

Die Geisteswissenschaften stehen erst am Anfang einer fundamentalen Veränderung durch digitale Technologien. Es zeigt sich aber bereits heute, dass diese Veränderung, der ›Digital Turn‹, für die Geisteswissenschaften neue, zusätzliche Möglichkeiten bieten wird, deren Potenzial noch längst nicht ausgelotet ist. Diese Digital Humanities verlangen ergänzend neue Methoden und Fachwissen.

The humanities are only at the beginning of a fundamental change through digital technologies. However, it is already clear today that this ‘digital turn’ will offer new, additional opportunities for the humanities, the potential of which is far from being fully explored. These Digital Humanities require new methods and expertise.

Die Digital Humanities sind keine junge Disziplin, denn die ersten Überlegungen zur Anwendung der Elektronischen Datenverarbeitung (EDV) in den Geistes-, Sozial- und Kulturwissenschaften gehen im deutschsprachigen Raum zurück in die 1970er-Jahre – die Frühzeit digitaler Computertechnologien. Es ist dennoch auch heute noch schwierig, den Begriff der Digital Humanities klar abzugrenzen. Es handelt sich um ein Fach und zugleich eine Form einer Methodik, die durch den zunehmenden gesellschaftlichen Stellenwert des Digitalen besonders in der geisteswissenschaftlichen Forschung unabdingbar wird. Nur durch den gezielten und reflektierten Einsatz der Computertechnologie kann ein Mehrwert für die geisteswissenschaftliche Forschung erzielt werden. Eine zentrale Rolle spielen in den Geisteswissenschaften die Informationsqualität und die Informationsverfügbarkeit. Dies zu realisieren und sicherzustellen stellt eine der wesentlichen Herausforderungen in der digitalen Domäne dar. Beides wird zudem immer stärker gefordert und auch gefördert, um so nicht zuletzt die Nachhaltigkeit von Forschungsinvestitionen sicherzustellen. Diese Bestrebungen haben zur Folge, dass Begriffe wie Open Access oder Open Data in den Geisteswissenschaften, zum Teil mehr als in anderen Disziplinen, zunehmend an Bedeutung gewinnen. Die Anforderungen an die Infrastruktur und die betreibende Organisation, um Forschungsdaten langfristig verfügbar zu halten, sind enorm. Alleine schon das Sicherstellen einer nachhaltigen Zitierfähigkeit von digitalen Dokumenten oder Daten ist keine einfache Aufgabe.

Digitale Quellen

Digitale Quellen stellen das wesentliche Substrat für die Forschung in den Digital Humanities dar. Sie werden oft durch den Prozess des Digitalisierens oder Abtastens, das heißt zum Beispiel durch Messen von Farbe oder Form eines Originals, erzeugt und stellen somit eine mehr oder weniger vollständige digitale Reproduktion eines Originals dar. Diese Reproduktion sind beispielsweise digitale Rasterdaten eines fotografisch erzeugten Bilds, Samples einer Tonaufzeichnung oder gar Vektordaten in einem dreidimensionalen Raum. Ebenso stellt ein digitaler Text, der beispielsweise durch OCR (Optical Character Recognition, also Texterkennung) aus einem digitalen Bild extrahiert wurde, eine digitale Quelle dar. Diese Reproduktionen können und sollen das Original nicht ersetzen, aber sie repräsentieren im Idealfall wichtige Eigenschaften des Originals, welche sich in der digitalen Domäne im Kontext der Forschung einfach transportieren, kopieren und auswerten lassen. Werden diese Objektdaten mit technischen oder kontextuellen Metadaten ergänzt, so entsteht ein vollständiges digitales Informationsset, welches über geeignete Schnittstellen – auch Interfaces genannt – zur Verfügung gestellt werden kann. Im Normalfall werden solch digitale Quellen in einer strukturierten Form über eine Datenbank für die Nutzung bereitgestellt. Werden die Daten im Internet zur Verfügung gestellt, wird von Open Data gesprochen. Solche Schnittstellen stellen Kommunikationsanschlüsse dar, welche in der korrekten Sprache – einem Protokoll – angesprochen werden müssen und welche die Daten in einem von vielen möglichen Formaten als Antwort auf

die Anfrage zur Verfügung stellen. Diese Kommunikationsanschlüsse, welche in ganz unterschiedlicher Form existieren, werden in der modernen Computertechnologie API (Application Programming Interface) genannt.

