

Zugang zu Informationen in digitalen Sammlungen: Fokus Archive

Tobias Hodel, Sonja Gasser, Christa Schneider, David Schoch¹

Inhalt

1. Archive: Endstation in Vielfaltigkeit	28
2. Zugänge als Herausforderung	30
3. Visuelle Zugangsformen	33
3.1 Karten und geografische Visualisierungen	41
3.2 Netzwerk	44
3.3 Zeitstrahl	46
3.4 Plots	56
3.5 Visualisierung der iterativen Digitalisierung	64
3.6 Kombinationen	65
3.7 PolyCube-System	68
3.8 Virtueller Raum	71
4. Semantische Zugangsformen	72
4.1 Storytelling	72
4.2 Identifizierbare Entitäten	73
4.3 Inhaltsextraktion	74
4.4 Text-Mining	77
5. Querschnittsthemen	78
5.1 Records in Contexts (Linked Open Data Modell für Archive)	78
5.2 Linked Data als game changer	80
5.3 Maschinelle Lernverfahren	81
5.4 Crowdsourcing & Citizen Science: Annotation und Kuratation	83
5.5 Standardisierte Publikation von Bildern via IIIF und Viewer	85
6. Fazit	85
Literaturverzeichnis	87

¹ Digital Humanities, Walter Benjamin Kolleg, Universität Bern.

1. Archive: Endstation in Vielfältigkeit

Die vorliegende Studie bietet eine aktuelle Literaturübersicht über den Zugang zu Informationen in digitalen Archiven.² Damit wird auf den gegenwärtigen Stand, sich abzeichnende Entwicklungen und vielversprechende Technologien für den Zugriff, die Analyse und Auswertung von digitalen Informationen durch Nutzende der digitalen Archive eingegangen. Der Fokus liegt auf Institutionen, die Dokumente unterschiedlichster materieller Ausprägung (von digitalen Daten über Papier, Pergament, Fotografie, Film- und Tondokumenten bis zu Objekten), aber auch diverser Formen (Digitalisate, born-digital) und Inhalte (Literatur, Kunst, Verwaltungsdokumente usw.) archivieren. Der Zugang zu öffentlichen Archiven und Archivmaterialien ist einerseits rechtlich zugesichert, andererseits zum Schutz von Rechtstiteln, insbesondere der Persönlichkeitsrechte, streng kontrolliert, weshalb komplexe Abwägungen zwischen offenen (Meta-)Daten und geschlossenen Beständen getroffen werden müssen. Auf die rechtlichen Belange und damit zusammenhängenden Probleme, insbesondere Schutzfristen, die gewisse Einschränkungen für die Nutzung der Daten mit sich bringen, wird in dieser Studie nicht weiter eingegangen.³

Die Annäherung an das Thema des Zugriffs erfolgt aus drei Richtungen, die zugleich die Gliederung dieses Beitrags vorgeben. Auf die Inhalte in digitalen Archiven kann zugegriffen werden:

1. über visuelle Zugangsformen,
2. über semantisch/textuelle Such- und Findwege und
3. über Querschnittsthemen.

Während es sich bei den unter Punkt Eins und Zwei genannten Formen um Zugangsarten handelt, beinhaltet Punkt Drei Ansätze, um Daten und Dokumente strukturell aufzubereiten und dadurch in einem informationstechnologischen Sinne zugänglich zu machen. Dieser dritte Punkt umfasst technische Ansätze wie *machine learning* und die Erschließung mit Linked Data sowie damit in Verbindung stehende Ansätze der Datenannotation.⁴

Die Studie ist in Kooperation mit dem Schweizerischen Bundesarchiv entstanden, um den aktuellen Stand der Forschung im Bereich des Zugangs zu digitalen Archiven zu erarbeiten. Dementsprechend geht es in diesem Beitrag darum, die

2 Das vorliegende Dokument entstand in der Verantwortlichkeit von Tobias Hodel unter Mitarbeit von Sonja Gasser (visuelle Zugriffe, Beschreibung des Schweizerischen Bundesarchivs, Recherche weiterer Zugänge), Christa Schneider (Lektorat, Recherche und Erstellung einer inhaltlichen Kohärenz) und David Schoch (Recherche und Zusammenfassung von Aufsätzen). Eine Bibliografie aller verwendeten Arbeiten ist online via Zotero verfügbar: https://www.zotero.org/groups/2386895/dh_unibe/collections/NM3SDA3A. Wir danken dem Schweizerischen Bundesarchiv, insbesondere Dr. Stefan Nellen, für die Kooperation und kritischen Rückmeldungen.

3 Für eine Auslegeordnung der rechtlichen Situation für Text und Data Mining in Deutschland sollte jedoch folgender Zeitschriftenartikel beachtet werden, in dem vorwiegend urheberrechtliche Abwägungen und nicht datenschutzrechtliche Einschränkungen vorgenommen werden: Schöch u. a., Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen.

4 Thematisiert wird insbesondere Records in Context (RIC): <https://www.ica.org/en/egad-ric-conceptual-model>.

Bandbreite an visuellen und semantischen Zugängen aufzuzeigen, die alternativ oder in Ergänzung zu textuellen Suchen («Volltextsuche» bzw. «Stichwortsuche») für historische oder kulturelle Bestände und Sammlungen in Betracht gezogen werden können und sollen. Berücksichtigt werden Zugangsformen, die den Blickwinkel erweitern und bewusst aus einem breiten Feld von Kulturinstitutionen wie Museen, Archiven und Bibliotheken stammen.

Da die einzelnen Kulturinstitutionen ähnliches mit unterschiedlichem Vokabular bezeichnen oder gewisse Begrifflichkeiten mit leicht unterschiedlicher Bedeutung verwendet werden, seien hier kurz einige Begriffe definiert, wie wir sie verstehen. Ist im vorliegenden Beitrag von «Sammlungen» die Rede, bezieht sich das generell auf eine unspezifizierte Menge an kulturellen Artefakten, die in Institutionen gesammelt und verwahrt werden. Darunter fallen etwa Kunstwerke und Objekte (typischerweise in Museen), Bücher und Sondersammlungen (in Bibliotheken) oder Archivalien (in Archiven). Ein «Bestand» ist ein ausgewählter Teil einer Sammlung, was sowohl auf die Gattung, die Systematik der Aufbewahrung als auch auf den Inhalt der Artefakte bezogen sein kann. Damit unterscheidet sich unser Verständnis eines Bestandes von engen archivischen Definitionen. Ein «Dokument» enthält schriftliche Informationen und stammt somit aus dem Archivkontext, während ein «Objekt» eher Teil einer Museumssammlung ist und die Materialität betont. «Material» meint mit Bezug zu Sammlungsobjekten, woraus das Medium (bspw. der Bildträger) besteht. Im Zusammenhang mit dem Archiv («Archivmaterial») sind jedoch generell Archivalien gemeint. «Digitales Dokument» oder «digitales Objekt» steht für Digitalisate von physischen Dokumenten und Objekten, wie sie in Datenbanken enthalten sind, worunter auch *digital-born* Dokumente oder Objekte fallen. «Digitale Informationen» sind die in einer Datenbank enthaltenen Datensätze, die die Beschreibung des Dokuments oder Objekts, für das sie stehen, in Form von Metadaten enthalten und auch eine Bild-, Text-, Video- oder Audiodatei umfassen können. Die Begriffe «Archivplan», «Dossier» und «Findmittel» verwenden wir entsprechend archivischer Gepflogenheiten. Der Archivplan bildet somit die Tektonik des Archivs ab. Ein Dossier (bzw. ein Band) bildet die kleinste verzeichnete Einheit. Findmittel sind in diesem Zusammenhang Datenbanken, die die Suche in den Beständen nach vorgegebenen Verzeichnungsstandards (aktuell ISAD (G)) erlauben und typischerweise aus Zettelkasten erwachsen sind.

Die Studie basiert folglich auf einer Vielfalt von institutionellen und kulturellen Voraussetzungen, auf denen die in kulturellen Datenbanken enthaltenen Inhalte beruhen. Bei der Untersuchung von vielversprechenden digitalen Zugängen wird auf grundlegende Prinzipien aus dem Bereich der Informationsvisualisierung zur Gestaltung von Interfaces genauso eingegangen wie auf dahinter liegende Technologien. Diese schaffen die nötige Datenstruktur und nehmen Kontextualisierungen vor, die zur Auffindbarkeit der digitalen Objekte und zur Darstellung von Suchresultaten beitragen. Neue Formen des Zugangs unterstützen Nutzende dabei, das Gesuchte zuverlässiger auffinden zu können und an unbekannte Inhalte herangeführt zu werden. Ein

damit unterschwellig verbundenes Ziel ist, bestenfalls neue Nutzungsgruppen zu erschliessen. Worüber die Vielzahl der gesichteten Projekte, Beispiele und analysierten Publikationen jedoch keine Auskunft geben, ist, wie hoch der zeitliche Aufwand zur Umsetzung der beschriebenen Anwendungen, Websites, Portale oder Angebote ist. Diese für die Praxis wirtschaftlich relevante Frage, stellt sich im akademischen Umfeld so nicht und fliesst deshalb auch nicht in Publikationen ein. Das Interesse in diesen Projekten bezieht sich vielmehr darauf, im Rahmen der Forschung ein Resultat zu präsentieren, optimalerweise zusammen mit kommentiertem Code.

2. Zugänge als Herausforderung

Kulturelle Daten (Bilder, Texte, Musik, Film, 3D-Objekte usw.), wie sie typischerweise in digitalen Sammlungen von Bibliotheken, Archiven und Museen zu finden sind, zeichnen sich durch eine Vielfalt an zugehörigen, oft heterogenen Metadaten (Autor- und Herausgeberschaft, Titel, Gattung, Beschreibung, Masse, Datierung, Lokalisierung, Provenienz usw.) aus. Die Einbettung der kulturellen Objekte, auf die sich diese Daten und Metadaten beziehen, in historische und kulturelle Kontexte führt dazu, dass solche Artefakte verschiedene Bedeutungs- und Betrachtungsebenen aufweisen. Kulturelle Daten sind daher deutlich komplexer als etwa naturwissenschaftliche Messdaten, deren Werte von Anfang an strukturiert in Tabellen abgebildet werden können. Mit der Digitalisierung von Sammlungen an Kulturinstitutionen und der Erfassung von Beständen in Datenbanken liegt aber zunehmend auch aus kultureller Sicht strukturiertes Datenmaterial vor, das für weiterführende digitale Projekte verwendet werden kann.

Im Fall der Archive liegen die einst in Katalogen und Zettelkästen erfassten Information mittlerweile digital in Datenbanken vor, sodass heutzutage die Findmittel für Archivalien online konsultiert werden können. Für eine Metasuche über verschiedene Schweizer Archive hinweg existiert mit *Archives Online* ein gemeinsames Suchportal.⁵ Die Archivalien selbst liegen nur in einzelnen Fällen und bei weitem nicht flächendeckend als digitalisierte Dokumente vor, auch wenn die Entwicklungen in diese Richtung deuten, wie das Schweizerische Bundesarchiv belegt.⁶ Archivdatenbanken haben ähnlich wie Bibliothekskataloge zum Ziel, die Informationen digital bereitzustellen, die zum Auffinden der Dossiers bzw. Dokumente notwendig sind. Zugriff auf die Materialien erhält man deshalb weitestgehend immer noch nur direkt in der Institution.

5 <https://archives-online.org/Search>

6 Mit dem im Herbst 2019 aufgeschalteten Online-Zugang zum Bundesarchiv (<https://www.recherche.bar.admin.ch/recherche/>) sind alle analogen oder digitalen Dossiers des Bundesarchivs über das Webportal zugänglich. Ein Besuch im Lesesaal ist, von Ausnahmen abgesehen, seither nicht mehr nötig. Benötigte, noch nicht digitalisierte Dokumente können über einen Digitalisierungsauftrag angefordert werden: <https://www.recherche.bar.admin.ch/recherche/#/de/informationen/bestellen-und-konsultieren>.

Es ist davon auszugehen, dass die Digitalisierung in allen Erinnerungsinstitutionen weiter voranschreitet. Damit werden zusätzlich zu den Metadaten stärker digitale Objekte in die Findmittel eingebunden oder über digitale Lesesäle zugänglich gemacht. Entsprechend liegt eine zentrale Herausforderung für Bibliotheken, Museen und Archive darin, die sinnvolle Verknüpfung von Daten(-beständen) herzustellen und um durchdachte Zugangsformen zu erweitern. Damit soll einerseits die Nutzung erleichtert, andererseits die Ausweitung der Nutzungsgruppe erreicht werden. Die entsprechenden Portale richten sich im Moment hauptsächlich an Fachpersonen, um das gezielte Aufspüren von Archivalien, Publikationen oder Objekten zum genaueren Studium zu ermöglichen. Mit den entsprechenden Anstrengungen ist indes auch eine weiterführende Nutzung denkbar, die etwa die Weiterverarbeitung von Daten und Metadaten beinhalten könnte. Durch attraktive Zugänge und die Aufbereitung der Inhalte wäre ausserdem die wünschenswerte Nutzung der Portale durch ein breites Publikum vorstellbar. Diese Entwicklungen gehen selbstredend mit verschiedenen Digitalisierungsvorgängen einher, im Zuge derer neben Metadaten auch digitale, bzw. digitalisierte Dokumente und Objekte bereitgestellt werden.

Der Zugang zu Beständen und Dokumenten folgt als letzter Schritt in einem Archivierungsprozess, der aus ausgesprochen unterschiedlichen und teilweise gegenläufigen Vorgängen und Aufbereitungsformen besteht. Einerseits müssen Metadaten zusammen mit digitalen Objekten nach gewissen Standards verfügbar gemacht werden und sollen andererseits entsprechend der technischen Möglichkeiten ausgewertet und angereichert werden. Technische und visuelle Erwartungen kollidieren so mit etablierten Standards und historisch gewachsenen Beschreibungsmethoden. Erschwerend kommt hinzu, dass unterschiedliche Nutzungsgruppen verschiedene, ausgesprochen spezialisierte Zugriffsformen formulieren, die untereinander oftmals unvereinbar sind.

Um vor allem aus technischer Perspektive und mit Fokus auf Bibliotheken einen Überblick zu neuen Lösungen für die eben skizzierten Probleme zu bieten, versucht der weit beachtete IFLA Trend Report (IFLA steht für International Federation of Library Associations and Institutions) den aktuellen Stand der Diskussion abzubilden.⁷ Kurz gefasst identifiziert der Report fünf Trends:

1. Neue Technologien werden den Zugang zu Informationen sowohl erweitern wie auch einschränken
2. Online-Lehre und Online-Bildung wird Lernen auf globaler Ebene demokratisieren und zerrütten (in der englischen Vorlage wird der Begriff *disrupt* gebraucht)
3. Die Grenzen der Privatsphäre und des Datenschutzes werden neu ausgehandelt
4. Hyper-vernetzte Gesellschaften werden neuen Stimmen und Gruppen Gehör verschaffen und diese ernst nehmen

7 Der Report ist online: <https://trends.ifla.org/>. Siehe einführend IFLA Trend Report, Neue Technologien erweitern den Informationszugang und schränken ihn gleichzeitig ein. Die deutsche Übersetzung des Reports findet sich hier: https://trends.ifla.org/files/trends/assets/ifla-trend-report_german.pdf.

5. Das globale Informationsökosystem wird durch neue Technologien transformiert

Für den vorliegenden Beitrag sind Trend 1 und 5 zentral, am Rande gestreift werden auch Trend 3 und 4. Ziel ist es entsprechend, den in Trend 1 geforderten Zugang zu Informationen in Aufbewahrungsinstitutionen zu skizzieren und weitere Möglichkeiten von Zugriff und Abfrage aufzuzeigen. Besonders augenfällig ist die Vogelperspektive auf die Trends, die zentrale Probleme wie Datenschutz und Kommodifizierung von Daten von Nutzerinnen und Nutzern aufzeigt, ohne dass dabei die Erinnerungsinstitutionen oder weitere Informationsdienstleister im Zentrum stehen.

Die Dialektik der Trends 1 und 2 bedürfen einer kurzen Vertiefung, da jeweils sowohl positive (erweitern/demokratisieren) als auch negative (einschränken/zerrütten) Begriffe mit ihnen zusammen genannt werden. Damit werden in Trend 1 neben den Versprechen einer technologisch aufgerüsteten Welt auch die Schattenseiten angesprochen, die etwa technische oder methodische Hürden andeuten, die den Zugriff auf den ersten Blick erleichtern, auf den zweiten jedoch Personen(-gruppen) häufig unabsichtlich ausschliessen. Zugang erhält, wer die entsprechende Infrastruktur nutzen kann (Stichwort Netzabdeckung und Kosten von Internet Providern), die notwendige Ausrüstung hat und die «Sprache» versteht (nicht nur im linguistischen Sinn, sondern auch aufgrund der Bildsprache eines Angebots oder der Führung von Nutzenden). Parallel dazu verhält es sich mit Bildungsformen, die aufgrund der Digitalisierung demokratischer (im Sinne von nutzbar unabhängig von sozialen oder monetären Voraussetzungen) zugänglich werden. Gleichzeitig führen neue digitale Angebote zu einer neuen Kommodifikation von Bildung und das nicht nur auf tertiärer Ausbildungsstufe. Neue (kommerzielle und non-profit) Einrichtungen drängen in einen Bildungsmarkt, der zu neuen Konkurrenzsituationen führt, die international und divers ausgerichtet sind.

Mit Überlegungen, wie Erinnerungsinstitutionen auf den erweiterten Zugang zu Informationen und die Transformation des Informationsökosystems durch neue Technologien reagieren können, verfolgen wir einen Ansatz, der verstärkt auf neuartige Formen von Wissensorganisation im Digitalen und anderen entsprechenden Zugängen setzt. In der Kontextualisierung von Dokumenten in Visualisierungen und durch semantische Verknüpfungen in den Metadaten wird es möglich, Wissen zu vermitteln und verschiedene Zugänge zum Archivmaterial zu schaffen, die über die Funktionsweisen gängiger Suchportale hinausgehen. Kontextualisierungen lassen sich bei kulturhistorischen Dokumenten und Objekten besonders gut über Personen, Themen, Ereignisse sowie geografische oder zeitliche Einordnungen vornehmen. Dadurch kann Wissen nicht nur konstruiert und vermittelt, sondern auch konkret an die Oberfläche gebracht und fassbar gemacht werden. Dabei darf die historische und kulturelle Bedingtheit von Zugangsformen weder vergessen noch ausser Acht gelassen werden. Einerseits sind die Erinnerungsinstitutionen historisch gewachsen, was ein bestimmtes Selbstverständnis sowie etablierte Strukturen und Aufgabenfelder mit sich bringt, an denen durch die digitale Transformation, die sich durch die

Gesellschaft zieht, teilweise gerüttelt wird. Andererseits werden auch die Erwartungen der Nutzenden durch Anspruchshaltungen geprägt, die aufgrund kultureller Bedingungen entstanden und aktuell durch digitale Entwicklungen verändert werden. Heutzutage wächst die Erwartung stetig, insbesondere im Zusammenhang mit jeglicher Art von Dienstleistungen, vieles digital vornehmen zu können, was lange auf analogem Weg vonstatten gegangen ist. Der Trend des *mobile-first* steht für ein (digitales) Archiv zwar nicht unbedingt an erster Stelle, dennoch ist zu bedenken, dass auch für kleine Bildschirmgrößen ein Zugang zu Web- und Visualisierungslösungen vorhanden sein sollte, was im unmittelbar damit verbundenen *responsive design* zu berücksichtigen ist.

