

# 1. Erfassen

Erfassen bedeutet im weiteren Sinne das Erzeugen digitaler Gegenstücke zu existierenden Kulturgegenständen oder die Darstellung existierender Artefakte als digitales Objekt ("Digitalisierung"). Erfassen kann eine manuelle Tätigkeit (siehe "Transkribieren") oder ein automatischer Prozess (siehe "Bilderzeugung" oder "Datenerkennung") sein. Systematisch betrachtet – aber auch in der praktischen Umsetzung – steht das Erfassen vor Tätigkeiten wie Anreicherung und Analyse.

## 1.1. Konvertieren

Konvertieren bedeutet, das Format eines Objektes in ein anderes (z.B. analog zu digital oder die Umwandlung von .wmv zu .mov Datei) umzuwandeln, ohne dass der Inhalt des Objektes verändert wird. Sofern es sich bei den zu konvertierenden Objekten lediglich um Metadaten handelt, umfasst dieser Schritt die Abbildung eines Metadaten-Schemas auf ein anderes (auch Mapping oder Crosswalk genannt). Grundlegenderes 'Konvertieren' wie das Überführen eines Scans in ein editierbares Textdokument werden passender mittels "Datenerkennung" referenziert.

## 1.2. Datenerkennung

Die Datenerkennung (z.B. OCR) bezieht sich auf Verfahren der digitalen Datenerfassung, die etwa aus digitalen Faksimiles von Texten oder Noten diskrete, maschinenlesbare Einheiten extrahiert (z.B. Klartext, Musiknoten) oder in bewegten Bildelementen beispielsweise Gesichter erkennt.

### **1.3. Entdecken**

Entdecken beschreibt das Identifizieren von Forschungsobjekten, Forschungsergebnissen oder weiterer Erkenntnisse während eines Forschungsprozesses. Das Entdecken schließt zielorientierte Techniken wie komplexe Datenbankabfragen ebenso mit ein wie einfache Suchanfragen und Zufallsfunde, wie sie z.B. bei (Web-) Recherchen vorkommen. Es unterscheidet sich von Information Retrieval, bei dem es sich um einen strukturierten Ansatz handelt eine bestimmte Information oder ein Subset von Objekten aus einer Ressource zu extrahieren.

### **1.4. Sammeln**

Sammeln beschreibt das Aggregieren entdeckter Ressourcen. Dies geschieht für gewöhnlich auf strukturierte Art, z.B. wenn alle Veröffentlichungen zu einem spezifischen Thema im Rahmen eines literarischen Reviews zusammengeführt werden oder wenn die Werke eines bestimmten Autors zur weiteren Analyse aus einer digitalen Sammlung herausgezogen werden. Verwandte Techniken sind web crawling und scraping.

### **1.5. Bilderfassung**

Bei der Bilderfassung werden Texte, Bilder, Artefakte oder räumliche Gliederungen durch optische oder akustische Mittel (z.B. Licht, Laser, Infrarot, Ultraschall) zwei- oder dreidimensional aufgenommen. Dieser Prozess allein erzeugt keine diskreten semantischen oder strukturellen Einheiten in den aufgenommenen Daten (wie Wörter oder Noten). Dazu ist die Aktivität "Datenerkennung" notwendig. Unter

Bilderfassung fallen auch Tätigkeiten wie Scannen oder Digitalphotographie.

### **1.6. Aufzeichnen**

Beschreibt das Aufzeichnen von Audio- oder Videodaten; das Ergebnis ist eine digitale Audio- (z.B. WAV, MP3, etc.) oder Videodatei (MP4, Quicktime, etc.), die bearbeitet, analysiert und/oder archiviert werden kann.

### **1.7. Transkription**

Transkription bedeutet die vollständige schriftliche Erfassung eines Manuskripts (mitunter in Verbindung mit Anreicherung) oder einer Audio- bzw. Videoaufnahme. Die Repräsentation erfolgt in der Regel (auch) bei verbalen Aspekten von Aufnahmen textuell und strukturiert, z.B. in Form von Sprechakten. Sie kann jedoch auch multimodale Informationen wie Gesten oder Ereignisse und multimediale Informationen wie zeitliche Ebenen oder Relationen zu Metadaten enthalten. Selektive oder unvollständige Transkriptionen sind hier nicht berücksichtigt und sind eher als Annotation zu verstehen.