Die Standardisierung solcher Schnittstellen ist von großer Wichtigkeit, um die Daten suchen, finden und abrufen zu können. Ein solcher Standard, der heute oft genannt und verwendet wird, ist Linked Open Data (LOD). LOD bezeichnet im Internet unter dem Service des World Wide Web (Web oder www) frei verfügbare Daten, die mit einem Uniform Resource Identifier (URI) per HTTP (dem Hypertext Transfer Protocol, dem Übertragungsprotokoll des Web) abgerufen und in einer semantischen ›Sprache‹, einem spezifischen Datenformat, zur Verfügung gestellt werden. Diese Sprache wird Resource Description Framework (RDF)¹ genannt. So zur Verfügung gestellte Information kann nun von Forschenden direkt oder aber durch andere Computersysteme verknüpft und ausgewertet werden. Die URI stellt hierbei sicher, dass die Daten über einen unbeschränkten Zeitraum unter derselben Adresse gefunden und genutzt werden können. Nur wenn diese Bedingung erfüllt und zudem sichergestellt ist, dass der Inhalt der so bereitgestellten Quelle sich nicht ändert, kann von einer Zitierfähigkeit der Daten gesprochen werden.

Digitale Werkzeuge

Werden Daten im Internet oder dem Web zur Verfügung gestellt, so ist eine wichtige Frage, wie sie genutzt werden können. Das FAIR²-Prinzip besagt zwar, dass Daten über eine Schnittstelle in standardisierter Form gesucht und zur Nutzung geladen werden können; die Form der Auswertung wird allerdings nicht explizit definiert. Dennoch kommt heute dem Begriff der Interoperabilität ein hoher Stellenwert zu. Unter Interoperabilität wird das Zusammenführen von Datenbeständen, die in verschiedenen Datenbanken gespeichert sind, verstanden. Dies ist besonders nützlich, wenn geeignete Werkzeuge zur Darstellung oder Auswertung der zur Verfügung gestellten Daten eingesetzt werden können. Dies erlaubt zum Beispiel das virtuelle Bilden und Auswerten von Sammlungen von Objekten, die sich physisch nicht am selben Ort befinden. Der Wunsch nach dem funktionalen Austausch von Daten, der Interoperabilität, führte zu einer Initiative, die durch die British Library, die Oxford Bodleian Library, die Harvard University, die Stanford University Library, die Cornell University sowie die Nationalbibliotheken von Frankreich und Norwegen, mit der Unterstützung der Andrew Mellon Foundation, ins Leben gerufen wurde. Die Initiative trägt den Namen IIRF (triple-i-f)³ oder International Image Interoperability Framework. IIRF ist ein Paradebeispiel für den Mehrwert, der mit dem Zusammenführen unterschiedlicher Datenquellen einhergeht.

Ein schönes Beispiel ist die virtuelle Bibliothek mittelalterlicher Handschriften e-codices⁴, welche mit einer geeigneten Web-Applikation wie dem Mirador-Viewer⁵ örtlich völlig ungebunden betrachtet werden kann. Dies erlaubt, dass die gespeicherten Handschriften, die geografisch an ganz unterschiedlichen Orten gelagert sind, zeitlich und örtlich vereinheitlicht betrachtet und genutzt werden können. Die eigentlichen Arbeitswerkzeuge können ganz unterschiedlich ausfallen: Es kann sich dabei beispielsweise um Applikationen zur statistischen Auswertung von großen Textkorpora oder um Annotations-Hilfen zur Beschreibung



von spezifischen Bildteilen in der Archäologie handeln; den möglichen Anwendungen sind kaum Grenzen gesetzt. Die Zurückführung von Erkenntnissen aus der vergleichenden Zusammenführung von Quellmaterial ist hierbei sehr wichtig, da nur so der inhaltliche Mehrwert anderen Forschenden zur Verfügung gestellt werden kann und somit echte ›Reusability‹ von Daten erreicht wird.

Stand heute

Viele mit öffentlichen Geldern finanzierte Forschungsprojekte in den Geisteswissenschaften resultieren heute in einem kleineren oder größeren digitalen Datenbestand. Das FAIR-Prinzip ist für diese Forschungsdatenbestände besonders wichtig, da die Wiederverwertbarkeit – die besagte Reusability – von den Förderorganisationen wie dem Schweizerischen Nationalfonds mit sogenannten Data-Management-Plänen (DMPs) eingefordert wird.⁶ Für die Forschenden und die Forschungsinstitutionen stellt hierbei die notwendige Infrastruktur zur langfristigen Sicherung und Verfügbarhaltung der Datenbestände eine Herausforderung dar. Die Kosten für die damit verbundenen Computer-, Software- und Speichertechnologien sind beträchtlich und die Gefahr von technisch weitgehend redundanten Insellösungen ist groß, was die Kosten langfristig weiter in die Höhe schnellen lässt. Redundanz ist zwar für die Sicherung der Daten von großer Wichtigkeit, Daten müssen redundant vorhanden sein, um so das Verlustrisiko durch den möglichen Ausfall einer Kopie der Daten zu minimieren; redundante, unkoordinierte Infrastrukturen sind hingegen nicht förderlich, da diese parallel betriebenen Plattformen gepflegt und technologisch auf dem neusten Stand gehalten werden müssen, was die Kosten massiv steigen lässt.