Im Sinne eines möglichst weit gefassten Verständnisses von Zugang zu historisch und kulturell relevantem Wissen, das in Erinnerungsinstitutionen gespeichert ist, beschreiben wir im Folgenden Zugangsformen, die teilweise experimentell, teilweise auf Basis einer Erwartungshaltung Material zugänglich machen. Dabei verfolgen wir das Ziel, eine möglichst grosse Bandbreite an Umsetzungsformen darzustellen. Diese Umsetzungsformen müssen jedoch im Einzelfall wiederum mit den zu präsentierenden Beständen oder Dokumentengattungen abgeglichen werden. Zu beachten sind deshalb auch medientheoretische Grundlagen, da jede mediale Umsetzung von den Aussagemöglichkeiten, die mit einem bestimmten Medium verbunden sind, geprägt ist, was letztlich Auswirkungen auf die Aussagekraft jeder eingesetzten Zugangsform hat. Dies wird vor dem Hintergrund der Reduzierung von Formen des Ausschlusses und der Erweiterung des (imaginierten) Zielpublikums diskutiert.

3. Visuelle Zugangsformen

Dieser Teil nimmt eine Auslegeordnung von verschiedenen visuellen Zugangsformen zu kulturellen Daten vor und benennt, was mit den Darstellungsweisen und Konzepten, auf denen diese Anwendungen fussen, beabsichtigt wird. Zunächst werden die Grenzen von gängigen Suchportalen dargelegt, um, nach einer kurzen theoretischen Auseinandersetzung mit einigen grundsätzlichen Voraussetzungen von Visualisierungen, alternative Ansätze des Zugangs zu digitalen Sammlungsinhalten vorzustellen. Die gewählten Beispiele berücksichtigen sowohl eine Vielzahl von Visualisierungsformen, Sammlungen unterschiedlichster Ausrichtung als auch den Einbezug verschiedener digitaler Technologien und Herangehensweisen.

Viele digitale Sammlungen, wozu auch digitale Archive gezählt werden, von Museen, Bibliotheken und Archiven stützen sich auf eine Stichwortsuche, die den Nachteil hat, dass grosse Teile der Sammlung für Nutzende unter der Oberfläche verborgen bleiben. Die Serendipität, wie sie durch visuelle und andere alternative Herangehensformen möglich wäre, bleibt auf diese Weise sehr beschränkt. Mit seiner Kritik an diesem Zustand hat Mitchell Whitelaw die Bezeichnung «Generous Interface» für eine reichhaltige Suchoberfläche eingeführt, die es Nutzerinnen und Nutzern ermöglicht, die enthaltenen Inhalte zu durchstöbern und zudem über den Umfang und

die Komplexität einer digitalen Sammlung gleichzeitig Aufschluss gibt. Statt einer Suchabfrage, hinter der eine bestimmte Suchabsicht stecken muss, werden verschiedene Suchwege angeboten, die sowohl das Entdecken von Inhalten als auch das fokussierte Suchen unterstützen.⁸

Die mit dem Zugang zu digitalen Sammlungen verbundenen Problematiken befinden sich an einer interdisziplinären Schnittstelle von Information Retrieval, Human-Computer-Interaction (HCI), Informationsvisualisierung und den Digital Humanities. Das Urban Complexity Lab (UCLAB) der Fachhochschule Potsdam hat sich in den letzten Jahren einen Namen mit innovativen Interfaces zur Visualisierung von kulturellen Sammlungen gemacht, weshalb in diesem Teil der Studie viele Beispiele davon versammelt sind. Das Team um Prof. Dr. Marian Dörk am UCLAB zeichnet sich dadurch aus, dass Informatik, Interface Design, Datenvisualisierung und Kulturwissenschaft zusammenkommen. Der Erfolg von Projekten des UCLAB lässt sich zudem darauf zurückführen, dass sie in enger Kollaboration mit der jeweiligen Kulturinstitution und den dort tätigen Fachleuten unterschiedlichsten Hintergrunds (z. B. Kunsthistoriker*innen, Sammlungskonservator*innen, Dokumentationsspezialist*innen, Numismatiker*innen, Wissenschaftler*innen, Marketing- und PR-Verantwortliche, IT-Verantwortliche) in gemeinsamen Workshops erarbeitet werden.⁹ Bereits 2013 haben sich Marian Dörk u. a. kritisch mit Visualisierungen auseinandergesetzt, die die visuelle Vermittlung als Praxis der Überzeugungsarbeit verortet.¹⁰ Sie analysieren darin eine Vielzahl von Visualisierungen nach den vier Kriterien Offenlegung, Pluralität, Kontingenz und Empowerment. Diese Kriterien eignen sich, um jegliche Zugangsform kritisch zu hinterfragen.

Johanna Drucker beschäftigt sich aus der Perspektive der Geisteswissenschaften mit der Frage, was es bedeutet, wenn die Digital Humanities Visualisierungstechniken aus Disziplinen wie den Sozial- und Naturwissenschaften einsetzen, bei denen empirische und vor allem quantifizierende Herangehensweisen an Wissen üblich sind. Sie legt Wert darauf, dass für die Geisteswissenschaften Überlegungen anzustellen sind, wie mit Mehrdeutigkeit, Widersprüchlichkeit und anderen Merkmalen interpretativer und hermeneutischer Methoden umgegangen und darüber reflektiert wird, wie Visualisierungen und bestimmte Erscheinungsformen zustande gekommen sind.¹¹ Diese grundsätzlichen Überlegungen müssen auch für archivisches Material und dessen digitale Zugänglichmachung angestellt werden. Bislang befinden sich jedoch die Forschungen von Johanna Drucker und weiteren Personen in diesem Forschungsfeld immer noch in einer initialen Phase mit vorwiegend experimentellen Resultaten, sodass es keine etablierte Technik gibt, die in Bezug auf Archivmaterial übernommen werden könnte. Aufgrund der Vielfältigkeit von kulturellem Erbe und den aus

8 Whitelaw, *Generous Interfaces for Digital Cultural Collections*, S. 2–3.

9 Glinka/Pietsch/Dörk, *Past Visions and Reconciling Views: Visualizing Time, Texture and Themes in Cultural Collections*, S. 18–20; 37.

10 Dörk u. a., *Critical InfoVis*.

11 Drucker, *Graphical Approaches to the Digital Humanities*.

geisteswissenschaftlicher Perspektive gleichzeitig zahlreichen Möglichkeiten zur Auseinandersetzung mit diesen Inhalten, dürfte es eine universell einsetzbare Patentlösung ohnehin nicht geben und wäre auch nicht zielführend. Vielmehr wird es auf einzelne Sammlungen, deren Inhalt und den gewählten Fokus ankommen, welche Arten des digitalen Zugangs aus welchen Gründen im einzelnen Fall Verwendung finden. Es ist deshalb wichtig, verschiedene Möglichkeiten des Zugangs zu kennen sowie zu wissen, welche Technologien wo eingesetzt werden können und welche grundlegenden Gestaltungsmöglichkeiten existieren.

Es gibt unzählige Möglichkeiten, wie Daten visualisiert werden können.¹² Sie hängen davon ab, welche Werte in den Daten vorhanden sind und was sichtbar gemacht werden soll. Teilweise können verschiedene Visualisierungsformen miteinander verknüpft werden. Im Unterschied zu analogen können digitale Diagramme um interaktive Komponenten und damit weitere Informationsschichten erweitert werden. Zudem sind sie wandelbar, indem die Daten nahtlos von einer Visualisierungsform in eine andere überführt und auf diese Weise unterschiedlich dargestellt werden können. Eine abschliessende systematische Aufstellung von Visualisierungen gibt es nicht, aber es existieren gewisse Grundtypen von Diagrammen, die häufig anzutreffen sind.

Untenstehend sollen deshalb einige der grundlegenden Visualisierungstypen und Gestaltungsmittel aufgeführt werden, die in den anschliessend besprochenen Beispielen in verschiedensten Kombinationen wiederzufinden sind. Für die Grundtypen von Diagrammen halten wir uns an die von Johanna Drucker aufgestellte Typologie,¹³ ergänzen diese aber mit weiteren Visualisierungstypen und greifen zur Illustration auf Abbildungsbeispiele von D3.js zurück, einer verbreiteten JavaScript-Bibliothek, die für statische oder interaktive Grafiken im Web eingesetzt wird.¹⁴ Da die einzelnen Visualisierungstypen unterschiedliche Werte (z. B. numerische Werte, Zeitverläufe usw.) voraussetzen, sind diese nur für bestimmte Daten geeignet, oder die Daten müssen erst aufbereitet werden. Eine Darstellung vermag zudem oft nur gewisse Aspekte aufzuzeigen, die in den Daten vorhanden sind.

12 Für einen Überblick siehe die Darstellung von Ralph Lengler und Martin J. Eppler, online: https://www.visual-literacy.org/periodic_table/periodic_table.html.

13 Drucker, *Graphical Approaches to the Digital Humanities*, S. 239.

14 <https://d3js.org/>

Diagramme¹⁵

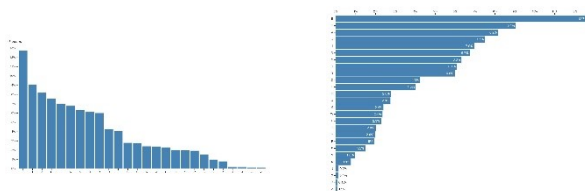


Abb. 1: Säulen- und Balkendiagramm. Werte vergleichen

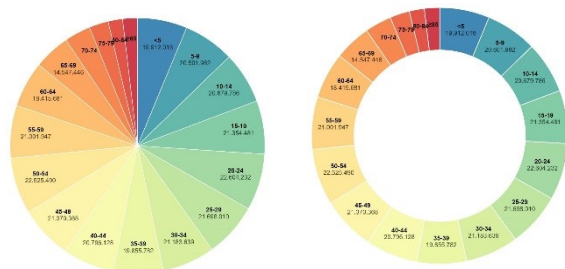


Abb. 2: Kreis- und Ringdiagramm. Prozentuale Werte

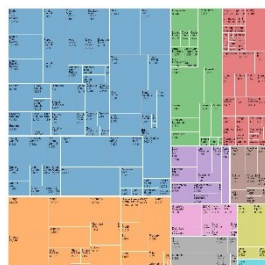


Abb. 3: Tree Map. Prozentuale Werte und hierarchische Strukturen

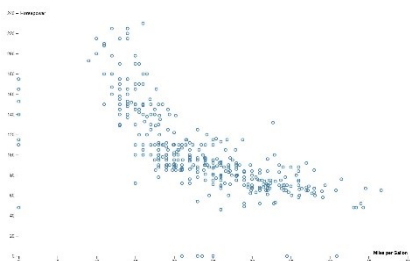


Abb. 4: Punktwolke (Scatter Plot). Diskrete Werte

15 Alle abgebildeten Beispiele in den Tabellen und zusätzliche Diagramme sind zusammen mit dem Code in der D3 Gallery zu finden, <https://observablehq.com/@d3/gallery>

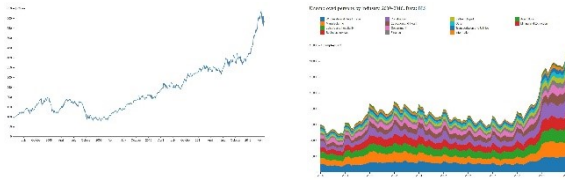


Abb. 5: Linien- und Flächendiagramm. Kontinuierliche Veränderungen

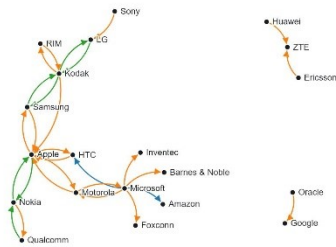


Abb. 6: Netzwerkdiagramm. Beziehungen und Verbindungen

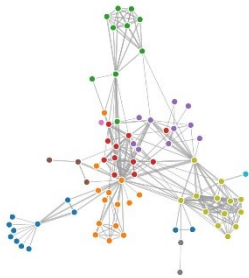


Abb. 7: Gerichteter Graph. Einfluss und Kräfte

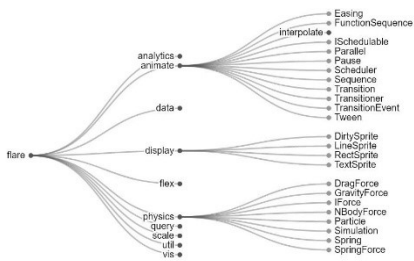


Abb. 8: Baumdiagramm. Hierarchische Beziehungen

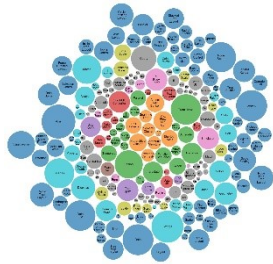


Abb. 9: Blasendiagramm. Relative Grössen und Werte

Weitere Visualisierungsmöglichkeiten

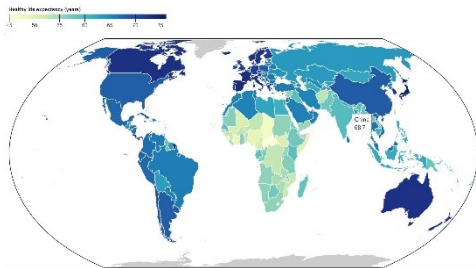


Abb. 10: Karten. Geografische Lokalisierung von Werten

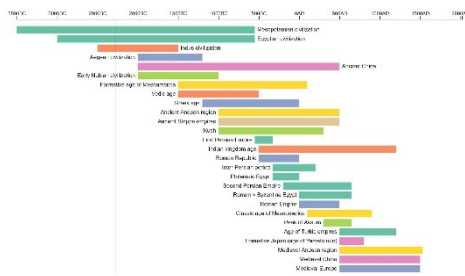


Abb. 11: Zeitstrahl. Zeitliche Werte



Abb. 12: Wordcloud. Worthäufigkeiten

Weiterführend bietet das *Financial Times Visual Vocabulary*¹⁶ eine hilfreiche Übersicht, welche Diagrammtypen zur Darstellung von Abweichung, Korrelation, Ranking, Verteilung, Veränderung über die Zeit, Menge, Teil des Ganzen, Raum oder einen Verlauf geeignet sind. Aufgrund der Vielgliedrigkeit kann diese Zusammenstellung hier nicht abgebildet werden, sondern wird am besten über die angegebene URL aufgerufen. Auch grafische Gestaltungsmittel wie Grösse, Helligkeitswert, Muster, Farbe, Richtung und Form können eingesetzt werden, um verschiedene Dimensionen, die in den Daten enthalten sind, wiederzugeben.¹⁷

Als Konsequenz der Vielfalt an Visualisierungsmöglichkeiten sehen wir bereits bei «standardisierten» Visualisierungen annähernd die Forderungen von Johanna Drucker nach einer auf geisteswissenschaftliche Sichtweisen ausgerichteten grafischen Wiedergabe erfüllt. Dies ist vor allem dann gegeben, wenn in solchen multiperspektivischen Systemen («point-of-view systems») die Repräsentation von Teilweisen, Massstabsverschiebungen, Mehrdeutigkeit, Unsicherheit und Betrachterabhängigkeit miteinbezogen oder wenigstens an geeigneter Stelle ausformuliert wird.¹⁸

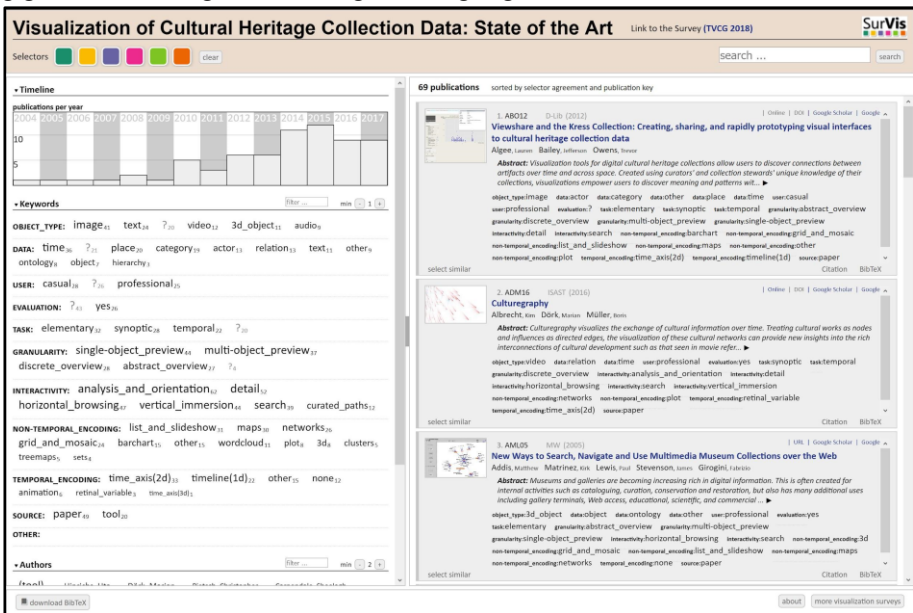


Abb. 13: Florian Windhager u. a., Visualization of Cultural Heritage Collection Data: State of the Art, <http://collectionvis.org/>

16 <https://ft.com/vocabulary>

17 Siehe dazu beispielhaft Bertin, Graphische Semiotik. Diagramme, Netze, Karten, S. 50.

18 Drucker, Graphical Approaches to the Digital Humanities, S. 248. Weiterführend siehe auch: Drucker, Visualization and Interpretation.

Florian Windhager u. a. haben in der Studie «Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges»¹⁹ eine umfassende Übersicht der Aspekte zusammengetragen, auf die es bei der Präsentation von kulturellen Sammlungen aus Bibliotheken, Archiven und Museen zu achten gilt. Ihr Augenmerk liegt dabei auf Interfaces, die mehr bieten als die bei digitalen Sammlungen weit verbreitete Stichwortsuche und der darauf folgenden Präsentation der Resultate im Kachellayout. Stattdessen ermöglichen Informationsvisualisierungen mit Sammlungsdaten neue Analyseformen und Zugänge sowohl für ein wissenschaftliches als auch ein breites Publikum. Dazu kommt, dass interaktive Visualisierungen es erlauben, komplexe und umfangreiche Informationsräume zu erkunden.²⁰ Die in der Studie berücksichtigten Beispiele von Informationsvisualisierungs-Systemen können auf <http://collectionvis.org/> selbst in einem interaktiven Browser erkundet und mit verschiedenen Schlagwörtern gefiltert werden. Die eingesetzte Webseiten-Struktur SurVis, die somit selbst eine Zugangsform darstellt, ist *open source* verfügbar.²¹

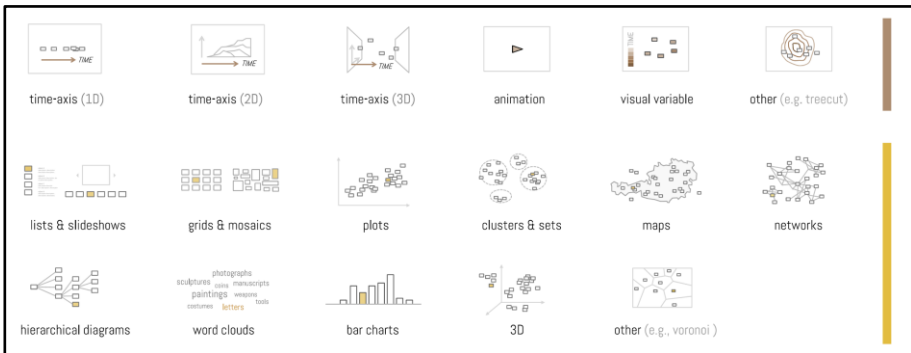


Abb. 14: Methoden zur Visualisierung zeitlicher Aspekte (oben) und Visualisierung nicht-zeitlicher Aspekte von digitalen kulturellen Sammlungen (Mitte und unten). Bildquelle: Windhager u. a. 2019, S. 2318.