## **2. Erzeugen**

Das Erzeugen von Dingen bedeutet hier in erster Linie die Erstellung von genuin digitalen Objekten (born-digital) und nicht die (Retro-)Digitalisierung bestehender analoger Objekte. Das Erzeugen ist zum Beispiel das Schreiben (vgl. Schreiben) eines Textes oder, im generellen Sinne des Erzeugens einer Aneinanderreihung einzelner Zeichen, Schreiben von Quellcode (vgl. Programmieren), Komponieren oder die Erstellung eines Bildes.

### **2.1. Design**

Design bezeichnet die Entwicklung einer Oberfläche, mittels derer der Benutzer mit einem Programm interagieren und Aufgaben ausführen kann. Es beschreibt auch die Entwicklung unter Berücksichtigung benutzerfreundlicher Aspekte, wie Nutzen, Einfachheit und Effizienz des Systems. Diese Aktivität umfasst nicht die Umsetzung dieses Designs in Code (siehe "Programmieren" oder "Webentwicklung"). Der Aufbau von Datenbanken ist als "Modellierung" zu verstehen.

### **2.2. Programmieren**

Das Erstellen von Quellcode (in Form von kleineren Skripten bis hin zu umfangreichen Programmen), der von einem Computer ausgeführt werden kann. Dies schließt auch die Erstellung von Prototypen mit ein, also die Erstellung von Quellcode zu Testzwecken oder Modellierung.

Programmieren ist eng verbunden mit der breiter angelegten Aktivität Toolentwicklung. Programmieren ist nicht dasselbe wie Kodieren (das Anreichern eines Dokumentes oder Teilen

davon mit strukturellen, layoutbezogenen, semantischen oder ähnlichen Daten).

### **2.3. Übersetzen**

Übersetzen bedeutet die Erstellung eines neuen linguistischen Objektes, das auf einem anderen Objekt basiert, aber in einer anderen Sprache verfasst ist. Dies bezieht sich sowohl auf natürlichsprachliche Dokumente als auch auf maschinenlesbaren Quellcode.

### **2.4. Webentwicklung**

Das Erstellen von Webseiten, entweder im Rahmen einer Plattform (z. B. Content Management Systeme wie Drupal, WordPress oder Omeka) oder durch Schreiben von HTML- und CSS-Dateien. Das Schreiben von Modulen oder Plugins für eine Plattform oder das Programmieren webbasierter Programme ist als "Programmieren" zu bezeichnen.

### **2.5. Schreiben**

Schreiben bezeichnet die Aktivität, bei der neuer Text erzeugt wird (im Unterschied zum Erfassen von existierendem Text). Im vorliegenden Kontext bedeutet dies das Verfassen von Aufsätzen und Berichten, schließt aber andere Schrifterzeugnisse nicht aus.

## **3. Anreichern**

Anreichern bezeichnet das Hinzufügen von Information zu einem bestimmten Objekt, zum Beispiel genauere Angaben zu Herkunft, Beschaffenheit, Struktur oder Bedeutung. Diese Aktivität folgt typischerweise auf das Erfassen eines Objekts.

### **3.1. Annotieren**

Annotieren bedeutet, Informationen bezüglich eines digitalen Objektes zu erheben, beispielsweise können Kommentare, Metadaten oder Schlagwörter einer digitalisierten Repräsentation beigefügt oder in einer separaten Datei gespeichert werden. Annotationen dienen dazu, einzelne Teile zu kommentieren oder zu kontextualisieren (erklärende Annotationen), um strukturelle oder linguistische Merkmale aufzuzeichnen (strukturelle/linguistische Annotation), um in Form von Linked Open Data die Beziehung zwischen verschiedenen Objekten maschinenlesbar zu machen oder, im Falle generischer Metadaten, Informationen über ein gesamtes Objekt zu speichern. Kodierung ist eine Technik, die, so wie auch POS-Tagging, Tree-Tagging und Georeferenzierung, dem Annotieren verbunden ist.

