft im 21. Jahrhundert: Interventionen zu aktuellen Debatten*, hg. von Cord Arendes, Karoline Döring, Claudia Kemper, Mareike König, Thorsten Logge, Angela Siebold und Nina Verheyen, 67–76. Berlin, Boston: De Gruyter Oldenbourg. Online: <https://blog.degruyter.com/geschichte-digital-zehn-herausforderungen/> (zugegriffen: 17. April 2024).

Krämer, Sybille. 2018. Der ‚Stachel des Digitalen‘ – ein Anreiz zur Selbstreflexion in den Geisteswissenschaften? Ein philosophischer Kommentar zu den Digital Humanities in neun Thesen. *Digital Classics Online* 4, Nr. 1: 5–11. <https://doi.org/10.11588/dco.2017.0.48490>.

Lialina, Olia. 2019. GeoCities’ Afterlife and Web History. In: *One Terabyte of Kilobyte Age*. <https://blog.geocities.institute/archives/6418> (zugegriffen: 17. April 2024).

Mayer, Katja. 2009. Zur Soziometrik der Suchmaschinen. Ein historischer Überblick der Methodik. In: *Deep Search: Politik des Suchens jenseits von Google*, hg. von Konrad Becker und Felix Stalder, 64–83. Innsbruck, Wien, Bozen: StudienVerlag.

Milligan, Ian. 2019. *History in the Age of Abundance? How the Web Is Transforming Historical Research*. Montreal u.a.: McGill-Queen’s University Press.

Putnam, Lara. 2016. The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast. *The American Historical Review* 121, Nr. 2: 377–402. <https://doi.org/10.1093/ahr/121.2.377>.

Ragnedda, Massimo und Glenn W. Muschert, Hrsg. 2013. *The Digital Divide: The Internet and Social Inequality in International Perspective*. Routledge Advances in Sociology. London: Routledge. <https://www.routledge.com/The-Digital-Divide-The-Internet-and-Social-Inequality-in-International/Ragnedda-Muschert/p/book/9781138960268> (zugegriffen: 17. April 2024).

Ries, Thorsten. 2022. Digital history and born-digital archives: the importance of forensic methods. *Journal of the British Academy* 10: 157–185. <https://doi.org/10.5871/jba/010.157>.

Ries, Thorsten und Gábor Palkó. 2019. Born-digital archives. *International Journal of Digital Humanities* 1, Nr. 1: 1–11. <https://doi.org/10.1007/s42803-019-00011-x>.

Rogerson, Simon. 2020. The digital divide is a multi-dimensional complex. *Journal of Information, Communication and Ethics in Society* 18, Nr. 3: 321–321. <https://doi.org/10.1108/JICES-05-2020-0060>.

Rosenzweig, Roy. 2003. Scarcity or Abundance? Preserving the Past in a Digital Era. *The American Historical Review* 108, Nr. 3: 735–762. <https://doi.org/10.1086/ahr/108.3.735>.

Rothenberg, Jeff. 1999. Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. CLIR Reports 77. Washington, D.C.: Council on Library and Information Resources. <https://www.clir.org/wp-content/uploads/sites/6/pub77.pdf> (zugegriffen: 17. April 2024).

—. 2000. Preserving Authentic Digital Information. In: *Authenticity in a Digital Environment*, hg. von Council on Library and Information Resources, 51–68. CLIR Reports 92. Washington, D.C.: Council on Library and Information Resources. <https://www.clir.org/wp-content/uploads/sites/6/pub92.pdf> (zugegriffen: 17. April 2024).

Sammons, John. 2014. *The Basics of Digital Forensics: The Primer for Getting Started in Digital Forensics*. 2. Aufl. Waltham, MA: Syngress.

Schöch, Christof. 2013. Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities* 2, Nr. 3: 2–13. Online: <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/> (zugegriffen: 17. April 2024).

Sulé, Andreu und Laia Lapeyra. 2016. Introduction to the Semantic Web and Linked Data. *DLIS. Digital Libraries and Information Science*. 21. Oktober. <https://dlis.hypotheses.org/788> (zugegriffen: 17. April 2024).

Thibodeau, Kenneth. 2002. Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In: *The State of Digital Preservation: An International Perspective*, hg. von Council on Library and Information Resources, 4–31. CLIR Reports 107. Washington, D.C.: Council on Library and Information Resources. Online: <https://www.clir.org/pubs/reports/pub107/thibodeau/> (zugegriffen: 17. April 2024).

W3C. 2014. Linked Data Platform Use Cases and Requirements. W3C Working Group Note 13 March 2014. W3C. <https://www.w3.org/TR/ldp-ucr/> (zugegriffen: 17. April 2024).

Willer, Christoph. 2012. *PC-Forensik: Daten suchen und wiederherstellen*. Böblingen: Computer und Literatur Verlag.

Zimmermann, Bernhard und Antonios Rengakos, Hrsg. 2014. *Handbuch der griechischen Literatur der Antike. Bd. 2: Die Literatur der klassischen und hellenistischen Zeit*. München: C.H. Beck. <https://www.chbeck.de/zimmermann-rengakos-handbuch-griechischen-literatur-antike-bd-2-literatur-klassischen-hellenistischen-zeit/product/8005213> (zugegriffen: 17. April 2024).

Endnoten

- ¹ Auch staatlich finanzierte Archive müssen für die Anforderungen der Langzeitarchivierung hohe Personal- und Energiekosten budgetieren.
- ² Siehe z. B. via Lialina 2019.
- ³ Nur wenige kommerzielle Softwareprogramme zeigen die Namen der beteiligten Programmierer:innen an. Bei Adobe Photoshop etwa erscheinen die Namen kurz und in sehr kleiner Schrift beim Programmstart, allerdings ohne Angabe der spezifischen Rolle.

Zitierweise

Feichtinger, Moritz; Althage, Melanie; Siebold, Anna (2024): Theoretische Grundlagen einer digitalen Quellenkritik. In: Living Handbook "Digitale Quellenkritik". Version 1.1. hrsg. v. Deicke, Aline; Geiger, Jonathan D.; Lemaire, Marina; Schmunk, Stefan. <https://doi.org/10.5281/zenodo.12647879>.

Metadaten

Autor:innen	Althage, Melanie ; Feichtinger, Moritz ; Siebold, Anna ;
Sprache	Deutsch
DOI	10.5281/zenodo.12647879
Creative Commons Lizenztyp	Attribution CC BY (4.0)
Verwandte Artikel	1. Einleitung: Was ist Digitale Quellenkritik? ;