Die digitale Repräsentation von Objekten oder Sammlungen kann über vier Ansichtstypen erfolgen, wie Windhager u. a. in ihrer Studie differenzieren: Vorschau eines einzelnen Objekts, Vorschau mehrerer Objekte, Sammlungsübersicht durch Verteilung der einzelnen Objekte anhand bestimmter Variablen (wie Grösse, Position, Farbe oder Form) und Sammlungsübersicht durch abstrakte Formen.²² Alle diese Formen werden in den unten folgenden Beispielen wiederzufinden sein und entweder einzeln vorkommen, in komplexeren Anwendungen kumuliert oder im Wechsel von einer Form in eine andere visuell Aussagen vermitteln. Auch bei den Aktionen, die

19 Windhager u. a., Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges.

20 Windhager u. a., Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges, S. 2311–2312.

21 Fabian Beck, SurVis – Visual Literature Browser, 2015, <https://github.com/fabian-beck/survis>

22 Windhager u. a., Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges, S. 2316.

Nutzende vornehmen können, lassen sich verschiedene Vorgehensweisen unterscheiden: Objektsuche, Überblick und Orientierung, Zoomen auf Abstraktionsebenen mit unterschiedlichem Detaillierungsgrad (*Vertical Immersion or Abstraction*), Zugang zu Objektangaben, Browsen nach verschiedenen Kategorien (*Horizontal Exploration*) und kuratierte Pfade.

In ihrer Analyse zeigen Windhager u. a. auf, dass die Visualisierung zeitabhängiger Daten innerhalb von sechs Kategorien erfolgen kann: Zeitstrahl (1D), Zeit als eine von zwei räumlichen Dimensionen (2D), Zeit als eine von drei räumlichen Dimensionen (3D), Animation, visuelle Variablen (Abbildung von Zeit durch Farbe, Grösse oder Textur) und andere Codierverfahren (beispielsweise Ringdiagramme). Weitere nicht-zeitliche Visualisierungsmethoden sind: Listen und Slideshows, Raster und Mosaik, Plots, Cluster und Sets, Karten, Netzwerke, hierarchische Diagramme, Wortwolken, Säulendiagramme, räumliche Visualisierungen und weitere Codierverfahren (beispielsweise Kreisdiagramme oder Liniendiagramme).²³

Wie bereits angemerkt, können einzelne Visualisierungsformen jeweils nur für eine Bandbreite an Datenpunkten optimale Aussagen machen. Dementsprechend eignen sich gewisse Visualisierungsverfahren für ausgesprochen viele Datenpunkte, während andere Verfahren nur wenige Datenpunkte sinnvoll abbilden können. Welche Verfahren zum Einsatz kommen, hängt von der Grösse der digitalen Sammlung und dem durch die Visualisierung verfolgten Ziel(e) ab.

3.1 Karten und geografische Visualisierungen

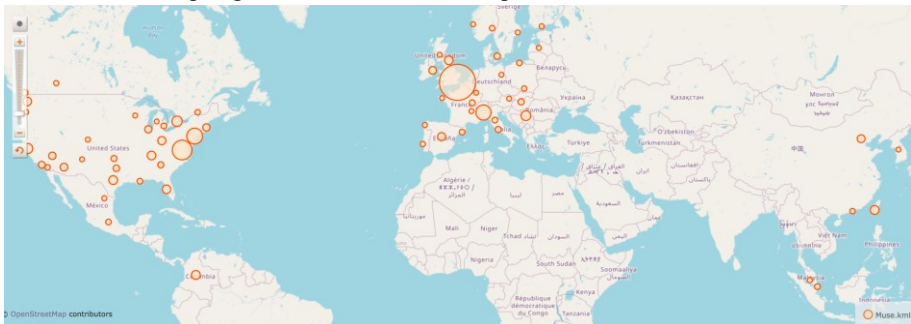


Abb. 15: Dariah Geobrowser, Visualisierung der Auftritte der britischen Rockband «Muse», die Grösse der Kreise korrespondiert der Anzahl Auftritte, Screenshot, <https://geobrowser.de.dariah.eu/>.

Die Kombination von Informationsvisualisierung und geografischen Karten ist bereits aus gedruckten Medien bekannt, wird in digitalen Anwendungen jedoch um Interaktivität erweitert. Heidmann etwa zeigt einerseits die historische Entwicklung der Kartografie bis zur interaktiven Karte auf und weist andererseits auf die politische

23 Windhager u. a., Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges, S. 2316–2319.

Entwicklung von Open Government Data hin, im Zuge derer Regierungen staatliche Datensammlungen allgemein und damit für vielfältige Einsatzbereiche verfügbar machen.²⁴ Als simples Beispiel lassen sich Daten, die mit sogenannten *geo tags* – Positionsangaben zur Verortung im geographischen Koordinatensystem – versehen sind, mit einem Geobrowser visualisieren.

Informationsvisualisierung in einer Karte kann in vielfältiger Weise erfolgen, wie eine weitere, statische Umsetzung zeigt. Durch das Nebeneinanderstellen desselben Kartenausschnitts, in den ein Ringdiagramm einbeschrieben ist, können bei Gregory Veränderungen der Kindersterblichkeit im viktorianischen Grossbritannien über Jahrzehnte hinweg nachvollzogen werden.²⁵ Dabei werden Aussagen zur Höhe der Kindersterblichkeit vorwiegend über die Farbgebung (rot = hoch, negativ; grün = niedrig, positiv) codiert, womit eine scheinbar intuitive Auswertung der Kartierung ermöglicht wird.²⁶ Grundsätzlich liesse sich eine solche Visualisierung auch interaktiv umsetzen, indem die Darstellung der Daten auf der Karte je nach gewähltem Zeitraum angepasst würde. Interaktive Karten kombinieren also unterschiedliche Visualisierungsformen, teilweise um die Möglichkeit erweitert, in die Dokumente hinein zu zoomen.²⁷

24 Heidmann, Interaktive Karten und Geovisualisierungen.

25 Gregory, *Different Places, Different Stories: Infant Mortality Decline in England and Wales, 1851–1911*.

26 Gregory, *Different Places, Different Stories: Infant Mortality Decline in England and Wales, 1851–1911*.

27 Zu Zooming siehe auch unten zu Plots (von Lev Manovich u. a.).

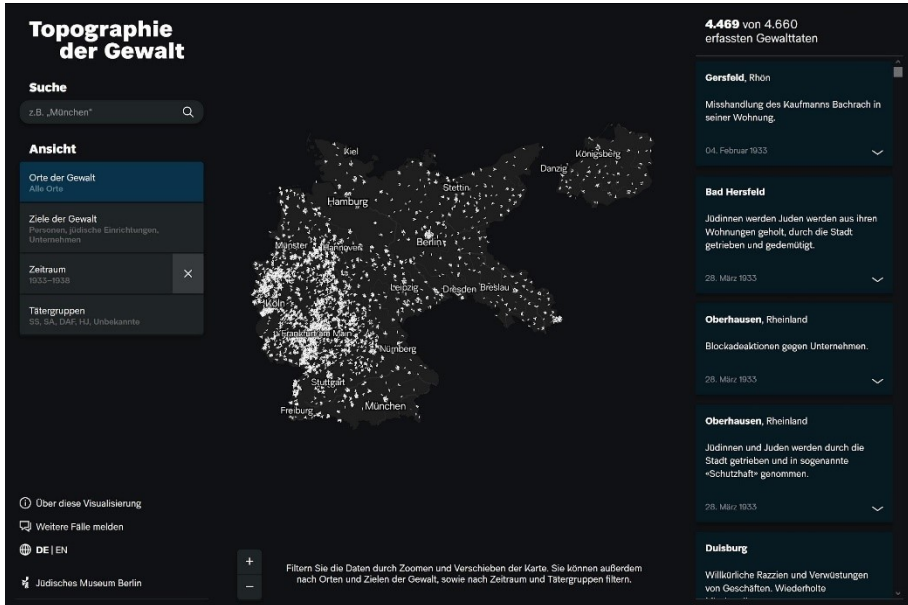


Abb. 16: Urban Complexity Lab, Fachhochschule Potsdam und Jüdisches Museum Berlin, *Topographie der Gewalt – Antisemitische Gewalttaten in Deutschland 1930–1938, 2019–2020*, <https://www.jmberlin.de/topographie-gewalt/>

Eine Vielzahl an Datenpunkten, die geografisch visualisiert werden, ist beispielsweise auch der Ansatz der Anwendung *Topographie der Gewalt – Antisemitische Gewalttaten in Deutschland 1930–1938*, die aus einer Projektkooperation zwischen dem Urban Complexity Lab (Fachhochschule Potsdam) und dem Jüdischen Museum Berlin hervorgegangen ist. In der Visualisierung werden Gewalttaten gegenüber Jüdinnen und Juden, jüdischen Einrichtungen, Geschäften und Unternehmen im Deutschen Reich aufgezeigt, die in den 1930er Jahren verübt wurden. Die vorhandenen Informationen der Gewalttaten werden auf einer Karte (wie oben abgebildet) oder in anderen Ansichten auch in einem Blasendiagramm, auf einem Zeitstrahl und in Ringdiagrammen visualisiert. Die verschiedenen Visualisierungsformen sind miteinander verknüpft, wodurch vorgenommene Eingrenzungen, z. B. des Zeitraums, sich auf die anderen Darstellungen auswirken. Zum bisher Beschriebenen kommt ein partizipativer Aspekt hinzu: Nutzerinnen und Nutzer der Anwendung können ihnen bekannte weitere Fälle unter Angaben von Quellen melden. Der dynamische Wechsel von einer Ansicht in eine andere ist ein Ansatz, der bezeichnend für die Visualisierungs-Projekte des Urban Complexity Labs ist, von denen in dieser Studie noch weitere besprochen werden.²⁸

28 Zur Umsetzung «Topographie der Gewalt» siehe <https://www.jmberlin.de/topographie-gewalt/>. Erarbeitet durch das Urban Complexity Lab, Fachhochschule Potsdam und Jüdisches Museum Berlin.

3.2 Netzwerk

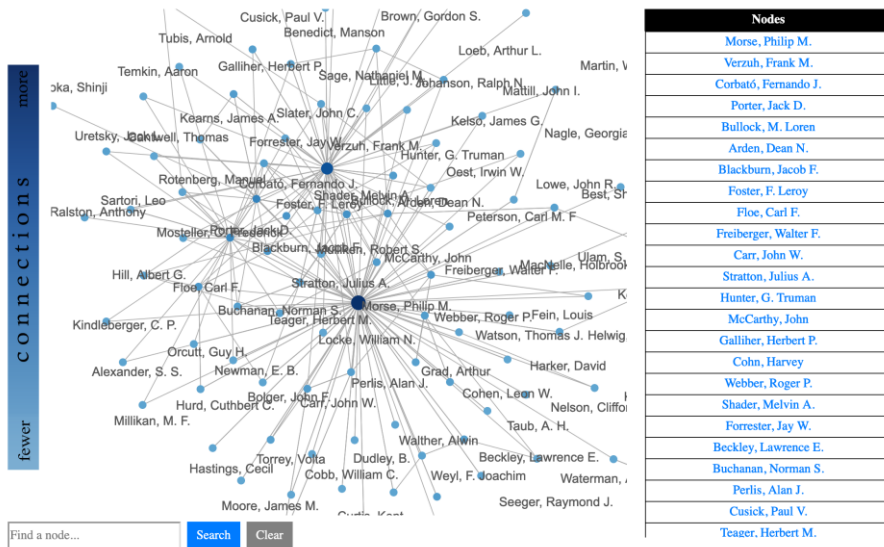


Abb. 17: Eine Netzwerkvisualisierung des Computation Archives am Massachusetts Institute for Technology (MIT), Screenshot, https://comphist.digitalhumanitiesmit.org/archives/net_viz

Ähnlich wie interaktive Karten gehören auch Netzwerke zu den weit verbreiteten Datenvisualisierungen. Insbesondere im Umkreis der Digital Humanities wurden bereits Stimmen laut, dass zu viele Netzwerkvisualisierungen produziert würden.²⁹ Dennoch kann die Visualisierung von Knoten und Kanten (*Nodes* und *Edges*) über Graphen (gerichtet oder ungerichtet) zur Identifikation von Beziehungen hilfreich sein. Aufgrund der Einführung von Graphen anstelle von relationalen Datenbanken zur Formulierung von Aussagen, etwa im Zusammenhang mit Linked Open Data,³⁰ ist der Einsatz von Netzwerkvisualisierungen ausgesprochen verlockend geworden. Insbesondere interaktive Ansätze, die Graphen an einem bestimmten Punkt (einer Person/einem Ort) anzusetzen und von dort aus Beziehungen aufzuzeigen, sind von hohem heuristischem Wert, da im Beziehungsnetz über mehrere Kanten und Knoten einer Spur gefolgt werden kann. Unter Diskussion der Datenmodellierung hat dies Andreas Kuczera am Beispiel der Regesta Imperii aufgezeigt.³¹

29 Siehe dazu und einführend Weingart, *Demystifying Networks, Parts I & II*.

30 Siehe dazu unten Record in Context (Linked Open Data Modell für Archive).

31 Kuczera, *Digitale Perspektiven mediävistischer Quellenrecherche*. Siehe auch Kuczera, *Graphdatenbanken*.

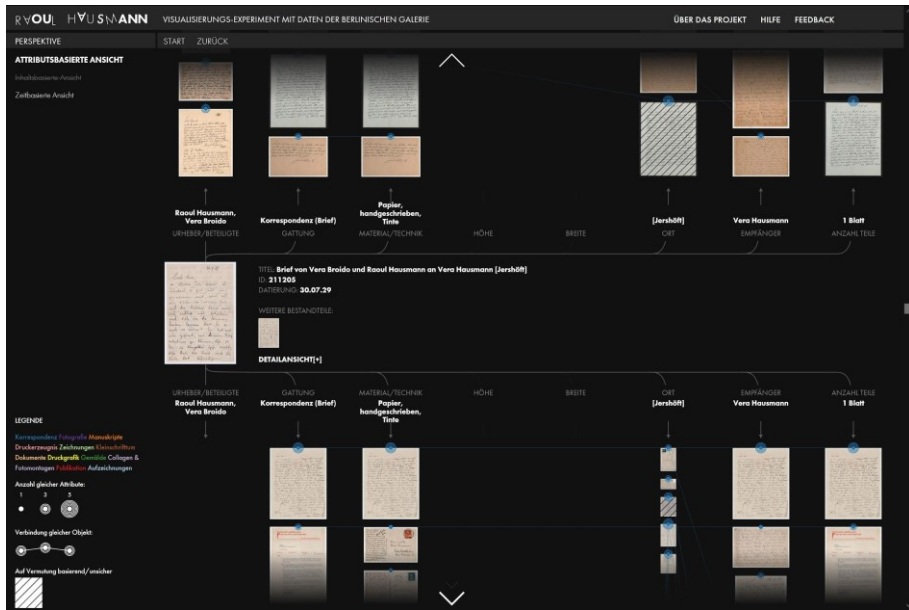


Abb. 18: Mark-Jan Bludau, Raoul Hausmann Sammlung – Ein Visualisierungs-Experiment mit Daten der Berlinischen Galerie, 2017, <https://uclab.fh-potsdam.de/hausmann/>

Mark-Jan Bludau, Marian Dörk und Frank Heidmann stellen im Zusammenhang mit der experimentellen Visualisierung *Raoul Hausmann* ein Konzept vor, gemäß dem mittels mehrerer Visualisierungen relationale Perspektiven von einzelnen Artefakten aus kulturellen Sammlungen aufgezeigt werden sollen. An der gleichzeitigen Darstellung der Komplexität und Einzigartigkeit von Artefakten scheitern viele gängige Visualisierungstypen.³² Die Visualisierung *Raoul Hausmann* ist als Masterarbeit von Mark-Jan Bludau an der Fachhochschule Potsdam entstanden. Visualisiert werden Dokumente aus dem Nachlass des Berliner Dadaisten Raoul Hausmann (1886–1971), der in der Berlinischen Galerie aufbewahrt wird und aus Gemälden, Zeichnungen, Fotografien, Collagen, Fotomontagen, Manuskripten, persönlichen Notizen, Zeitungsausschnitten und Briefen besteht. In der Visualisierung geben nach Gattungen aufgeteilte Raster zunächst eine grobe Übersicht über die ganze Sammlung, während die Balken am linken Seitenrand die chronologische Verteilung der Dokumente auf einem Zeitstrahl aufzeigen. Am rechten Seitenrand werden gleichzeitig vorkommende Personen und Körperschaften aufgelistet und ihre Beziehungen in einem Bogendiagramm dargestellt. Sowohl die Zeitleiste als auch die Personenliste dienen als Filter für die Miniaturvorschauen der Dokumente. Sobald ein Dokument angeklickt wird, wechselt das Interface zu einer Ansicht, in der – wie in der Abbildung oben – vielfältige Beziehungen zwischen verschiedenen Dokumenten hergestellt

32 Heidmann/Bludau/Dörk, Relational Perspectives as Situated Visualizations of Art Collections.

werden. Jedes sichtbare Dokument kann zum neuen Mittelpunkt werden, wodurch sich das Interface dynamisch anpasst und Nutzende interessengesteuert durch die Sammlung navigieren können, während durch die visualisierten Beziehungen immer ein Kontext zwischen verschiedenen Dokumenten hergestellt wird. Beim ausgewählten Dokument kann zudem ein Fenster mit einem grösseren Bild sowie den Objektinformationen und der Transkription aufgerufen werden. In diesem Fenster ist auch der Link zur Online-Sammlung der Berlinischen Galerie und damit des von der Institution gepflegten Objekteintrags angegeben. Mehrseitige Dokumente werden sowohl im Fenster als auch in der Visualisierung stets in ihrem Zusammenhang angezeigt. Bei diesem Visualisierungsbeispiel wird also das Netzwerk aus relationalen Daten hergestellt, dessen Knoten von einem Vorschaubild dargestellt und die Verbindungen dazwischen aus den Metadaten generiert werden. Damit wird ein Einblick in einen eher übersichtlichen, thematisch zusammenhängenden digitalisierten Bestand möglich, der über vielseitige Kontexte explorativ erkundet werden kann.