### **3.2. Bereinigen**

Das Bereinigen von Daten hat zum Ziel, die Güte eines bestehenden digitalen Objekts zu verbessern. Darunter fallen Aktivitäten wie Fehlerkorrektur in geschriebenen Texten oder OCR-Ergebnissen, Debuggen von Code oder Verbesserung der Qualität von Video-, Audio- oder Bilddateien.

### **3.3. Bearbeiten**

Bearbeiten bezeichnet das Explizit-Machen struktureller, Layout-spezifischer, semantischer oder sonstiger Informationen über einen bestimmten Teil eines Dokuments durch das Hinzufügen (inline oder stand-off) von Markup. Häufig erfolgt es im Rahmen der übergeordneten Aktivität des wissenschaftlichen Edierens textlicher, musikalischer oder sonstiger Quellen. Es baut auf der Transkription eines Dokuments auf (dem Ergebnis von Data Recognition) und wird vom Modell des Dokuments geleitet (das Ergebnis der Modellierung).

## **4. Analyse**

Dieses allgemeine Forschungsziel bezieht sich auf die Aktivität des Extrahierens jeglicher Information aus offenen, geschlossenen, strukturierten oder unstrukturierten Datensammlungen, auf das Entdecken wiederkehrender Phänomene, Einheiten, Elemente, Muster, Gruppen und ähnlichem. Es kann sich auf strukturelle, formale oder semantische Aspekte von Daten beziehen. Die Analyse schließt auch Methoden ein, welche zur Visualisierung der Ergebnisse verwendet werden. Methoden und Techniken zur Analyse schließen oftmals an die Ergebnisse von Erfassen und Anreichern an; das Anreichern hängt dabei jedoch von Vorannahmen, Forschungsfragen und Analyse-Ergebnissen ab.

### **4.1. Inhaltsanalyse**

Bei der Inhaltsanalyse handelt es sich um eine Methode, die das Ziel verfolgt, Aspekte digitaler Objekte im Hinblick auf ihre Bedeutung, z.B. das Identifizieren von Konzepten oder bedeutungstragender Einheiten, zu analysieren. Einschlägige Techniken sind u.a. Topic Modeling, Sentiment Analysis, Information Retrieval, Discourse Analysis sowie Named Entity Recognition.

### **4.2. Netzwerkanalyse**

Netzwerkanalyse ist eine Methode zur Untersuchung von (realen oder fiktionalen) Beziehungen zwischen Akteuren oder anderen Entitäten eines gegebenen Netzwerks, das die Form eines sozialen, akademischen oder online-Netzwerks, einer Korrespondenz oder eines literarischen Werkes haben

kann. Das resultierende Netzwerk besteht für gewöhnlich aus Knoten (Entitäten) und Kanten (Beziehungen). Eine einschlägige Technik ist Named Entity Recognition. Sind hingegen die Artefakte (Texte, Bilder etc.) selbst und ihre Bezüge untereinander Gegenstand der Untersuchung, ist 'Analyse von Relationen' die entsprechende Forschungsaktivität.

### **4.3. Analyse von Relationen**

Die Analyse von Relationen bezieht sich auf computergestützte Verfahren, die dazu dienen, spezifische Relationen zwischen mehreren Forschungsobjekten aufzudecken. In textbasierten Untersuchungen könnte dies z.B. das Identifizieren von Gemeinsamkeiten verschiedener Texte (textuelle Nachnutzung / Plagiat), das Aufdecken von Versionen eines Textes (Kollationierung) oder die Bewertung im Hinblick auf Textähnlichkeit mittels stilistischer Merkmale (Stilometrie) beschreiben. In Analogie können diese Verfahren auch auf andere kulturelle Artefakte angewandt werden, z.B. Musik, Film oder Malerei. Zugehörige Techniken umfassen Sequence Alignment, Kollationierung und Techniken, die zur Stilanalyse gezählt werden.