Eine Lösung für das Bereitstellen von Forschungsdaten sind vertrauenswürdige nationale, langfristig finanzierte digitale Infrastrukturen, welche die Aufgabe der Verfügbarhaltung der digitalen Quellen übernehmen. Diese digitale Forschungs-Grundversorgung ist besonders in den Geisteswissenschaften erst spät geschaffen worden, da die Computersysteme den Anforderungen für die Verfügbarhaltung von qualitativen Forschungsdaten vor einigen Jahren noch nicht leistungsfähig und flexibel genug waren. Heute ist die Technologie allerdings so weit fortgeschritten, dass geeignete Systeme betrieben werden können. Eine solche nationale Forschungsinfrastruktur für den nachhaltigen Erhalt von geisteswissenschaftlichen Forschungsdaten ist das DaSCH, das Data and Service Center for the Humanities.⁷ Das DaSCH, welches von der Schweizerischen Akademie der Geisteswissenschaften finanziert wird, hat zum Ziel, den Forschenden genau diese Grundversorgung – also die langfristige Speicherung, den Zugriff sowie Werkzeuge zur Bearbeitung der Daten – bereitzustellen.

Eine zweite, deutlich zu erkennende aktuelle Tendenz ist der funktionale und standardisierte Datenaustausch zwischen unterschiedlichen Systemen, die Interoperabilität. Dies stellt zugleich eine Herausforderung und ein großes Innovationspotenzial mit Transfermöglichkeiten in Bereiche außerhalb der Humanities dar, da die in den Geisteswissenschaften vorliegenden Datenbestände stark heterogen sind und aus ganz unterschiedlichen Domänen stammen können. Sollen solch heterogene Datensätze miteinander verknüpft oder in Verbindung gebracht werden,



so bedarf dies besonderer Anstrengungen. Digitale Metainformation wird mit sogenannten Metadaten-Schemen dargestellt. Sollen diese Meta-informationen aus mehreren Repositorien zusammengeführt werden, ist eine einheitliche Sprache notwendig. Linked Open Data versucht, durch »selbsterklärende« Datenstrukturen diese Interoperabilität über das World Wide Web zu ermöglichen. Zudem werden mit LOD Relationen zwischen einzelnen Datenelementen darstellbar, was erlaubt, Daten mit einer semantischen Bedeutung zu versehen. Um dies in der Praxis realisieren zu können, sind allerdings bei allen beteiligten Repositorien entsprechende Schnittstellen notwendig, was heute noch nicht immer gegeben ist. Ist allerdings ein domänenübergreifender Datenaustausch gewährleistet, so entsteht eine große Ansammlung von hochinteressanter digitaler Information, »Big Data in den Humanities«, welche mit geeigneten Algorithmen verarbeitet werden kann – was zu ganz neuen geisteswissenschaftlich relevanten Resultaten führt.

Eine Schwierigkeit angesichts von Open Data im World Wide Web stellt allerdings die rechtliche Situation dar. Sollen Daten öffentlich zugänglich gemacht werden, muss gewährleistet sein, dass dies auch gestattet ist. Hierbei sind Aspekte des Urheberrechts genauso zu berücksichtigen wie die Regeln des Persönlichkeitsschutzes, wenn es um Personendaten geht. Um diese Rahmenbedingungen sicherzustellen, sind unterschiedliche Informationsquellen vorhanden, wie zum Beispiel die entsprechende Seite des Schweizerischen Nationalfonds.⁸ Zudem bieten die meisten Universitäten entsprechende Leitfäden und teilweise auch persönliche Beratung an.