3.3 Zeitstrahl

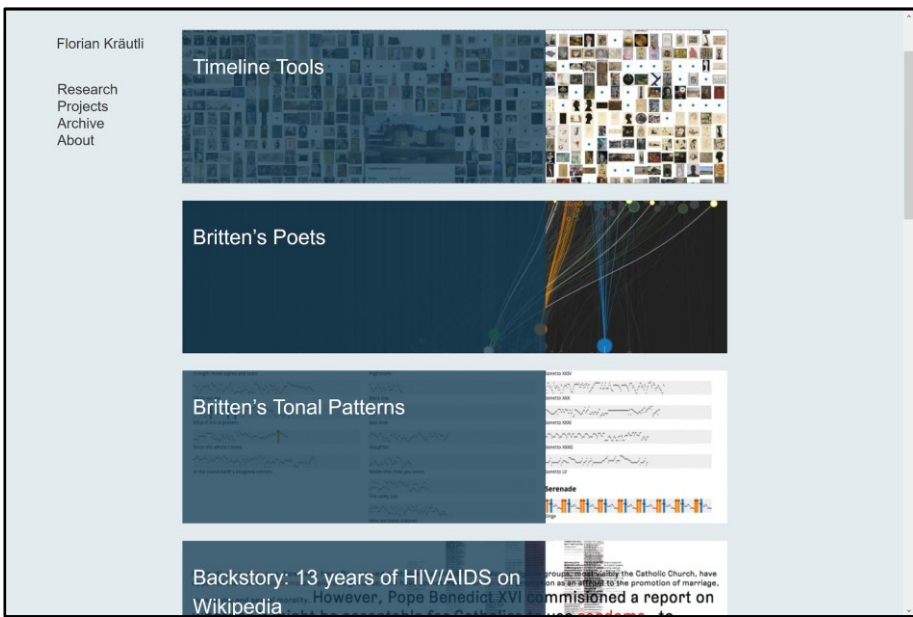


Abb. 19: Dokumentation der Visualisierungsprojekte von Florian Kräutli: <http://www.kraeutli.com/index.php/category/projects/>

Historische Dokumente und Artefakte sind häufig mit einem bestimmten Datum oder Zeitraum verbunden. Bei einem Fokus auf die zeitliche Dimension, die in kulturellen Daten enthalten ist, bietet sich als Darstellungsform ein Zugang über einen Zeitstrahl an. In der 2016 erschienenen Dissertation *Visualising Cultural Data:*

*Exploring Digital Collections Through Timeline Visualisations*³³ hat Florian Kräutli untersucht, wie zeitbasierte Formen der Datenvisualisierung ermöglichen, in digitalen Sammlungen Wissen zu entdecken. Interaktive Zeitstrahle gehen über ein lineares Storytelling hinaus und werden zu Werkzeugen für visuelle Analysen. Ausgehend von den in digitalen Sammlungen erfassten Datierungen hat Kräutli die Auswirkung kuratorischer Entscheidungen auf die Sammlungsdaten untersucht und auch gezeigt, wie diese Entscheidungen in einem Zeitstrahl sichtbar gemacht werden können. In enger Zusammenarbeit mit Kulturinstitutionen und den zuständigen Kuratorinnen und Kuratoren hat er funktionsfähige Prototypen erstellt, um aufzuzeigen, wie umfangreiche digitale Sammlungen visualisiert und das darin enthaltene Wissen zugänglich und nutzbar gemacht werden kann. Die Repräsentation von ungenauen, beschreibenden Zeitangaben, wie sie in digitalen Sammlungen oft vorkommen, haben in diesen Prototypen einen besonderen Stellenwert erhalten.³⁴ Kräutli betont die Wichtigkeit des frühen Einbezugs von auf Datenvisualisierung spezialisierten Designerinnen und Designern in Digital Humanities Projekte: Idealerweise sind sie von Anfang an dabei und gestalten sowohl die Konzeption als auch die spätere Umsetzung der Projekte mit. Nur in der interdisziplinären Zusammenarbeit von Designer*innen mit Geisteswissenschaftler*innen (womit sowohl Digital Humanists als auch Fachwissenschaftler*innen gemeint sind) können die Methoden der jeweiligen Disziplin wechselseitig fruchtbar eingesetzt werden. So ist der Beitrag, der vonseiten des Designs zur visuellen Repräsentation von Wissen geleistet werden kann, nicht zu unterschätzen.³⁵

33 Kräutli, *Visualising Cultural Data: Exploring Digital Collections Through Timeline Visualisations*.

34 Kräutli, *Visualising Cultural Data: Exploring Digital Collections Through Timeline Visualisations*, S. 4–5.

35 Kräutli, *Visualising Cultural Data: Exploring Digital Collections Through Timeline Visualisations*, S. 246.

The screenshot displays the Olivia Vane website interface, which features a navigation menu at the top with links for PROJECTS, WRITING, SPEAKING, PHD, and ABOUT. The main content area is divided into several sections, each showcasing a different digital visualization tool:

- Press Picker:** A tool for making sense of the British Library's newspapers from 2020, featuring a timeline from 1700 to 2020 and a list of newspaper titles with corresponding red waveform visualizations.
- Faces of Sweden:** A tool for exploring portraits in the Nordic Museum from 1700 to 1850, displaying a grid of portrait images.
- Visualising the V&A Museum collection:** A tool for visualizing patterns across historical photographs from 2018, showing a dense, colorful grid of small image thumbnails.
- Dive into Color:** A tool for exploring color and color harmonies in the Smithsonian Design Museum collection from 2018, featuring a hand interacting with a large digital screen displaying a color palette and image thumbnails.
- Waterfall:** A tool for visualizing WWI portraits from the Imperial War Museum, London, from 2018, showing a vertical waterfall of portrait images.
- Visualising the Smithsonian Design Museum collection:** A tool for visualizing timelines and tags from 2017, displaying a grid of image thumbnails with associated tags.
- Stepstext:** A text visualization tool for exploring digitised historical documents from 2016, showing a vertical timeline of text snippets from 1800 to 1900.

Abb. 20: Vane, Timeline Design for Visualising Cultural Heritage Data.

Olivia Vane hat sich in ihrer 2019 veröffentlichten Dissertation *Timeline Design for Visualising Cultural Heritage Data*³⁶ ebenfalls mit der Visualisierung von Daten aus digitalisierten Sammlungen von Museen, Archiven und Bibliotheken in interaktiven Zeitstrahlen befasst. Während Florian Kräutli den Fokus auf den Zeitstrahl als Tool zur Analyse von digitalen Sammlungen gelegt hat, steht bei Olivia Vane der Zeitstrahl als Übermittler von Narration im Vordergrund. Sie hat den Schwerpunkt auf das Konzept von Narration in und durch Zeitstrahle gelegt sowie auf die Darstellung von Unsicherheit bei der Visualisierung von Zeit.

Durch Digitalisierung allein werden kulturelle Sammlungen weder automatisch zugänglich, auffindbar und verständlich, noch unterstützen Standard-Interfaces notwendigerweise die Interaktionen, die Nutzerinnen und Nutzer vornehmen möchten. Visualisierungen ermöglichen sowohl Fachleuten als auch einem breiten Publikum Zugang zu digitalen Abbildungen aus Sammlungen, um diese zu analysieren und zu erkunden. Interaktive Datenvisualisierungen geben einen Überblick über die Inhalte und Breite von Sammlungen und ermöglichen das Erkennen von Mustern.³⁷ Wie unten aufgelistet, sind dazu verschiedene Formen von Zeitstrahlen erkennbar, die das Gewicht bei der Visualisierung auf unterschiedliche Schwerpunkte legen und verschiedene Grade des Kuratierens von Inhalten oder automatisierter Darstellung aufweisen³⁸:

1. Ausgeprägt kuratiertes Narrativ (multimediale Zeitstrahle, die weitgehend manuell konzeptioniert worden sind). Beispiele:
 - Alphonse Mucha Timeline: <http://www.muchafoundation.org/en/timeline>
 - Bud's Story: <https://projects.propublica.org/graphics/mia-letters>
 - Met's Timeline of Art History: <https://www.metmuseum.org/toah/chronology/>

37 Vane, *Timeline Design for Visualising Cultural Heritage Data*, S. 4.

38 Die Kategorisierung und die Auswahl der Beispiele folgt der Zusammenstellung von Olivia Vane: Vane, *Timeline Design for Visualising Cultural Heritage Data*, S. 25–34.

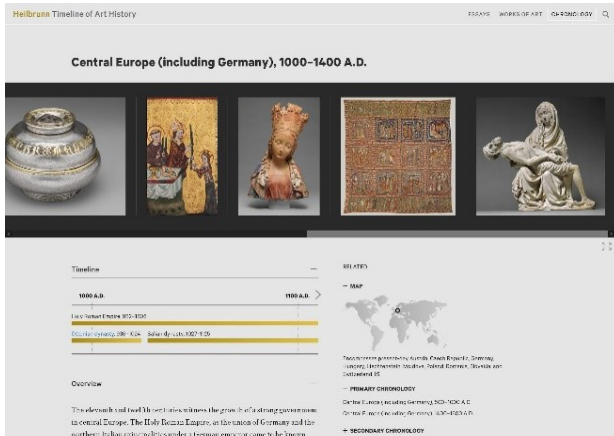


Abb. 21: Met's Timeline of Art History

2. Kuratierte Gruppen (eine kuratierte Auswahl von Objekten mit einer bestimmten Eigenschaft oder zu einem bestimmten Thema). Beispiel:

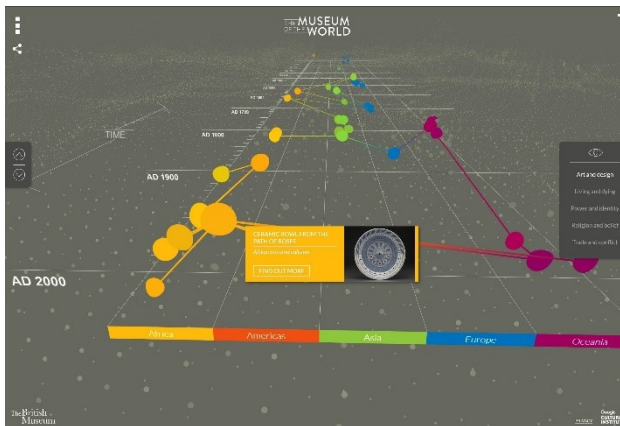


Abb. 22: British Museum's The Museum of the World: <https://britishmuseum.withgoogle.com/>

3. Quantitative Darstellung (Diagramme für eine quantitative Übersicht im Verlauf der Zeit, oft verbunden mit Filtermöglichkeiten und Annotationen). Beispiele:
- Darwin's letters: a timeline: <https://www.darwinproject.ac.uk/letters/darwins-letters-timeline>
 - Ashmolean Museum's Pilgrimage: http://jameelcentre.ashmolean.org/collection/6/10239/10402/all/sort_by/seqn.
 - Deutsche Digitale Bibliothek visualisiert: <https://uclab.fh-potsdam.de/ddb/epochen/>
 - Tate Explorer: <https://gravitron.com.au/tate/tate.html>



Abb. 23 : Deutsche Digitale Bibliothek visualisiert

4. Chronologisches Raster (Aneinanderreihung von kleinen Bild- oder Dokumentvorschauen in einem Raster nach gewissen Kriterien wie z. B. Helligkeit).
Beispiele:

- 4535 Time Magazine Covers, 1923–2009: <http://lab.culturalanalytics.info/2016/04/timeline-4535-time-magazine-covers-1923.html>
NYPL Public Domain Visualization: <https://publicdomain.nypl.org/pd-visualization/>
- Below the Surface: <https://belowthesurface.amsterdam/en/vondsten>,

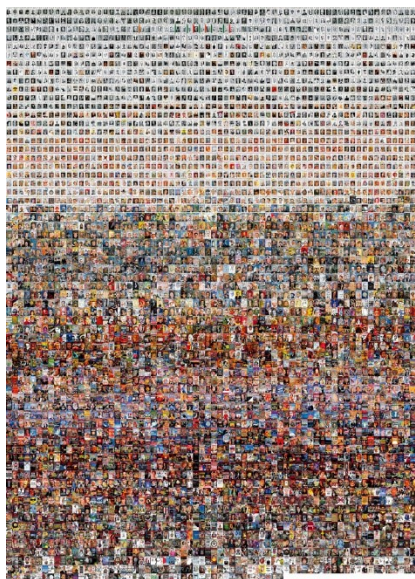


Abb. 24: 4535 Time Magazine Covers, 1923–2009

5. Zuordnen der einzelnen Objekte zu bestimmten Zeitpunkten (verschiedene Abstraktionsgrade, oft jedoch besteht der Datenpunkt aus den Vorschaubildern). Beispiele:

- VIKUS Viewer: <https://vikusviewer.fh-potsdam.de/fw4/>
- Forgotten Heritage: <https://www.forgottenheritage.eu>
- Coins – A journey through a rich cultural collection: <https://uclab.fh-potsdam.de/coins/>

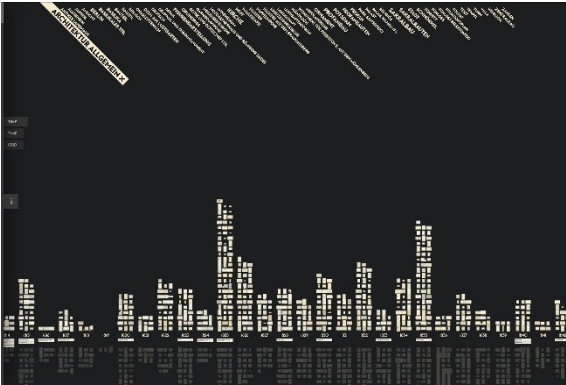


Abb. 25 : VIKUS Viewer

6. Plots (Darstellung der digitalen Bilder aufgrund von quantitativ messbaren Werten in einem Koordinatensystem, wodurch die Verteilung der Objekte sichtbar wird). Beispiel:



Abb. 26: 4535 Time Magazine Covers, 1923–2009. <http://lab.culturalanalytics.info/2016/04/timeline-4535-time-magazine-covers-1923.html>

7. Objekt-zu-Objekt-Verbindungen (bei der Auswahl eines Objekts werden Verbindungen zu anderen Objekten aufgezeigt). Beispiel:

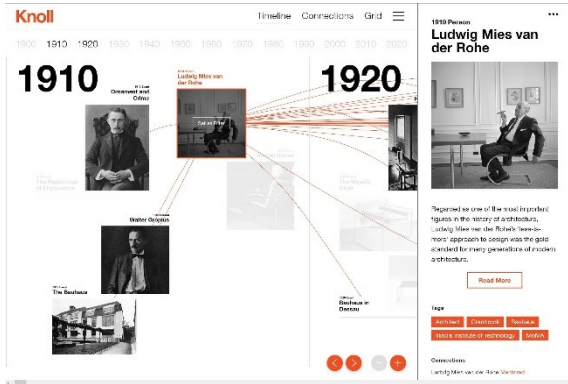


Abb. 27: Zeitstrahl Knoll-Archiv: <https://www.knoll.com/the-archive/>

8. Sampling (Darstellung einer repräsentativen Objektauswahl für einen bestimmten Zeitpunkt zur Komplexitätsreduktion). Beispiele:

- DigitaltMuseum: <https://digitaltmuseum.no/>
- Timeline of Modern Art: <https://www.framestore.com/timelineofmodern-art?language=en>



Abb. 28: DigitaltMuseum

Die Sichtbarkeit von Bildern in den Visualisierungen ermöglicht einen direkten qualitativen Zugang zu den Objekten in der digitalen Sammlung und deren umgehenden Vergleich. In den Daten enthaltene Kategorien können zur Gruppierung und zum Filtern der Objekte verwendet werden, numerische Werte ermöglichen die Visualisierung bestimmter Eigenschaften in einem Koordinatensystem oder das Skalieren von Datenpunkten.³⁹

39 Vane, *Timeline Design for Visualising Cultural Heritage Data*, S. 34.



Abb. 29: Urban Complexity Lab, Fachhochschule Potsdam, Deutsche Digitale Bibliothek visualisiert, 2014, <https://uclab.fh-potsdam.de/ddb/>

Die *Deutsche Digitale Bibliothek visualisiert* setzt auf mehrere parallele Zeitstrahlen, um eine Übersicht über grössere Bestände zu geben. Als eines der frühen Projekte, die am Urban Complexity Lab entstanden sind, bietet diese interaktive Visualisierung ein alternatives Interface als Zugang zu den Beständen der Deutschen Digitalen Bibliothek (DDB)⁴⁰. Die DDB verfolgt als zentrales nationales Portal das Ziel, die digitalen Angebote von deutschen Kultur- und Wissenschaftseinrichtungen miteinander zu vernetzen und so einen freien digitalen Zugang zu Büchern, Archivalien, Bildern, Skulpturen, Musikstücken, Tondokumenten, Filmen und Noten zu ermöglichen. In der Visualisierung steht jeder Zeitstrahl für eine Sparte (Archiv, Bibliothek, Mediathek, Forschung, Museum, Denkmalpflege und Sonstige) und ermöglicht, auf einen Blick zu erkennen, wovon es zu welchem Zeitpunkt eine wie grosse Menge an Dokumenten in der DDB gibt. Eine Zeitperiode, die von besonderem Interesse ist, kann markiert werden, wodurch sich sowohl die darunter angeordneten Word Clouds der Auswahl anpassen, als auch die weissen Säulen, die die Medien anzeigen (Ohne, Bild, Text, Audio, Video) sowie die bunten Balken daneben, die für die Anzahl der Einträge stehen. Falls diese verschiedenen Darstellungen, aus denen sich die Visualisierung zusammensetzt, zu wenig selbsterklärend sind, können über den Info-Button Labels eingeblendet werden, die in aller Kürze die verschiedenen Elemente erläutern. Alle um die Zeitstrahlen angeordneten Visualisierungselemente, d.h. die

40 <https://www.deutsche-digitale-bibliothek.de/>

Stichwörter in den Word Clouds, die weissen Säulen und die Bunten Balken führen mit einem Klick zum klassisch aufgebauten Suchportal der Deutschen Digitalen Bibliothek. Dort werden direkt diejenigen Datensätze als Suchresultat angezeigt, die der Auswahl entsprechen, die zuvor mit den vorhandenen Filtermöglichkeiten in der Visualisierung vorgenommenen wurde. Der Zeitstrahl dient in dieser Visualisierung als Indiz, um den Medienbestand in seiner Quantifizierung darzustellen. Ziel ist, wie Marian Dörk, Christopher Pietsch und Gabriel Credico festhalten, unterschiedliche Blickwinkel zu eröffnen, weil eine einzige Visualisierung die Komplexität einer kulturellen Sammlung nicht wiedergeben kann.⁴¹ Die Visualisierungen sollen nach Kathrin Glinka, Sebastian Meier und Marian Dörk auch genutzt werden, um nicht direkt sichtbare Zusammenhänge darzustellen und konfligierende Narrative zu verdeutlichen.⁴²

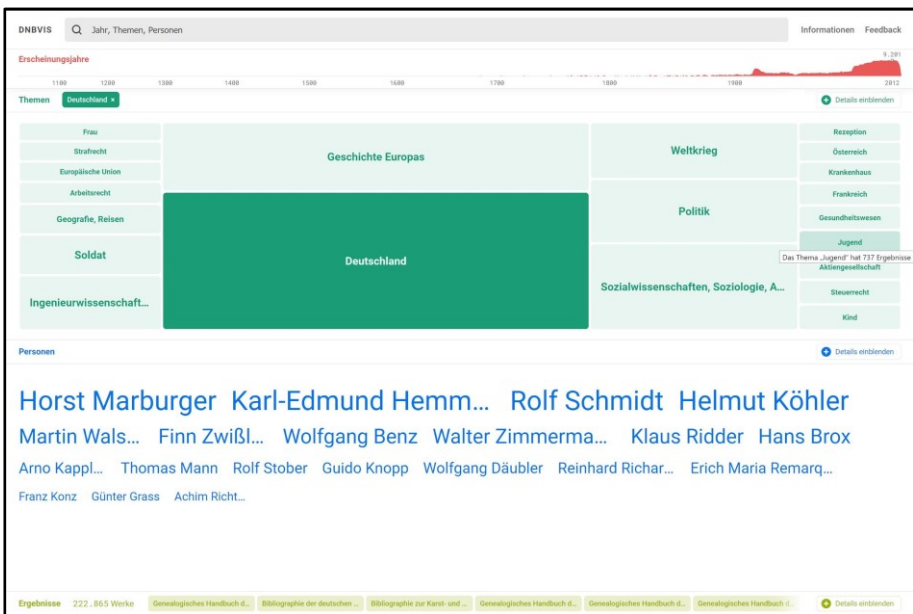


Abb. 30: Urban Complexity Lab, Fachhochschule Potsdam in Kooperation mit der Deutschen Nationalbibliothek, DNBVIS – Visualisierung bibliografischer Daten und Inhalte, 2017, <https://dnbvis.fh-potsdam.de/>

Aufbauend auf die Timelines wurden TreeMaps sowie Wordclouds für die Bestände der Deutschen Nationalbibliothek (DNB) erstellt. Sie visualisieren zusätzlich zu den chronologischen Daten weitere Informationen wie Schlagwörter und stellen somit zusätzliche Zugangsformen bereit. Diese Informationen werden logarithmisch

41 Dörk/Pietsch/Credico, One View is Not Enough: High-Level Visualizations of a Large Cultural Collection.

42 Glinka/Meier/Dörk, Visualising the »Un-seen«: Towards Critical Approaches and Strategies of Inclusion in Digital Cultural Heritage Interfaces.