#### **4.4. Räumliche Analyse**

Die Räumliche Analyse hat das Entdecken von Trends und Mustern in Daten zum Ziel, die räumliche oder geographische Aspekte betreffen. Eine Räumliche Analyse basiert häufig auf mithilfe von Annotationstechniken, wie Georeferenzierung oder Named Entity Recognition aufbereiteten Daten und kann Visualisierung oder Modellierung in Form von Kartierung zur Folge haben.

#### **4.5. Strukturanalyse**

Strukturanalyse betrifft die Analyse von Objekten auf der Ebene der Relationen struktureller Elemente eines kulturellen Artefakts (z.B. die Morphologie oder Syntax in der Linguistik). Relevante Techniken umfassen POS-Tagging, syntaktisches Parsen, Kollokationsanalyse, Konkordanzen.

#### **4.6. Stilistische Analyse**

Die Stilistische Analyse umfasst das Identifizieren stilistischer oder formaler Eigenschaften digitaler Objekte. Auch wenn rechnergestützte Stilistik in vielen Fällen auf Texte angewendet wird und auf linguistischen Eigenschaften aufbaut, kann sie auch auf andere Medien angewandt werden wie z.B. physikalische Artefakte, Malerei, Musik oder Film. Einschlägige Techniken sind Stilometrie, Hauptkomponentenanalyse, Clusteranalyse oder paleographische Analysen.

## **4.7. Visualisierung**

Visualisierung bezieht sich auf Aktivitäten, die zur Zusammenfassung und grafischen Präsentation von Daten dienen und diese grafischen Formen analytisch verwenden, um z.B. Muster, Strukturen oder interessante Aspekte in den dahinterliegenden Daten zu entdecken. Dabei kann jegliche Art von Daten visualisiert werden und Visualisierungen können in Form von Bildern, Karten, Zeitleisten, Graphen, Tabellen o.ä. umgesetzt werden. Einschlägige Techniken umfassen Plotten und Kartieren.

## **5. Interpretation**

Durch die Interpretation wird den im Zuge der Analyse betrachteten Phänomenen Bedeutung zugeschrieben. Für gewöhnlich folgt Interpretation daher auf die Analyse, wobei man auch der Auffassung sein kann, dass die Interpretation die hermeneutische Perspektive jeglicher Analysemethode definiert.

### **5.1. Kontextualisieren**

Kontextualisieren bezeichnet das Schaffen von Assoziationen zwischen einem betrachteten Objekt und anderen etablierteren oder besser verstandenen Objekten in einer bspw. geografischen, zeitlichen oder thematischen Nähe zum Untersuchungsgegenstand. Ziel ist es, diesem eine Bedeutung zuzuschreiben. Diese Art der Kontextualisierung kann auf vorhandene Annotationen und/oder Metadaten aufbauen.

### **5.2. Modellieren**

Modellieren ist die Aktivität des Schaffens einer abstrakten Repräsentation eines komplexen Phänomens, meist maschinen-lesbar, ggfs. aber auch auf interaktive Art (es umfasst z.B. die Simulation). Modelle werden maschinenlesbar, wenn die Modellierung ein Schema produziert, welches die Elemente und Struktur des Forschungsobjekts explizit beschreibt. Modellierung kann sich auch auf die Aktivität einer Transformation oder Veränderung eines digitalen Objekts beziehen, sofern es dazu dient, Kompatibilität zu einem zuvor geschaffenen Modell oder Schema zu erreichen. Kartieren wäre ein Beispiel für ein

räumliches Modell. Workflow Design ist Teil von Modellieren, indem es den Prozess als Objekt untersucht.

### **5.3. Theoriebildung**

Theoriebildung ist eine Methode, die darauf abzielt, eine Anzahl von Elementen oder Ideen in einem kohärenten System miteinander in Beziehung zu setzen. Das System basiert dabei auf allgemeinen Prinzipien und ist in der Lage, relevante Phänomene und Beobachtungen zu erklären. Theoriebildung stützt sich auf Techniken wie Argumentation, abstraktes Denken, Konzeptualisierung und Definition. Eine Theorie kann in Form eines Modells implementiert werden, ein Modell kann jedoch auch den Anstoß zur Formulierung einer Theorie geben.