Ausblick

Die Geisteswissenschaften stehen erst am Anfang einer fundamentalen Veränderung durch digitale Technologien. Es zeigt sich aber bereits heute, dass diese Veränderung, der »Digital Turn«, für die Geisteswissenschaften neue, zusätzliche Möglichkeiten bieten wird, deren Potenzial noch längst nicht ausgelotet ist. Genauso wie Bibliotheken, Archive und Museen schon früher die Bezugsorte für forschungsrelevante Information in physischer Form waren, so stellen im Netz verfügbare Daten eine wichtige und in Zukunft essenzielle Quelle für die geisteswissenschaftliche Forschung dar.

Aus diesem Grund sind Anstrengungen nötig, um diese in Zukunft vermehrt in digitaler Form vorliegenden Quellen in Bezug auf ihre Qualität, Aussagekraft und Verfügbarkeit in ein zweckmäßiges und nutzbare Format zu bringen und diese in geeigneten Infrastrukturen zu speichern und sie langfristig bereitzustellen.

Nationale Initiativen und Lösungen stellen hier wichtige Säulen dar, die durch zunehmende interdisziplinäre Zusammenarbeit weiter gestärkt werden können. Durch den internationalen Austausch und die Nutzung internationaler Standards kann zudem die Interoperabilität und somit der internationale Austausch über die Grenzen der Disziplinen der geisteswissenschaftlichen Forschung hinaus weiter gesteigert werden.

Dieser digitale Wandel wird in Zukunft noch stärker zeigen, wie wichtig die geisteswissenschaftliche Forschung ist und sein wird: Nebst der Auswertung und Verknüpfung historischer Information mit aktuellen Fragestellungen stellt die digitale Unterstützung bewährter geisteswissen-

schaftlicher Forschungsmethoden einen Katalysator dar, welcher neue Antworten auf gesellschaftlich hoch relevante Fragen liefern kann. Jede Software, jede Applikation benötigt jedoch Daten, um diese zu verarbeiten und Resultate liefern zu können. Aus diesem Grund ist offene Information, Open Data, essenziell für die digitalen Geisteswissenschaften.

Somit müssen Studierende in den Humanities den Umgang mit digitalen Inhalten von Grund auf lernen, wozu nebst der Fähigkeit zur Nutzung von Daten und Werkzeugen auch die kritische Bewertung vorhandener Quellen und digitaler Arbeitswerkzeuge gehört. Aus diesem Grund wird an der Universität Basel seit dem Herbstsemester 2019 ein Masterstudienfach Digital Humanities angeboten, welches den Studierenden diese Grundlagen vermittelt.⁹



- 1 www.w3.org/RDF/ (alle Links in diesem Beitrag zuletzt aufgerufen am 17. Januar 2020).
- 2 Findable, Accessible, Interoperable, Reusable; Definition gemäß Mark D. Wilkinson / Michel Dumontier / IJsbrand Jan Aalbersberg / Gabrielle Appleton / Myles Axton / Arie Baak / Niklas Blomberg / Jan-Willem Boiten / Luiz Bonino da Silva Santos / Philip E. Bourne / Jildau Bouwman / Anthony J. Brookes / Tim Clark, Mercè Crosas / Ingrid Dillo / Olivier Dumon, Scott Edmunds / Chris T. Evelo / Richard Finkers / Alejandra Gonzalez-Beltran / Alasdair J. G. Gray / Paul Groth / Carole Goble / Jeffrey S. Grethe / Jaap Heringa / Peter A. C't Hoen / Rob Hooft / Tobias Kuhn / Ruben Kok / Joost Kok / Scott J. Lusher / Maryann E. Martone / Albert Mons / Abel L. Packer / Bengt Persson / Philippe Rocca-Serra / Marco Roos / Rene van Schaik / Susanna-Assunta Sansone / Erik Schultes / Thierry Sengstag / Ted Slater / George Strawn / Morris A. Swertz / Mark Thompson / Johan van der Lei / Erik van Mulligen / Jan Velterop / Andra Waagmeester / Peter Wittenburg / Katherine Wolstencroft / Jun Zhao & Barend Mons: The FAIR Guiding Principles for scientific data management and stewardship, in: *Scientific Data* 3 (2016), Nr. 160018, doi: 10.1038/sdata.2016.18.
- 3 <https://iiif.io>.
- 4 www.e-codices.unifr.ch/de.
- 5 www.e-codices.unifr.ch/en/mirador.
- 6 www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/data-management-plan-dmp-leitlinien-fuer-forschende.aspx.
- 7 <http://dasch.swiss>.
- 8 www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/default.aspx#Grundsatzerkl%E4rung%20des%20SNF%20zu%20Open%20Research%20Data.
- 9 <https://philhist.unibas.ch/de/studium/studienangebote/master/digital-humanities-master>.