(Zeitstrahl), als Kacheldiagramm (Themen) sowie als Formwolke (Personen) dargestellt und sind interaktiv. Eingrenzungen in einem Bereich verändern wiederum die Auswahl eines anderen Bereichs und beeinflussen schlussendlich das Suchergebnis.

Wie Brüggemann, Dittrich und Dörk nach Fokusgruppenbefragungen feststellen, sind diese Zugänge für explorative erste Erkundungen hilfreich, jedoch nicht für zielgerichtete Suchen. Das Designziel, nämlich die Integration von Sichtung und Suche wird zwar angepeilt, verfehlt aber wohl seine letztliche Intention, Suchvorgänge zu erleichtern. Gleichzeitig werden aber das Auffinden zufälliger Befunde und Einsichten aus quantifizierenden Ansätzen befördert.⁴³

3.4 Plots

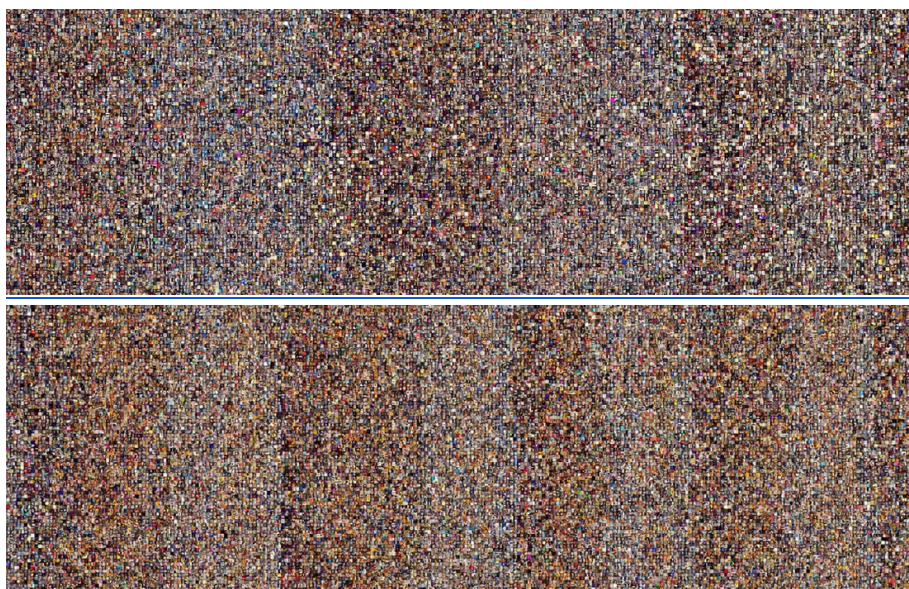


Abb. 31: Cultural Analytics Lab, City University of New York, Phototrails, 2013. Zwei Raster im Vergleich zur Visualisierung von Instagram-Fotos, die über vier aufeinanderfolgende 24-Stunden-Zeiträume in New York City (oben) und Tokio (unten) geteilt wurden. Die Fotos sind nach Veröffentlichungsdatum (von oben nach unten) und -zeit (von links nach rechts) sortiert, <http://phototrails.info/>. Bildquelle und hochauflösende Bilder: <http://phototrails.info/visualizations/montage-visualization/>

Lev Manovich hat das Cultural Analytics Lab (bei der Gründung 2007 noch unter der Bezeichnung Software Studies Initiative an der University of California, San Diego) aus dem Interesse heraus ins Leben gerufen, grosse Mengen an Bildern, Videos und interaktiven visuellen Medien – wie sie täglich zu tausenden im Web in Blogs, auf Flickr, Instagram, YouTube, Facebook, Twitter entstehen – zu

43 Brüggemann u. a., Die bibliografischen Daten der Deutschen Nationalbibliothek entfalten.

untersuchen.⁴⁴ Mittels automatischer Analyse werden aus den digitalen Bildern verschiedene numerische Werte und visuelle Charakteristika extrahiert, die dazu dienen, sämtliche Bilder aus einem Datensatz anhand dieser Merkmale zu visualisieren. Die dafür notwendigen Methoden und Techniken bezeichnet Lev Manovich als Cultural Analytics, deren zugrundeliegende Idee die Nutzung von Visualisierungen ist, um grosse Mengen von digitalen Bildern und Videos zu erkunden.⁴⁵

Zur Erforschung grosser Bildersammlungen arbeitet Lev Manovich mit Plots, die einzelne Bilder nach bestimmten Schemata anordnen. Dadurch werden Strukturen sicht- und fassbar, die kulturelle Aussagen über ganze Bildkorpora erlauben. Bei *Phototrails*, an diesem Projekt waren zudem Nadav Hochman und Jay Chow beteiligt, wurden Fotografien von Instagram verwendet, die in New York und in Tokio hochgeladen wurden. Die Fotografien sind in der Reihenfolge des Uploads (oben gegen unten) und nach Veröffentlichungszeit (links nach rechts) geordnet, so dass die zu unterschiedlichen Tageszeiten hochgeladenen Fotos als Bänder sichtbar werden.⁴⁶ Im Vergleich der beiden Plots wird deutlich, dass der Farbton der beiden Städte leicht voneinander abweicht und eine unterschiedliche Frequenz von ungleich breiten Bändern auftritt. Welche sozialen Bedingungen hinter dem beobachteten Effekt stecken, muss bei näherer Betrachtung der einzelnen Fotos untersucht werden. Eine Rolle spielt sicherlich das aufgenommene Objekt und ob die Aufnahmen in geschlossenen Räumen, unter freiem Himmel, bei Tag oder bei Nacht gemacht wurden. Davon ausgehend lassen sich Aussagen darüber treffen, in welchen Städten beispielsweise das Nachtleben besonders ausgeprägt ist und zudem dokumentiert wird, oder wo eher andere Objekte und Themen ins Gewicht fallen, wie beispielsweise die Abbildung von Speisen, Landschaften oder Personen (auch Selfies).

Ordnungen lassen sich mit Bilddaten beliebiger Objekte in unterschiedlichsten Arrangements erstellen. Die Anordnung in einem Raster wie im obigen Beispiel von Lev Manovich und damit das Aufspannen eines Bilderteppichs zwischen einer unsichtbaren X- und Y-Achse ist nur eine Möglichkeit. Eine andere Möglichkeit ist, aufgrund bestimmter Kategorien in den Metadaten, Cluster zu bilden.

44 Manovich, *How to Compare One Million Images?*, S. 250–251.

45 Manovich, *How to Compare One Million Images?*, S. 262.

46 Hochman/Manovich, *Zooming into an Instagram City: Reading the Local through Social Media*.

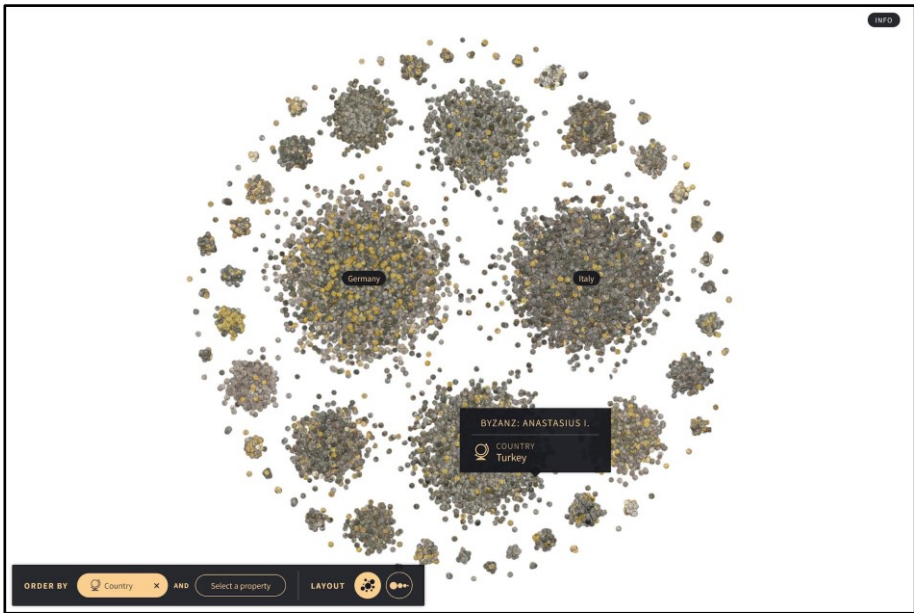


Abb. 32: Urban Complexity Lab, Universität Potsdam, Coins – A journey through a rich cultural collection, 2017–2018, <https://uclab.fh-potsdam.de/coins/>

Besonders gelungen und mehrfach prämiert ist die Visualisierung historischer Münzen, *Coins*, die am Urban Complexity Lab der Fachhochschule Potsdam in Kollaboration mit dem Münzkabinett Berlin entstanden ist: Rund 26000 Münzen aus dem Münzkabinett können mit verschiedenen Layouts und Filtern arrangiert und interaktiv erkundet werden. Der Gesamtbestand der Münzen kann so beispielsweise in nach Ländern sortierte, kleinere Haufen aufgeteilt werden. Auf diese Weise werden konventionelle Displays der Präsentation von Münzen im Museumskontext aufgebrochen. In einer Ausstellung im Museum steht üblicherweise die einzelne Münze, ihre Beschaffenheit und das darauf abgebildete Motiv im Vordergrund. Daneben bestehen digitale Sammlungspräsenzen in den meisten Fällen aus einer durchsuchbaren Datenbank, die aus einer Abbildung des Objekts und Objektinformationen (Bezeichnung/Titel, Datierung, Material, Masse usw.) bestehen. Die einzelnen Münzen in *Coins* sind ebenfalls mit zusätzlichen Informationen (wie Land, Material, Durchmesser, Gewicht oder Datierung) zur näheren Bestimmung verknüpft. Das sind jedoch ausgewählte Kategorien, die vor allem deshalb vorhanden sind, weil sie auch zum Durchstöbern, Arrangieren und Filtern der Sammlung verwendet werden. Bei Interesse an ausführlichen Informationen zu einer einzelnen Münze wird man jedoch direkt aus der Anwendung heraus zum entsprechenden Eintrag im *Interaktiven Katalog des Münzkabinetts*⁴⁷ weitergeleitet. Die 2018 mit dem *Information is Beautiful Award*

47 <https://ikmk.smb.museum/home>

sowie dem *Digital Humanities Award* ausgezeichnete Anwendung bietet also ein Interface, das Präsentationszwecken gleichermassen wie dem neugierigen Aufspüren von Zufallsfunden dient. In spielerischer Interaktion wird man an eine vermutlich für viele eher trockene Materie herangeführt, indem wissenswerte Informationen weitergegeben und bei den Nutzenden ein gewisses Interesse für diese Sammlung geweckt wird. Erklärtes Ziel von Flavio Gortana u. a. war es denn auch, in Abkehr von einem datenbanktypischen Kachellayout Visualisierungen zu entwickeln, die der materiellen und semantischen Reichhaltigkeit der ganzen Sammlung gerecht werden und auch Personen einen Zugang bieten, die kein numismatisches Vorwissen mitbringen.⁴⁸ Es ist vorwiegend die dynamische Anordnung bei der gleichzeitigen Möglichkeit, die zunächst eher als Datenpunkte visualisierten Objekte mittels Zoom zur genaueren Inspektion heranzuholen, die den Reiz der Visualisierung ausmacht, wie von Flavio Gortana u. a. beschrieben. Die iterativ, im engen Austausch mit Numismatikerinnen und Numismatikern und mit echten Münzen spielerisch vorgenommene Designfindung hat dazu geführt, dass ein Zugriff auf die digital repräsentierten Münzen möglich ist, wie er in der Sammlung oder im Museum nicht erlaubt wäre.⁴⁹ An diesem Beispiel wird besonders evident, dass nicht zuletzt auch der Inhalt und die Charakteristiken der Daten zu einer bestimmten visuellen Umsetzung führen, die jedoch für Daten mit einer Herkunft aus einem ganz anderen Kontext völlig ungeeignet sein mag.

Die beiden letzten besprochenen Beispiele legen nahe, dass im Umgang mit Daten die Fernsicht für einen umfassenden Blick auf eine Sammlung eine besondere Stellung einnimmt, ohne dabei den Fokus auf das einzelne Objekt zu verlieren. Zwischen diesen beiden Polen liegt eine Bewegung, die entweder durch die vergrösserte Darstellung eines Bildes zustande kommt oder indem für eine kleinere Objektgruppe oder ein Einzelobjekt detailreichere Informationen eingeblendet werden. Das Konzept des Zooms von einer Sicht aus der Ferne hin zur einzelnen Fotografie ist bereits in den zweidimensionalen Plots von Lev Manovich sichtbar geworden, tritt aber noch häufiger in späteren Arbeiten auf, in denen die unterschiedlichen Ebenen oder Schichten noch stärker verschränkt und teilweise direkt sichtbar gemacht werden (siehe dazu die Legende zu *On Broadway*).

48 Gortana u. a., Off the Grid: Visualizing a Numismatic Collection as Dynamic Piles and Streams.

49 Gortana u. a., Off the Grid: Visualizing a Numismatic Collection as Dynamic Piles and Streams.

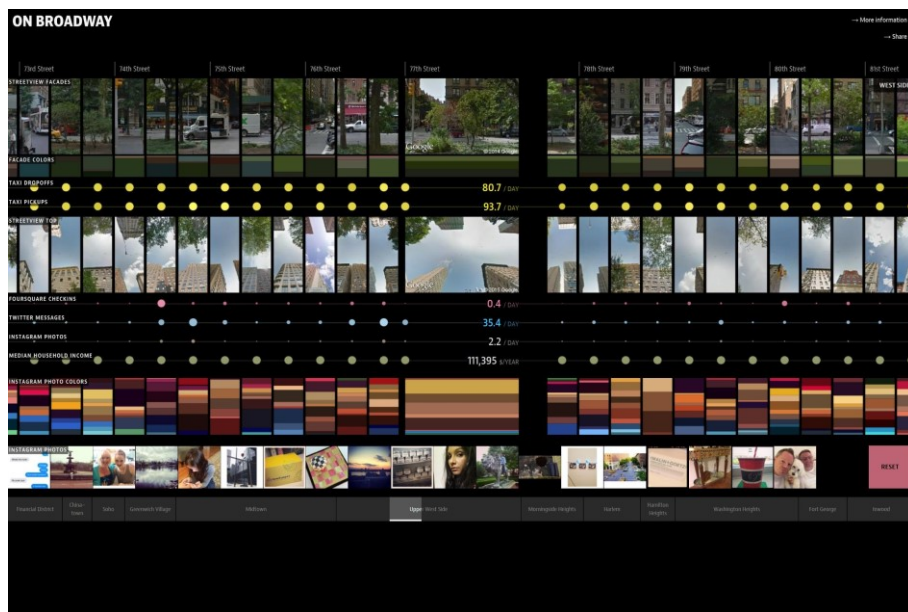


Abb. 33: Cultural Analytics Lab, City University of New York, On Broadway, 2014–2015, <http://on-broadway.nyc/>

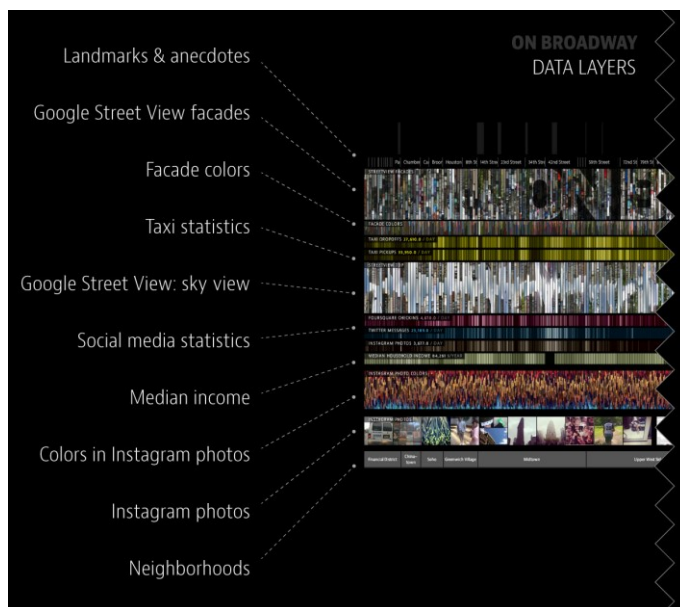


Abb. 34: Legende zu On Broadway, Bildquelle: http://on-broadway.nyc/img/broadway_layers.jpg.