## **6. Speicherung**

Speicherung (Storage) bezieht sich auf die Aktivität des Herstellens digitaler Kopien von untersuchten Objekten, Forschungsergebnissen oder Software und Diensten sowie ihre nachhaltige Erreichbarkeit. Die gespeicherten Dateien müssen nicht notwendigerweise öffentlich zur Verfügung stehen.

### **6.1. Archivieren**

Archivieren schließt den Prozess ein, Daten und weiteren Ressourcen einen gesonderten Ort der Speicherung zuzuweisen. Wenn es um Langzeitarchivierung geht, spielt auch die Aktivität Konservierung eine Rolle.

### **6.2. Identifizieren**

Identifizieren meint die Aktivität des Benennens und/oder Zuschreibens von (möglicherweise eindeutigen und/oder persistenten) Signaturen/Identifiern zu Forschungsgegenständen oder jeglicher Form digitaler Objekte. Das Hinzufügen einer Metadatenbeschreibung des Objekts ist Bestandteil von Annotation.

### **6.3. Organisieren**

Organisieren beschreibt das Ordnen von Objekten (Forschungsmaterialien, Datensets, Bilddateien, etc.) auf eine Art, die weitere Forschungsaktivitäten erleichtert. Es kann auch Aktivitäten einschließen, die das Entdecken unterstützen, wie z.B. Erzeugen von Metadaten und Anreichern.

## **6.4. Konservierung**

Konservierung ist das Anwenden bestimmter Strategien, Aktivitäten und Technologien mit dem Ziel, eine genaue Wiedergabe des digitalen Content auch zukünftig zu gewährleisten. Es erleichtert das Wiederverwenden von Forschungsdaten, (Forschungs)Objekten und verwandten Ressourcen und kann Aktivitäten einschließen, die sich mit Nachhaltigkeit und Interoperabilität befassen. Zu den relevanten Techniken gehören unter anderem Bitstream Preservation, Durable Persistent Media, Emulation, Hinzufügen von Metadaten, Datenmigration, Replikation, Technologiekonservierung, Versionierung, der Gebrauch von Offenen Archiv-Informationen-Systemen und Standards, die Interoperabilität unterstützen.

## **7. Dissemination**

Dissemination bezieht sich auf die Aktivität, Untersuchungsgegenstände, Forschungsergebnisse, Software oder Services anderen Forschenden oder einer breiteren Öffentlichkeit auf mehr oder weniger formale Weise zur Verfügung zu stellen. Dissemination beruht auf der Speicherung und kann das Veröffentlichen oder Teilen von Daten beinhalten, wobei eine große Bandbreite von Methoden und Techniken relevant ist, zu denen unter anderem die Anwendung von Linked Open Data gehört.

### **7.1. Kollaboration**

Kollaboration spielt immer dann eine Rolle, wenn eine beliebige Forschungsaktivität von verschiedenen Wissenschaftlern gemeinsam ausgeübt wird, möglicherweise auch an unterschiedlichen Orten und zu unterschiedlichen Zeiten. Forschungsorientierte Kollaboration wird insbesondere durch umfassende Digitale Forschungsumgebungen ermöglicht, es kann sich dabei aber auch um spezifische Aktivitäten handeln, wie etwa Kommunikation oder das Teilen von Ressourcen.

### **7.2. Kommentieren**

Kommentieren beschreibt das Hinzufügen von Informationen zu einer Dateneinheit, für gewöhnlich auf eine Art, die zwischen den Daten, auf die sich der Kommentar bezieht, und dem Kommentar selbst unterscheidet. Es dient für gewöhnlich dazu, eine Meinung auszudrücken, Kontextinformationen beizufügen oder Kommunikation und Kollaboration über das kommentierte Objekt anzuregen. Es

unterscheidet sich daher vom "Annotieren" (wie hier beschrieben), das sich auf das Hinzufügen deskriptiver oder erläuternder Informationen zu Sektionen eines Objektes bezieht und das Ziel hat, inhärente Qualitäten, Strukturen oder Bedeutungen dieser Sektion zu explizieren.

### **7.3. Kommunizieren**

Kommunizieren bezieht sich auf die Aktivität des Ideenaustauschs mit anderen Leuten, meistens, aber nicht ausschließlich, mit sprachlichen Mitteln. Relevante Techniken schließen E-Mail, Chat und Audiokonferenzen ein.