Mit *On Broadway* wird die Grenze der Visualisierung von Bildern überschritten, indem von Manovich zusätzliche Datenpunkte mit dargestellt wurden, etwa Daten

zu Steuerabgaben, aber auch Angaben zu Social Media Aktivitäten (Twitter und Foursquare) sowie die Mitnahme durch Taxifahrerinnen und Taxifahrern. Aufgrund der unterschiedlichen Datenquellen und der Vermengung von medial stark divergierenden Formaten (unterschiedliche Bildtypen, Tabellen, etc.) handelt es sich bei der Arbeit von Manovich wahrscheinlich um eine der multimedialsten und «datengesätigtsten» Ansätze, die in diesem Text besprochen werden.⁵⁰

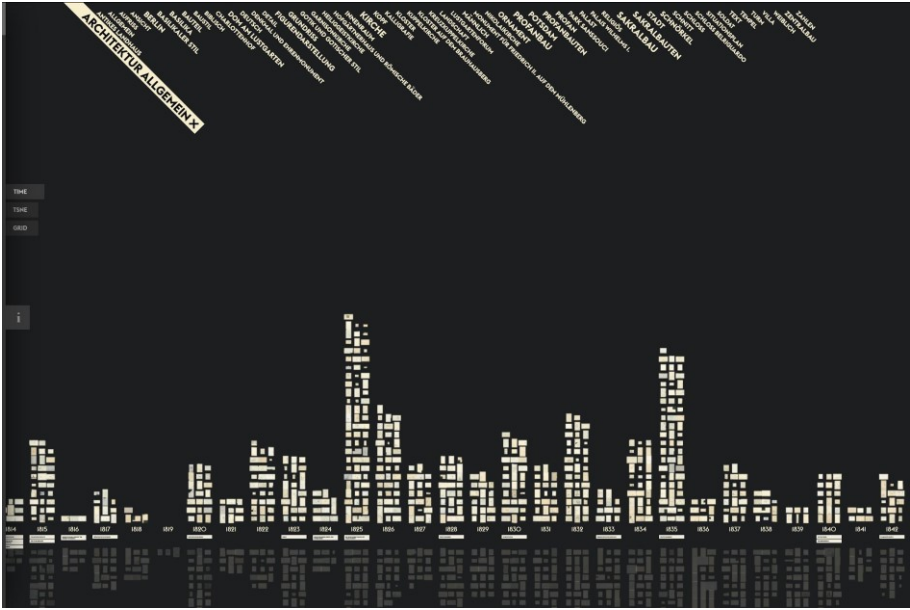


Abb. 35: Urban Complexity Lab, Fachhochschule Potsdam, *Vergangene Visionen* – Aus der Feder von Friedrich Wilhelm IV., 2017–2018, <https://vikusviewer.fh-potsdam.de/fw4/>

Weitere Formen von Plots haben Katrin Glinka, Christopher Pietsch und Marian Dörk mit der Anwendung *Vergangene Visionen* – auch unter dem englischen Titel *Past Visions* bekannt – vorgestellt. Dort wird die Anordnung von Zeichnungen über einem Zeitstrahl in der Kombination mit Metadaten dynamisch angepasst.⁵¹ Zeichnungen von Friedrich Wilhelm IV. aus der Graphischen Sammlung der Stiftung Preußische Schlösser und Gärten Berlin-Brandenburg werden nach Entstehungsjahr mit kleinen Vorschaubildern in einer Art Säulendiagramm dargestellt. Je grösser ein Begriff am oberen Rand angezeigt wird, desto häufiger kommt er als Schlagwort in den Metadaten von Zeichnungen vor. Gleichzeitig dienen diese Begriffe als Filter. Wird eines dieser Schlagwörter angeklickt, fallen alle digitalen Objekte, die dieses nicht enthalten, unter den Zeitstrahl und werden ausgegraut. Wird in die Darstellung

50 Manovich, Exploring Urban Social Media: Selficity and On Broadway.

51 Glinka/Pietsch/Dörk, Past Visions and Reconciling Views: Visualizing Time, Texture and Themes in Cultural Collections.

hineingezoomt oder ein Werk angeklickt, erscheint die Zeichnung als hochaufgelöstes Digitalisat, während am rechten Rand weiterführende Informationen wie Beschreibungen, Werkangaben und die zugeordneten Schlagworte eingeblendet werden. Zur groben historischen Orientierung befinden sich unterhalb des Zeitstrahls kurze Beschreibungen verschiedener Stationen im Leben von Friedrich Wilhelm IV. Aus diesem und weiteren vergleichbaren Sammlungsvisualisierungs-Projekten, die auf diese Weise zwischen 200 und 6900 Digitalisaten darstellen, ist der VIKUS Viewer⁵² hervorgegangen, der open source verfügbar ist.⁵³ Die Kombination zwischen Zeitstrahl und Plot im VIKUS Viewer und die dynamische Anpassung der Auswahl in der Visualisierung mit Filtern dient, wie von Glinka, Pietsch und Dörk beschrieben, der Übersichtlichkeit und Information, indem von einer Sicht auf einen ganzen Sammlungsbestand aus der Ferne auf das einzelne Objekt und die Angaben dazu gewechselt werden kann.⁵⁴ Die Nutzerführung ist sehr intuitiv. Beim Aufrufen der Anwendung wird am linken Seitenrand ein Informationsbereich eingeblendet, der die speziellen, jedoch überschaubaren Navigationsmöglichkeiten erklärt. Ein solches Vorgehen empfiehlt sich bei Interfaces, die Interaktionsmöglichkeiten einsetzen, die neu oder selten sind und aufgrund der Navigationsgewohnheiten auf Websites, in verbreiteten Programmen oder im alltäglichen Gebrauch der Geräte nicht als allgemein bekannt vorausgesetzt werden können. Bei interaktiven Anwendungen, der Navigation durch 3D-Welten (Virtual Reality), Games und Computersoftware ist es deshalb mittlerweile verbreitet, die Nutzenden zu Beginn auf eine virtuelle Tour mitzunehmen oder ein paar einfache Aufgaben zu stellen, um diese mit den vorgesehenen Navigations- und Interaktionsmöglichkeiten vertraut werden zu lassen.

52 <https://vikusviewer.fh-potsdam.de/>

53 <https://github.com/cpietsch/vikus-viewer>

54 Glinka/Pietsch/Dörk, *Past Visions and Reconciling Views: Visualizing Time, Texture and Themes in Cultural Collections*. Zu *Past Visions* siehe auch <https://uclab.fh-potsdam.de/wp/wp-content/uploads/dhd2017-aus-programm.pdf> und DHD-2016 Konferenzband sowie Glinka u. a., *Linking Structure, Texture and Context in a Visualization of Historical Drawings by Frederick William IV (1795–1861)*. und Glinka/Dörk, *Museum im Display: Visualisierung kultureller Sammlungen (Vikus)*.

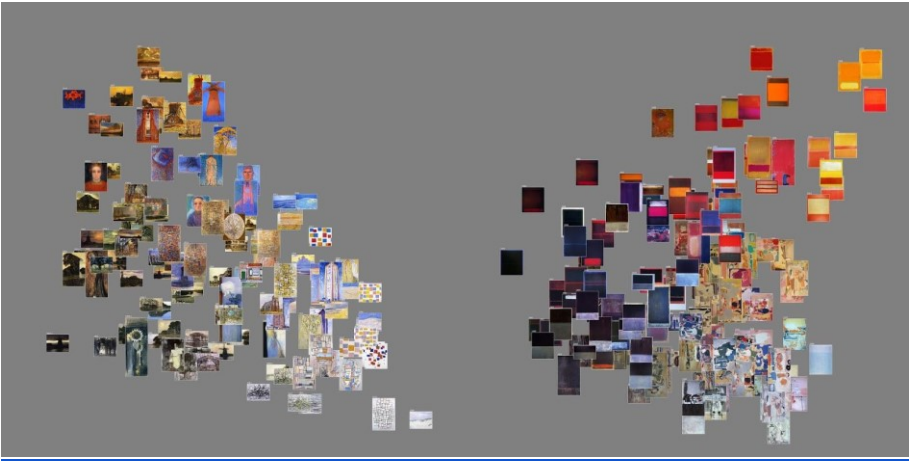


Abb. 36: Lev Manovich, *Mondrian vs Rothko: Revealing the Comparative «Footprints» of the Modern Painters*, 2011. Zwei Image Plots: links 128 Gemälde von Piet Mondrian (entstanden 1905–1917), rechts 151 Gemälde von Mark Rothko (entstanden 1944–1957). Die X-Achse gibt jeweils den Helligkeitsmittelwert, die Y-Achse den Sättigungsmittelwert wieder, <http://lab.culturalanalytics.info/2016/04/mondrian-vs-rothko.html>. Bildquelle: <http://lab.culturalanalytics.info/2016/04/mondrian-vs-rothko.html>

Neben Datenpunkten können auch Farbskalen und allgemein Clustering-Algorithmen zur visuellen Darstellung von Bildern verwendet werden. Lev Manovich spricht bei dieser Visualisierungstechnik, die entweder in einem Liniendiagramm oder in einer Punktwolke die Datenpunkte mit Vorschaubildern überlagert, von *image plots*.⁵⁵ Ein auf einem Liniendiagramm aufbauender Image Plot war bereits in *Photo-trails* zu beobachten. *Mondrian vs Rothko* ist nun ein Beispiel von zwei relativ simplen Plots im Vergleich, die auf einer Punktwolke aufbauen.⁵⁶ Die nach Helligkeit und Sättigung angeordneten Werke lassen deutlich werden, dass bei Piet Mondrian viel mehr dunkle Gemälde einen hohen Sättigungsgrad aufweisen, während bei Mark Rothko die hellen Gemälde die am stärksten gesättigten sind.

55 Manovich, *How to Compare One Million Images?*, S. 264.

56 Manovich, *How to Compare One Million Images?* Zugang zum Manuskript: http://manovich.net/content/04-projects/072-how-to-compare/68_article_2011_sm.pdf, siehe auch einführend: <http://manovich.net/index.php/projects/style-space> sowie die Dokumentation von Manovich zu ImagePlots: https://docs.google.com/document/d/1zkeik0v2LJmi1TOK4OxT7dVKJO7oCmx_fNP8SYdTGU/edit?hl=en_US#heading=h.muw9srkh6ku1 und <http://lab.softwarestudies.com/p/imageplot.html>.

3.5 Visualisierung der iterativen Digitalisierung

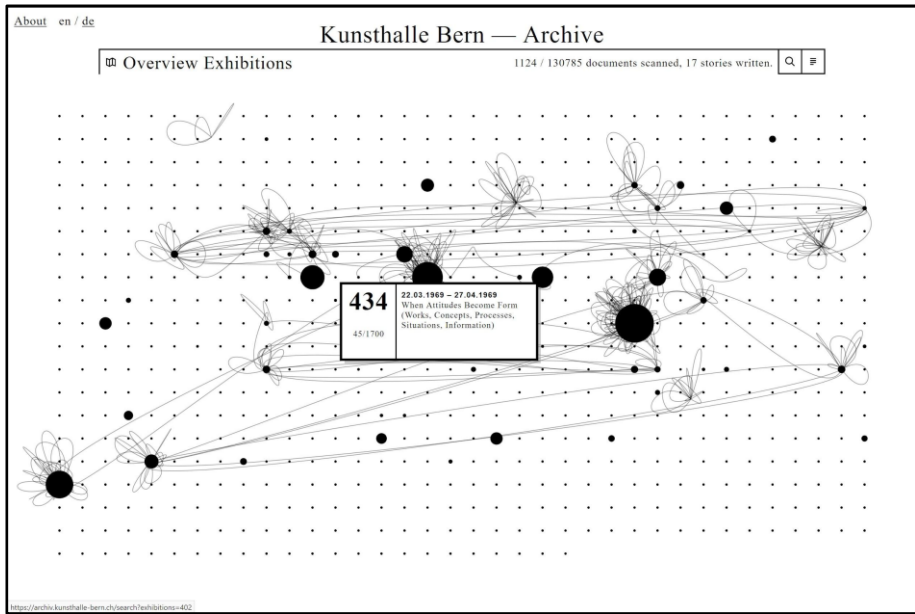


Abb. 37: Kunsthalle Bern – Archiv, 2018, <https://archiv.kunsthalle-bern.ch/de/overview>

Einen besonderen Weg geht die Kunsthalle Bern mit ihrem Archiv, das seit Juli 2018 online ist.⁵⁷ Eine interaktive Datenvisualisierung zeigt laufend an, welche Dokumente aus den über 700 Ausstellungen, die seit der Eröffnung der Institution 1918 gezeigt wurden, bereits digitalisiert vorhanden sind. Forschende, die das Archiv zu Recherchezwecken besuchen, werden von der Kunsthalle Bern aufgefordert, an einem bereitgestellten Arbeitstisch mit einer installierten und über ein Computer-Interface gesteuerten Kamera digitale Abbildungen der benötigten Dokumente anzufertigen und die wichtigsten Metadaten zu erfassen.⁵⁸ Die Archivbesuchenden stellen damit nicht nur zu ihren eigenen Zwecken Reproduktionen her, sondern digitalisieren sie so, dass sie in guter Qualität auch weiteren Nutzenden zur Verfügung stehen. Die Digitalisierung erfolgt also schrittweise und wird gleichzeitig über die Website für Externe sichtbar gemacht, die zunehmend einen grösseren Einblick ins Archiv der Kunsthalle erhalten. Die Punkte in der Visualisierung stehen für die einzelnen Ausstellungen und wachsen an, je mehr Dokumente digital vorliegen. Verbindungslinien stellen die Forschungs- bzw. Rechercheinteressen einzelner Archivnutzender dar. Die digital vorhandenen Dokumente können direkt über die Visualisierung aufgerufen und eingesehen werden.

57 <https://archiv.kunsthalle-bern.ch>

58 Eine ausführlichere Erläuterung des Vorgehens ist unter <https://kunsthalle-bern.ch/online-archiv/> zu finden.

3.6 Kombinationen

Die oben beschriebenen visuellen Ansätze, aber auch textuell/semantische Zugänge können über Kombinationen zusammengebracht werden. Als Erweiterung werden auch sogenannte Dashboards eingesetzt. Sie kommen zum Einsatz, damit die Steuerung mehr oder minder intuitiv erfolgt und dennoch die unterschiedlichen Zugangsmöglichkeiten visuell sichtbar bleiben. Sobald mehrere Visualisierungen in Kombination gezeigt werden, kann von einem Dashboard gesprochen werden, das durchaus auch als Frontend geeignet ist, um eine andere Sicht auf die Inhalte einer Datenbank von Archiven und anderen Kulturinstitutionen zu geben.⁵⁹ Solche Kommandozentralen sind in ihrem Aufbau nicht unproblematisch, da Vorwissen vorausgesetzt wird, sodass die Navigation (annähernd) selbsterklärend bleibt.



Abb. 38: Urban Complexity Lab, Fachhochschule Potsdam in Kooperation mit dem Theodor Fontane Archiv, Fontanes Handbibliothek, 2018–2019, <https://uclab.fh-potsdam.de/ff/>

Dieser Ansatz wird von Viktoria Brüggemann u. a. beschrieben, die vor allem explorative Aspekte und zufälliges Auffinden über Möglichkeiten der Betrachtung aus Distanz und Nähe formulieren.⁶⁰ Anhand der Handbibliothek Theodor Fontanes wird eine Kombination von Zugriffen ebenfalls durch das UCLAB (Jan-Mark Bludau u. a.) umgesetzt, womit Ähnlichkeiten sichtbar und die Lesespuren Fontanes nachverfolgbar gemacht werden sollen. Die visuelle Umsetzung wird als «elastisch» beschrieben, wodurch Zoom-Vorgänge eine neue Qualität erhalten, da nicht nur vergrößert,

59 Siehe einführend Few, *Information Dashboard Design*.

60 Brüggemann/Bludau/Dörk, *Zwischen Distanz und Nähe: Formen der Betrachtung und Bewegung in (digitalen) Sammlungen*.

sondern auch auf einen bestimmten Ausschnitt fokussiert werden kann.⁶¹ Grundsätzlich erhalten Nutzende über zwei Modi Zugang zum Materialfundus. Der erste Modus visualisiert die Verteilung der Lesespuren entlang einer linearen Lesereihenfolge der Bücher. Durch die Scrollfunktion kann fließend zwischen den Granularitätsebenen gewechselt werden, wobei das Scrollen nach oben die Abstraktion, nach unten den Detailgrad erhöht. Der zweite Modus stellt die Gesamtverteilung der Lesespuren aller Bücher einzelner Autoren dar. Dabei werden entsprechend der Granularitätsebene Bücher oder Autoren mit einem ähnlichen Lesespurvorkommen räumlich näher dargestellt.

Interfaces wie *Fontanes Handbibliothek*, *Topographie der Gewalt* [siehe dazu auch oben zu Karten und Geographical Information Systems], *Raoul Hausmann*, *Deutsche Digitale Bibliothek visualisiert*, *Coins* oder *Vergangene Visionen* schaffen einen Zugang zu kulturellen Inhalten. Die in einem ansprechenden Design umgesetzten Visualisierungen ermöglichen einen Überblick, nehmen vielfältige Kontextualisierungen vor und führen an einzelne Dokumente, Kunstwerke oder Artefakte heran. Wie Katrin Glinka und Marian Dörk betonen, liegt die Stärke von Informations- und Datenvisualisierung darin, komplexe Zusammenhänge umfangreicher Daten in visuelle Arrangements zu übersetzen, indem ausgewählte Strukturen und Dimensionen der Daten in optisch schnell erfassbare Muster überführt werden. Auf diese Weise können vorgängig formulierte Hypothesen überprüft werden. Die Übersetzung unübersichtlicher Zahlenreihen und Tabellen in visuell fassbare Formate ist Voraussetzung für neue Erkenntnisse, die anders nicht möglich wären.⁶² Durch Interaktion erhalten Visualisierungen weitere Informationsebenen. Auf diese Weise dienen Visualisierungen nicht nur als Mittel zur Präsentation von Ergebnissen, sondern können genauso als eine Methode verstanden werden, die Forschungs- und Erkenntnisprozesse unterstützt.⁶³ Zur Visualisierung von Daten können entweder abstrakte Formen (wie beispielsweise bei der *Deutschen Digitalen Bibliothek visualisiert*) oder Objektbilder selbst (wie bei den Plots von Lev Manovich) eingesetzt werden, oder beides in Kombination (wie in *Vergangene Visionen*). Bei kombinierten Darstellungen können visuelle Anordnungen sowohl aufgrund von Bild- als auch aufgrund von Metadatenähnlichkeit zustandekommen. Algorithmen wie MDS (*multidimensional scaling*) oder t-SNE (*t-distributed stochastic neighbor embedding*) können für grosse Bildbestände durch Dimensionsprojektionen einfache, zusammenhängende Layouts basierend auf Ähnlichkeit generieren und ermöglichen so die Repräsentation grosser Bilddatensätze in dynamischen Layouts.⁶⁴ Im Kontext von digitalen Archivzugängen bieten

61 Bludau u. a., Reading Traces: Scalable Exploration in Elastic Visualizations of Cultural Heritage Data.

62 Glinka/Dörk, Zwischen Repräsentation und Rezeption: Visualisierung als Facette von Analyse und Argumentation in der Kunstgeschichte, S. 237.

63 Glinka/Dörk, Zwischen Repräsentation und Rezeption: Visualisierung als Facette von Analyse und Argumentation in der Kunstgeschichte, S. 244.