### **7.4. Crowdsourcing**

Crowdsourcing bezieht sich auf das Paradigma eines von Usern generierten Inhalts in einem Web-2.0-Kontext, der sich hier auf die Domäne digitaler geisteswissenschaftlicher Forschung bezieht. Crowdsourcing kann Gamification einschließen, das als Form der Motivierung zur Beteiligung an Crowdsourcing-Ansätzen verstanden werden kann.

### **7.5. Veröffentlichen**

Veröffentlichen bezeichnet die Aktivität, die Objekte jeglicher Art für die Öffentlichkeit formal verfügbar macht. Dabei kann es sich um Forschungsobjekte, Forschungsdaten, Forschungsergebnisse oder Tools und Dienste handeln. Veröffentlichen kann geschützt oder per Open Access / Open Source erfolgen und Forschungsergebnisse können in Print- oder digitalen Formaten veröffentlicht werden.

## **7.6. Teilen**

Teilen meint das öffentliche Verfügbarmachen von Objekten mittels informeller Kanäle wie Blogs, Code-Sharing-Seiten wie GitHub oder sonstigen Social-Media-Plattformen.

## **8. Rahmenaktivitäten**

Bei Rahmenaktivitäten handelt es sich um solche Aktivitäten, die sich, anders als reguläre Aktivitäten, nicht unmittelbar auf ein Forschungsobjekt beziehen, sondern auf eine Kombination von Forschungsaktivität und Forschungsobjekt. Beispiele wären das Unterrichten digitaler Musikauszeichnung oder die Überprüfung der Qualität von Mustererkennung in Bildern. Rahmenaktivitäten-Tags können hinzugefügt werden, um zusätzlichen Kontext für typische Aktivität+Objekt Tag-Paare zu schaffen. In einigen Fällen können sich Rahmenaktivitäten auch auf Objekte beziehen, etwa bei abstrakten Objekten wie "Infrastruktur" oder "Digitale Geisteswissenschaften".

### **8.1. Prüfen**

Prüfen bezieht sich auf die Aktivität, die die Existenz bestimmter Eigenschaften verifiziert, meist im Hinblick auf die Qualität der Ergebnisse einer Aktivität. Es kann sich auf das Begutachten von Aufsätzen oder Konferenzbeiträgen, auf die Evaluierung der Kohärenz von Annotationen audio-/visueller Materialien oder auf die Beurteilung des Werts der Digitalen Geisteswissenschaften beziehen.

### **8.2. CommunityBuilding**

CommunityBuilding beschreibt die Aktivität des Schaffens oder Ausbaus einer Community mit gemeinsamen Interessen. Es kann Dissemination und Lehre ebenso umfassen wie das Fördern bestimmter Aktivitäten, Praktiken oder Werte.

### **8.3. Überblick Geben**

ÜberblickGeben bezieht sich auf das Liefern von Informationen, die verhältnismäßig generell sind oder einen historischen oder systematischen Überblick zu einem gegebenen Thema verschaffen. Es kann sich dabei jedoch sowohl an Experten als auch an Novizen eines Feldes, einer Domäne oder eines Fachgebiets richten.

### **8.4. Projektmanagement**

Projektmanagement umfasst Aktivitäten wie das Entwickeln von Strategien und das Verantworten der Projektleitung ebenso wie Aufgabenverwaltung, z.B. das Dokumentieren von Arbeitsschritten, Terminen und weiterer relevanter Informationen. Es kann Aktivitäten wie Planen, Dokumentieren und das Einwerben/Beziehen von Fördergeldern einschließen, aber auch das Versenden von Erinnerungen und Statusberichten. Projektmanagement steht in Bezug zu Kollaboration.

### **8.5. Lehren/Lernen**

Lehren/Lernen umfasst eine Gruppe von Leuten, die auf interaktive Art einer anderen Gruppe dabei hilft, Fähigkeiten, Kompetenzen und Wissen zu erlangen und/oder zu entwickeln, welche es ihnen erlauben, Probleme in einem bestimmten Forschungsfeld zu lösen.