64 Glinka/Dörk, Zwischen Repräsentation und Rezeption: Visualisierung als Facette von Analyse und Argumentation in der Kunstgeschichte, S. 245–246.

interaktive Datenvisualisierungen einen explorativen Zugang zu den Beständen. Da nicht zwingend ein Vorwissen und eine bestimmte Suchabsicht vorhanden sein müssen, nimmt die Attraktivität zur Interaktion mit einer Sammlung auch für ein breiteres Publikum zu. Unbekannte Bestände gelangen über Visualisierungen eher an die Oberfläche, wodurch neue Interessen und Forschungsfragen entstehen können.⁶⁵

In der Beschäftigung mit Interfaces zur Exploration, Navigation und Analyse von Sammlungen weisen Florian Windhager u. a. kritisch auf die Grenzen von interaktiven Datenvisualisierungen hin.⁶⁶ Diese Einschränkungen lassen sich sowohl auf der Ebene der Informationsvisualisierung, der digitalen Modellierung, der Kurations- und Sammlungspraxis als auch generell an den kulturellen Bedingungen festmachen, unter denen Sammlungen zustande kommen. Konkret hängt jede Visualisierung von Entscheidungen ab, was wie, mit welcher Technologie und aus welchem Interesse heraus umgesetzt wird, ohne dass dies für Nutzende an der Oberfläche zwingend erkennbar sein muss, wenn nicht entsprechende Kontextinformationen darüber Aufschluss geben. Den Einschränkungen, was einzelne Visualisierungsformen darzustellen in der Lage sind, kann entgegengewirkt werden, indem multiple Ansichten zur Verfügung gestellt werden. Transparenz entsteht, wenn der Programmiercode offen zugänglich ist, sodass nachvollzogen werden kann, was bei einer Anwendung im Hintergrund ausgeführt wird. Bereits die Generierung, Aufbereitung und Modellierung der Daten hat einen Einfluss darauf, wie diese genutzt, verarbeitet und dargestellt werden können. Ins Bewusstsein gerufen werden sollte ausserdem die Tatsache, dass Selektivität auch auf historisch gewachsene Strukturen zurückgeht. Dies tritt beispielsweise dann in Erscheinung, wenn Informationen aus historischen Sammlungskatalogen aus Zeitgründen ohne Anreicherung in digitale Erfassungs- und Sammlungsverwaltungssysteme übernommen worden sind. Dadurch kann es passieren, dass die über die Jahre entstandenen Kanonisierungen und kuratorische Entscheidungen unreflektiert zementiert werden. Sammlungen, schliesslich, befinden sich in bestimmten kulturellen und sozialen Kontexten und sind unter bestimmten Bedingungen entstanden. Deshalb können sie nur dann differenziert interpretiert werden, wenn diese Zusammenhänge berücksichtigt bleiben.⁶⁷ Sowohl im Projekt *Deutsche Digitale Bibliothek visualisiert*, als auch in weiteren Projekten des Urban Complexity Labs, wurde nach Lösungen gesucht, die spezifisch auf die Eigenschaften der Daten einer bestimmten Sammlung Rücksicht nehmen.

Ein Dashboard erlaubt, grosse Mengen von komplexen Daten in Echtzeit zu überwachen und über verschiedene Formen der Datenvisualisierung die Werte aus unterschiedlichen Quellen parallel in der Übersicht zu behalten. Dashboards werden

65 Glinka/Dörk, Zwischen Repräsentation und Rezeption: Visualisierung als Facette von Analyse und Argumentation in der Kunstgeschichte, S. 249.

66 Windhager u. a., Zur Weiterentwicklung des "cognition support": Sammlungsvisualisierungen als Austragungsort kritisch-kulturwissenschaftlicher Forschung.

67 Windhager u. a., Zur Weiterentwicklung des "cognition support": Sammlungsvisualisierungen als Austragungsort kritisch-kulturwissenschaftlicher Forschung, S. 343–344.

oftmals im Backend von Systemen eingesetzt, was ein Monitoring des Erfolgs von veröffentlichten Beiträgen genauso wie die Anzahl und Herkunft von Zugriffen auf Websites (z. B. mit Google Analytics) oder Social Media Accounts (z. B. mit Twitter Analytics) über unterschiedliche Diagramme erlaubt. Mit solchen statistischen Auswertungen werden die Komplexität von unübersichtlichen Daten reduziert und einzelne Parameter über aussagekräftige Visualisierungen für die menschliche Wahrnehmung erst fassbar. Monica G. Maceli and Kerry Yu beschreiben im Kontext von Archiven ein Dashboard, mit dem die von den Sensoren in den Magazinen gemessenen klimatischen Bedingungen (z. B. Temperatur und Feuchtigkeit) ständig überwacht werden können. Das ist an sich nicht neu, da zu diesem Zweck bereits verschiedene kommerzielle Monitoringsysteme eingesetzt werden. Bemerkenswert an diesem Projekt ist jedoch, dass bewusst auf die Verwendung von open source Tools gesetzt wurde. Das hat den Vorteil, dass Anpassungen entsprechend der Bedürfnisse von Archivarinnen und Archivaren vorgenommen werden können.⁶⁸ Die Flexibilität eines eigens gestalteten Interfaces ist, dass nicht auf eine Standardlösung zurückgegriffen werden muss, die für die eigenen Daten nur eine mässig befriedigende Funktionalität bietet, sondern ein massgeschneidertes Layout möglich wird.

Schliesslich ein weiteres Beispiel für ein Dashboard, das im Frontend zum Interface einer Sammlung wird, ist die weiter unten bei den «Semantischen Zugangsformen» besprochene Impresso-App. Dort werden verschiedene Visualisierungen eingesetzt, um im Zeitungsportal Filtermöglichkeiten vorzunehmen und Zusammenhänge erkennen zu können.

3.7 PolyCube-System

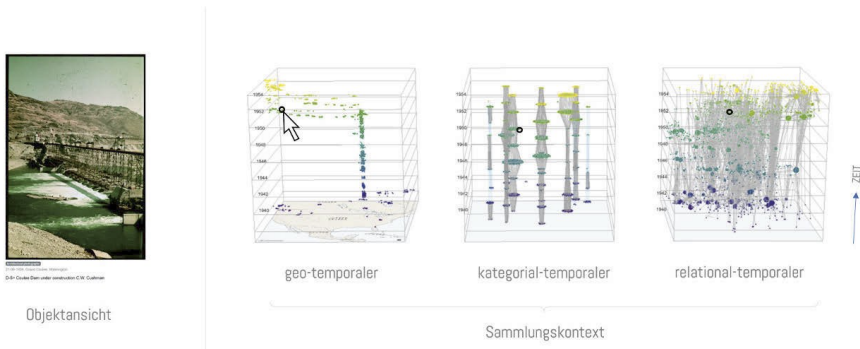


Abb. 39: PolyCube-System zur polykontextualen Visualisierung einzelner Objekte (links) in den multidimensionalen Informationsdimensionen von kulturellen Sammlungen (rechts). Bildquelle: Mayr/Windhager 2020, S. 239. Der Artikel und die Abbildung sind lizenziert unter einer CC-BY Lizenz.

68 Maceli/Yu, Usability Evaluation of an Open-Source Environmental Monitoring Data Dashboard for Archivists.

Eva Mayr und Florian Windhager gehen davon aus, dass kulturelle Objekte nie nur in einem Kontext erschöpfend verortet werden können, sondern «polykontextual» repräsentierbar sind.⁶⁹ Der digitale Raum eröffnet, unabhängig von den Restriktionen des Realraums, einen Informationsraum, in dem Inhalte aus Datenbanken anders gruppiert und kontextualisiert werden können. In ihrem Projekt *PolyCube* haben sie «synoptische Optionen der Kontextualisierung von Objekten innerhalb einer Sammlung» untersucht. Multiperspektivische Ansichten bilden ein Interface, das mehrere Datendimensionen von Objekten gleichzeitig darstellt. Das PolyCube-System kombiniert geo-temporale, kategorial-temporale und relational-temporale Perspektiven mit nahtlosen Übergängen zur Darstellung einer kulturellen Sammlung, wodurch ein Verständnis von komplexen Kontextualisierungen möglich wird. Im Rahmen der Forschungen zu PolyCubes wurde festgestellt, dass Sammlungsinterfaces makroskopische Kontextualisierungen vorwiegend mit Blick auf eine Sammlung ermöglichen. Mit PolyCube sollen aber nicht nur die in einer Sammlung enthaltenen Objekte selbst in den Fokus kommen, sondern Informationen und Kontext (allgemein aus dem Web oder aus anderen Sammlungen) miteinbezogen werden. Gerade in der Verweisstruktur von kulturellen Objekten steckt das Potenzial, diese Informationen im Digitalen zu verknüpfen und die Daten in einem semantischen Netz zu strukturieren, das als Grundlage für Visualisierungen und Interfaces dient. In Bezug auf Kunstsammlungen unterscheiden sie eine biografische, eine kunsthistorische und eine makrohistorische Kontextualisierung. Für eine polykontextuale Visualisierung bedeutet das also, dass neben der Werkanalyse mit Bezug auf das Künstlerleben und der Verortung eines Objekts in grösseren kunsthistorischen Entwicklungssträngen auch parallele gesellschaftliche Prozesse (z. B. in Politik, Medien, Religion, Wirtschaft, Technologie, Kultur oder Wissenschaft) berücksichtigt werden müssen. Daraus entsteht ein Modell, das Übergänge von Close-Reading- und Distant-Reading-Perspektiven über mehrere Ebenen miteinander verbindet.⁷⁰

69 Mayr/Windhager, Vor welchem Hintergrund und mit Bezug auf was? Zur polykontextualen Visualisierung kultureller Sammlungen, S. 235.

70 Mayr/Windhager, Vor welchem Hintergrund und mit Bezug auf was? Zur polykontextualen Visualisierung kultureller Sammlungen, S. 236–242.

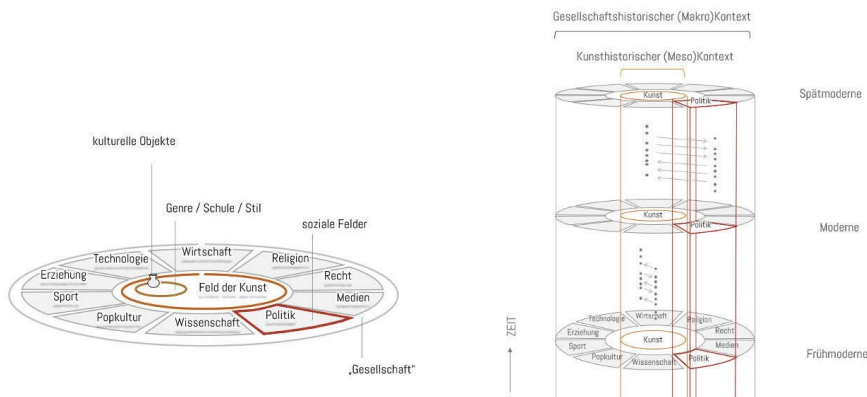


Abb. 40: Makrosoziologische Kontextualisierung von Objekten im Feld der Kunst (links) und zeitorientierte Darstellung zur Visualisierung soziokultureller Entwicklungen (rechts). Bildquelle: Mayr/Windhager 2020, S. 242.

Eva Mayr und Florian Windhager verstehen Sammlungsvisualisierungen im digitalen Raum als ein zukunftsgerichtetes Feld zur Kunst- und Kulturvermittlung. Solche Datenvisualisierungen sollten so einfach, flexibel, attraktiv und generös wie möglich umgesetzt werden, um einen benutzerfreundlichen Zugang zu bieten, der eine leichte Orientierung, verschiedene Zugänge für unterschiedliche Bedürfnisse und Interessen bietet und zudem ästhetisch überzeugend an die strukturell komplexen Zusammenhänge kultureller Objekte heranführt.⁷¹ Der Code von PolyCube steht auf GitHub zur Verfügung.⁷²

71 Mayr/Windhager, Vor welchem Hintergrund und mit Bezug auf was? Zur polykontextualen Visualisierung kultureller Sammlungen, S. 242–243.

72 <https://github.com/bigdata-vis/polycube>

3.8 Virtueller Raum



Abb. 41: Rachel Hendery, Western University of Sydney, und Andrew Burrell, University of Technology Sydney, Glossopticon VR, 2017, <http://glossopticon.com/>

Neben flachen Interfaces auf einem Bildschirm weisen virtuelle Umgebungen ein räumliches Interface auf, das einen dreidimensionalen Zugang zu kulturellen Daten ermöglicht. Bei *Glossopticon VR* können mit einem VR-Headset historische Audioaufnahmen von über 1500 Sprachen aus dem Pazifikraum erkundet werden. Damit wird ein Zugang zur Sammlung von linguistischen Archivdaten des *Pacific and Regional Archive for Digital Sources in Endangered Cultures* (PARADISEC) geschaffen. Die Anwendung wurde sowohl als Forschungstool als auch für eine Ausstellung, und damit für eine breite Öffentlichkeit, entwickelt. In der Anwendung lokalisieren Lichtsäulen die einzelnen Sprachen auf einer dreidimensionalen Karte, dabei korrespondiert die Größe der halbtransparenten Halbkugeln mit der Anzahl Sprechenden; zusätzlich werden weiterführende Angaben eingeblendet. Beim Navigieren durch die virtuelle Umgebung werden verschiedene Soundfiles ausgelöst und parallel abgespielt, sodass ein Klangteppich aus verschiedenen Stimmen erkundet werden kann.⁷³ Das Besondere an diesem Beispiel ist nicht nur, dass es für die Informationsvisualisierung auf Virtual Reality Technologie setzt, sondern aufzeigt, dass auch abstrakte Daten wie Audiospuren visualisiert werden können. Zudem wird deutlich, dass Interaktion bei Tondokumenten auch anders gestaltet sein kann, als diese in einem Player als einzelne Audiostreams abzuspielen. Der Programmcode steht zur Nachnutzung in eigenen Projekten zur Verfügung.⁷⁴

73 Burrell/Hendery/Thieberger, Glossopticon.
74 <https://github.com/clavis-magna/glossopticon-webvr>

4. Semantische Zugangsformen

Neben visuellen Zugangsformen werden aktuell auch textuelle Zugänge diskutiert, die über eine reine Volltextsuche hinausgehen. Zentral in diesem Zusammenhang sind semantische Aufbereitungen, die manuell oder automatisiert geschehen. Die textuellen Zugangsformen sind naturgemäss aus visueller Perspektive weniger beeindruckend. Aufgrund der historischen Einbettung in den Umgang mit archivischem Material – Editionen etwa basieren grösstenteils auf semantischen Zugängen – lohnt es sich jedoch darauf einzugehen. An dieser Stelle steht dementsprechend nicht die Umsetzung der Zugangsform im Fokus, sondern die aufbereiteten Informationen. Semantische Zugänge können über Markierungen in Texten (etwa Hervorhebungen von Eigennamen) oder in klassischer Listenform (analog zu gedruckten Registern) visuell umgesetzt werden. Es ist selbstredend auch möglich, dass semantische Informationen in visueller Form dargestellt werden, womit sich semantische und visuelle Zugänge überschneiden.

Im folgenden Teil werden deshalb extrahierte Informationen thematisiert, die gezielte Zugriffe auf gesamte Archive oder Archivbestände erlauben. Zusätzlich ist es aber auch denkbar, dass ganze Dokumentengruppen mittels Narration abgedeckt werden.

4.1 Storytelling

Eine häufig nur implizit reflektierte Zugangsform ist das sogenannte Storytelling. Bereits die Aufbereitung von Dokumenten, unabhängig davon ob visuell oder textuell, ist eine Form von Narration. Am einfachsten lässt sich dies im archivischen Kontext anhand von Archivplänen nachvollziehen, die häufig historisch gewachsen sind und eine sehr abstrakte Erzählung des Zusammenkommens der Dokumente ergibt. Es ist in dem Zusammenhang kein Zufall, dass bei der Einführung in Archive die historischen Zusammenhänge erläutert und somit die Archivtektonik in einen Kontext gesetzt werden. Gemäss Ann Laura Stoler hilft das Verfolgen «along the archival grain» der Identifikation von archivischen Leerstellen (was verschwiegen wird) und *top-down* Setzungen aufgrund einer Verwaltungsperspektive, die sich selbst legitimiert und dadurch eine narrative Form einnimmt.⁷⁵

Storytelling kann aber auch viel gezielter und bewusster eingesetzt werden, um auf einzelne Inhalte aufmerksam zu machen, die nicht nur als Teile einer Sammlung präsentiert, sondern als zentrale Stücke hervorgehoben und in ein gemeinsames Narrativ eingebunden werden. Im musealen Kontext wird Storytelling vielfach in virtuellen Ausstellungen oder gleich als eigene Vermittlungsform eingesetzt, wie etwa das Rijksmuseum demonstriert (<https://www.rijksmuseum.nl/en/stories>). Storytelling findet sich aber auch in einigen Archiven, etwa als «Vitrinen» oder «Objekte des Monats», die einzelne Archivadokumente kontextualisieren und einbetten.⁷⁶ Die enger

75 Stoler, *Along the Archival Grain*.

76 Siehe bspw. das «Schaufenster» des Staatsarchiv Luzern: <https://staatsarchiv.lu.ch/schaufenster>

werdende Verknüpfung von (wissenschaftlicher) Literatur mit Objekten aus Erinnerungsinstitutionen, sei das über Links oder andere Verknüpfungsformen (bspw. Pop-ups mit Verweisen oder eingebettete Informationen), kann zusätzlich verschiedene Ausprägungen von Storytelling näher an Archive und andere Erinnerungsinstitutionen heranbringen.⁷⁷

Die Plattform OMEKA verfolgt eigens das Ziel, die digitale Narration zu unterstützen. Dazu wird ein Content Management System (CMS) zur Verfügung gestellt, das klar strukturierte Metadaten beinhaltet und die digitalen Objekte als eigenständig manipulierbare Objekte versteht, die in unterschiedliche Kontexte gestellt und verwendet werden können.⁷⁸ Wie von H. E. Green beschrieben liegt der Vorteil der Nutzung von OMEKA und vergleichbaren Content-Management-Systemen zur Verwaltung und Veröffentlichung von Sammlungsobjekten im Web in der Möglichkeit, die Komplexität der Objekte abzubilden, während gleichzeitig die technischen Voraussetzungen relativ niedrig angesetzt werden bzw. durch Spezialistinnen und Spezialisten übernommen werden können.⁷⁹

Narrative Zugänge lassen sich auch visuell umsetzen, wie mit Bezug auf Erzählungen von Hannah Schwan u. a. demonstriert wurde.⁸⁰ Sprünge und Verwicklungen in der Erzählung werden durch visuelle Marker vermittelt, damit lassen sich auch in der Umkehrung Erzählstränge nicht nur textuell nachvollziehen, sondern visuell analysieren.

Gerade im Zusammenhang mit der aktiven Nutzung von Social Media wird Storytelling zentral, da die unkontextualisierte Publikation von Objekten oder Dokumenten nur zu wenig Echo führt und dadurch insgesamt weder ein grösseres noch ein neues oder anderes Publikum angesprochen wird. Bereits die strategische Verwendung von hashtags wie #otd (für *on this day*) oder #archivCH (für Archive Schweiz) führt zu einer breiteren Rezeption und der Einbettung in laufende Diskussionen. Bei aller Kürze, in der hier das Thema Social Media behandelt wird, sei an der Stelle noch angemerkt, dass erfolgreich betriebene Accounts einer aktiven Betreuung und einer Strategie bedürfen.

4.2 Identifizierbare Entitäten

Näher an klassischen Suchzugängen befinden sich Register, die etwa Personen, Orte, Organisationen oder Begriffe (sog. identifizierbare Entitäten) auffindbar machen, bzw. visuell in Dokumenten markieren. Der Vorteil derart aufbereiteter Daten ist die

77 Siehe dazu auch Marsh u. a., die Formen der Digitalisierung in Erinnerungsinstitutionen bevorzugt über die Narration vermittelt sehen: Marsh u. a., *Stories of Impact: The Role of Narrative in Understanding the Value and Impact of Digital Collections*.

78 Zu OMEKA siehe: <https://omeka.org/about/project/>.

79 Green, *Publishing Without Walls: Building a Collaboration to Support Digital Publishing at the University of Illinois*. Siehe auch Fitzpatrick, *Giving It Away: Sharing and the Future of Scholarly Communication*; Fitzpatrick, *Scholarly Publishing in the Digital Age*.

80 Schwan u. a., *Narrations – Visualizing Narrative Levels and their Correlations with Temporal Phenomena*.

erhöhte Präzision durch eine eindeutige Namenszuweisung und die gezielte Führung von Suchzugriffen. Dies trifft besonders im Vergleich zu Volltextsuchen zu, die nicht mehr liefern (können) als den Abgleich von *strings*, also von Folgen von Buchstaben. Die Identifikation der Entitäten kann sowohl händisch als auch unterstützt durch *machine learning* Algorithmen erfolgen.⁸¹ Für beide Vorgänge ist es zentral, dass eine Begriffsklärung erfolgt, um gleich oder ähnlich benannte Entitäten auseinanderzuhalten.

Als Beispiel für diese Herausforderung gilt die Erkennung von Personen- und Ortsnamen sowie Daten in einem Text. Damit wird nachvollziehbar wer, welche Orte und welche Datumsangaben in einem Text genannt werden. Diese Informationen können nachfolgend für die Erstellung von Registern oder für die Visualisierung von Zusammenhängen (wer tritt wann wo mit wem auf) genutzt werden.

Dieses automatisierte, computerwissenschaftliche Problem wird in zwei unterschiedliche Aufgaben aufgeteilt: Erstens in die *Named Entity Recognition* (Erkennung) und zweitens ins *Named Entity Linking* (Zuordnung zu einer spezifischen Entität). Diese erfolgreichen Systeme basieren auf neuronalen Netzen nach dem *deep learning* Prinzip und sind abhängig von Sprachmodellen. Mit Ausnahme des unten beschriebenen Impresso-Projekts gibt es bislang noch wenig Erfahrungen im Einsatz der Technologie in grossem Umfang für Daten aus dem GLAM⁸²-Bereich. Gleichzeitig werden vermehrt Frameworks zur Verfügung gestellt, sodass vor allem *Named Entity Recognition* Probleme mit überschaubarem Aufwand bearbeitet werden können.⁸³

4.3 Inhaltsextraktion

Analog zur Identifikation von Entitäten können auch Inhalte extrahiert und registerähnlich aufbereitet werden. Aktuell wird dabei häufig auch auf maschinelle Lernverfahren zurückgegriffen, wobei die Dokumente seltener über Annotationsverfahren verschlagwortet, sondern vielfach über Clustering-Ansätze in Themenfelder gruppiert werden.⁸⁴ Das momentan am häufigsten angewandte Verfahren wird unter dem Begriff des *Topic Modeling* zusammengefasst, worunter eine *Latent Dirichlet Allocation* (LDA) verstanden wird.⁸⁵ Die Anwendung von *topic modeling*-Algorithmen führt dabei nicht zur Identifikation eines begrifflich gefassten Themas, sondern zur Ausgabe einer Liste der häufigsten Zeichenfolgen, die in einem Cluster auftauchen bzw.

81 Siehe dazu unten Annotation, dabei müssen insbesondere die Voraussetzungen des Ansatzes, etwa das Vorhandensein von adäquaten Sprachmodellen sowie Trainingsdaten berücksichtigt werden.

82 GLAM steht für Galleries, Libraries, Archives und Museen.

83 Bekannt und bewährt haben sich vorwiegend flairNLP: <https://github.com/flairNLP/flair> sowie spaCy: <https://spacy.io/>, beide Frameworks sind Open Source.

84 Siehe dazu unten Clustering.

85 Zu LDA siehe v. a. die Arbeiten von David Blei, insbesondere: Blei/Ng/Jordan, Latent Dirichlet Allocation; Blei, Introduction to probabilistic topic models.

typisch für das Cluster sind. Aufgrund dieser Liste können Themenbereiche interpretativ identifiziert werden.⁸⁶

Für geschlossene Korpora eignet sich die Aufbereitung mit der Methode, da häufig auftauchende Wortkombinationen die Behandlung eines (mehr oder minder) fixen Themas hindeuten. Dadurch erlaubt eine Aufbereitung über *topic model* einen thematischen Zugang. Wie bereits erwähnt, findet *Topic Modeling* vorwiegend bei geschlossene Dokumentenkorpora Anwendung, die aufgrund des Umfangs an Textdokumenten nicht vernünftig überblickt werden können und oft nur in Ausschnitten interessant sind. In Protokollreihen lassen sich beispielsweise mittels *Topic Modeling* spezifische Themen auffindbar machen, die nicht mittels eines Suchbegriffs aufgefunden werden können.

Die Anwendung von *topic models* führt entweder auf Seiten der Institution oder der Nutzenden zu einem Interpretationsschritt, in dem mögliche Themen in einem Korpus identifiziert werden können.⁸⁷ *Topic Modeling* eignet sich daher nicht für ganze Archivbestände, da die Breite der vorkommenden Themen so weit ist, dass Modelle mit weniger als 100 Cluster nicht zu Themen, sondern wahrscheinlicher zu Dokumentengattungen führen würden, die ohnehin schon bekannt sind.

86 Siehe dazu als Anleitung Graham/Weingart/Milligan, *Getting Started with Topic Modeling and MALLET*.

87 Siehe dazu Weingart, *Topic Modeling for Humanists*.

The screenshot displays the Impresso project interface. On the left, a search query for "Clara Schumann" is active, with filters for "service - langue - poste - travail - p" and "article text" containing "Clara Schumann". Below this, a list of topics and persons is shown, with "FR télévision - radio - chaîne - système - publicité" selected. On the right, the "Inspect search queries" panel shows a line graph of results from 1980 to 1998, with 5 results identified for Clara Schumann. A detailed view of the current selection shows a profile for Clara Schumann (Q132232), a German musician and composer, with a source link to Wikidata.

Abb. 42: Impresso – Media Monitoring of the Past, Named Entity Recognition und Linking der Person «Clara Schumann» in Zeitungen, Screenshot, <https://impresso-project.ch/app/compare?left=CgIYAgoYCAEQAhGIAEqDkNsYXJhIFNjaHVtYW5uCh4IARA-CGBIqFrNtLWZyLWFsbC12Mi4wX3RwNtFfZnl%3D>.

Im Rahmen des Projekts *Impresso – Media Monitoring of the Past* wurden die zwei Formen der Entitäten-Identifikation und die Themenextraktion anhand eines dreisprachigen Zeitungskorpus aus der Schweiz und Luxemburg durchgeführt.⁸⁸ Bereits für den, aus der Perspektive der meisten Archive, verhältnismässig kurzen Zeitraum von 200 Jahren, stellt die Identifikation von benannten Entitäten ein Problem dar, wobei sich die Mehrsprachigkeit als grössere Herausforderung herauskristallisiert hat.⁸⁹ Aufgrund des erst kürzlich abgeschlossenen Projekts kann in naher Zukunft von weiteren aufschlussreichen Publikationen zum Thema ausgegangen werden.

88 Siehe dazu <https://impresso-project.ch/app/>, zum Projekt siehe: <https://impresso-project.ch/project/objectives/>.

89 Umgesetzt im Rahmen des Projekts in Form eines Wettbewerbs: Ehrmann u. a., Introducing the CLEF 2020 HIPE Shared Task.

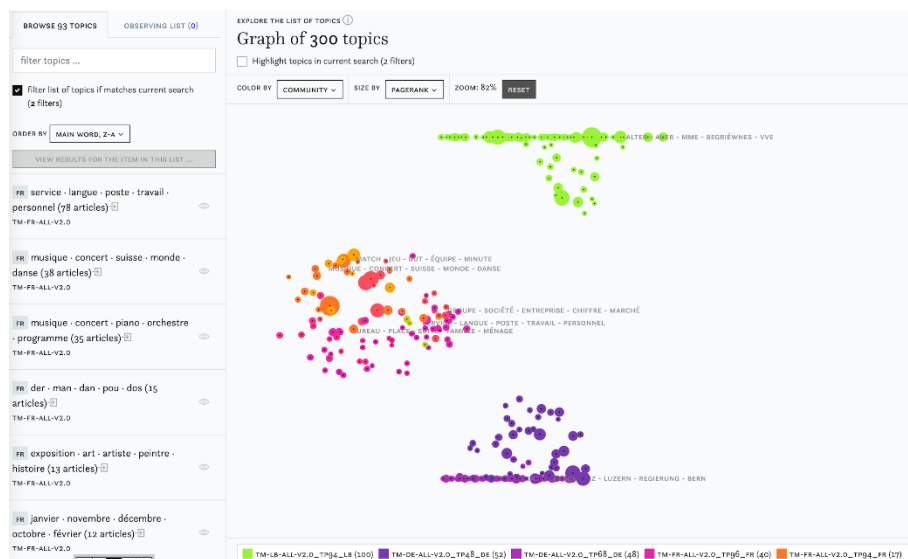


Abb. 43: Impresso, mit Topic Modeling aus Zeitungskorpora extrahierte Themenfelder, Screenshot, <https://impresso-project.ch/app/topics/?sq=CglYAgOXEAIYByABKq9Sb2JlcnQgU2NodW1hbm4KHgg-BEAIYEioWdG0tZnltYWxsLXYyLjBfdHA1MV9mcg%3D%3D>.

4.4 Text-Mining

Begrifflich eng verwandt mit der Entitäten-Identifikation und der Inhaltsextraktion ist das Text-Mining. Unter diesem Begriff versteht man die digitale Analyse von Textkorpora, etwa durch die Quantifizierung von vorkommenden Wörtern (*token*) oder die Bildung und Auszählung von Wortgruppen (*n-gram*). Dies geschieht alles vor dem Hintergrund, die durchsuchten (*mined*) Dokumente in ihrer Masse besser verstehen und einordnen zu können.

Der Zugriff auf Dokumente/Textkorpora erfolgt typischerweise über publizierte Dumps (meistens über Datenpublikationen in Repositorien) oder über APIs. Der Vorteil des Dumps ist die Möglichkeit zur selbständigen Weiterverarbeitung, während die API dem Datenhost die Option zur Anpassung und Erweiterung der Daten gibt, die bei den Nutzende in angepasster Form ankommen. Beide Formen sind gleichermaßen legitim und ergänzen sich.

Aufgrund der vielfältigen Möglichkeiten, die durch Text-Mining eröffnet werden und die unbedingt entlang der Fragestellung eingesetzt werden müssen, kann dieser Themenbereich nicht als eigenständiger Zugang beschrieben werden, sondern als Werkzeugkasten, der potentiell neue Zugänge eröffnet.

5. Querschnittsthemen

5.1 Records in Contexts (Linked Open Data Modell für Archive)

Unter Linked Data versteht man die Modellierung von strukturierten und miteinander verknüpften Daten nach Standards des World Wide Web Consortium (W3C).⁹⁰ Daten müssen dazu adressierbar und möglichst nach definierter Ontologie veröffentlicht werden. Aus technischer Sicht existiert die Möglichkeit der Publikation im Format RDF (*Resource Description Framework*) oder über Graphendatenbanken mit Schnittstellen (bspw. neo4j und SPARQL-Schnittstellen).⁹¹ Der Vorteil von Linked Data besteht sowohl in der Anschlussfähigkeit der Daten an andere Wissenssysteme als auch in der Möglichkeit, bestehende Informationen zu erweitern. Im Unterschied zu relationalen Datenbanken muss somit nicht die gesamte Architektur der Datenbank angepasst werden. Des Weiteren werden Überlegungen zur Ontologie eines Datenmodells sichtbar und damit nachvollziehbar gemacht.⁹²

Um die Vorteile von Linked Data zu nutzen, wird in der Archivwelt aktuell der Standard *Records in Contexts* (RiC) diskutiert, der den Archivgutbeschreibungsstandard ISAD (G)⁹³ sowie verwandte Standards, etwa zur Definition von Urhebern von Archivgut (ISAAR CPF),⁹⁴ ablösen soll.

Der aktuelle Stand von Records in Contexts wird durch Merzaghi beschrieben, wobei vorwiegend die Genese des Standards thematisiert wird.⁹⁵ Der Vorteil von Linked Data, vor allem mit Bezug zu Suchzugriffen, die auch über einzelne Archive hinaus weisen, wird von demselben Autor in einem 2017 erschienen Artikel beschrieben.⁹⁶ Einen Schritt weiter geht Fornaro mit der Vernetzung von Linked Data und Bildern nach dem IIIF-Standard (International Image Interoperability Framework).⁹⁷ Die Vorgehensweise macht insbesondere auf die konsequente Nutzung von APIs aufmerksam, die die Nachnutzung und die nahtlose Weiterverarbeitung ermöglichen, ohne dass die genutzten Inhalte selbst lokal abgespeichert werden müssen. Von Seiten der Anbietenden können die APIs auch entsprechend aufgerüstet werden, sodass Zugriffe kontrolliert oder spezifische Suchen angeboten werden können. Die Umsetzung

90 <https://www.w3.org/standards/semanticweb/data>.

91 Siehe dazu: <https://neo4j.com/>.

92 Für eine grundlegende Einführung aus archivischer Perspektive jedoch vor der Entwicklung von RiC siehe Gracy, *Archival Description and Linked Data: A Preliminary Study of Opportunities and Implementation Challenges*.

93 ISAD(G): General International Standard Archival Description – Second edition | International Council on Archives.

94 ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, 2nd Edition | International Council on Archives.

95 Merzaghi, *Informationen finden und Wissen verlinken: Der Weg der Metadatenstandards vom Archivregal zu den Linked Data*.

96 Merzaghi, *Neue Standards erarbeiten und Informationsnetze bauen: Verbesserung von Dienstleistungen in Archiven durch Normen und Standards*.

97 Zu IIIF siehe unten: *Bildpublikation via IIIF und Mirador*. Fornaro, *Zugang zum digitalen Archiv*.

von Linked Open Data (LOD) im archivischen Kontext ist bereits weit fortgeschritten und erste Pilotprojekte finden sich auch für die Schweiz.⁹⁸

Ein Beispiel für die Nutzung einer API ist die Erstellung einer Weboberfläche, die angebotene Informationen und Inhalte (etwa mit Datierung versehene Dossiers) aus einer Ressource abrufen und in ein eigenes Design mit einer eigenen Navigation einbindet. Darunter ist beispielsweise ein Service zu verstehen, der ein Dossier präsentiert, das sich von der Datierung her am aktuellen Tag zum 30., 50. oder 100. Mal jährt.

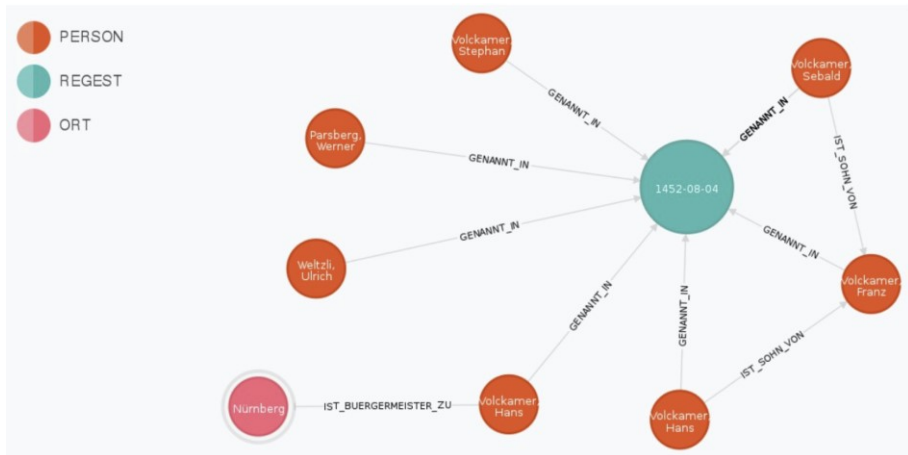


Abb. 44: Abbildung eines Linked Open Data Netzwerks. Bildquelle: Kuczera 2015, S. 4. Die Abbildung ist lizenziert unter einer CC-BY-SA Lizenz.

Im Zusammenhang mit der Umstellung auf LOD entstehen neue Zugriffsformen, weil die Datensätze durch die Verlinkung auf ein weites Netz an Informationen zugreifen und damit angereichert werden können. Die nur im Netzwerk existierenden Daten können innerhalb des Netzwerks von jedem beliebigen Punkt aus betrachtet werden. Aus diesem Grund gehören Netzwerkvisualisierungen zu den beliebten Umsetzungen mit LOD. Ausgehend von einem Dokument kann so in unterschiedliche Richtungen nach Dokumenten gesucht werden, die mindestens eine gemeinsame Verbindung aufweisen. In der Technologie ist somit ein Zugang angelegt, der unerwartete Treffer mit einer hohen Ähnlichkeit zu den gesuchten bzw. betrachteten Treffern intendiert.

Alle oben beschriebenen Visualisierungen und semantischen Zugriffe lassen sich entsprechend auf Aspekte anwenden, die in einer Linked Data Ontologie hinterlegt sind. Im Hinblick auf die Planung neuer Zugriffsformen sollte daher überlegt werden, inwiefern der Spielraum von Records in Contexts (RiC)/LOD genutzt werden

98 Siehe die Arbeitsgruppe ENSEMEN (<https://vsa-aas.ch/arbeitsgruppen/projektgruppe-ensem/>), die sich die Umsetzung von RiC für die Schweiz zum Ziel gesetzt hat und auch die Slides von Tobias Wildi und Alain Dubois: Wildi/Dubois, Alain Dubois, Tobias Wildi. Le Matterhorn RDF Data Model.

kann, um die mehr oder minder tief erschlossenen Daten als LOD für weitere Zugänge zu nutzen. Die Umsetzung von Ontologien ist dabei zeitaufwändiger als der Aufbau eigener Datenbanken, weshalb idealerweise auf bestehende Ontologien aufgebaut wird. Die Nachnutzung von Ontologien gewährleistet zudem die Verlinkung mit anderen Datenbeständen und die Anbindung an die Linked Open Data Cloud.⁹⁹ In den vergangenen Jahren wurden einige Initiativen geschaffen, die die Harmonisierung und Diskussion von Ontologien vorangetrieben hat. Daraus sind Ressourcen entstanden, die einen guten Überblick über geisteswissenschaftlich genutzte Ontologien geben:

- Für die Beschreibung von Texten (im Kontext von Editionen): <http://e-editions.ch/ontology> (im Rahmen von NIE-INE entstanden; Schichtenmodell, das einerseits bestehende Ontologien nachnutzt und andererseits die Möglichkeit für die Erstellung spezifischer Ontologien zu zentralen Informationen eines Projekts offen lässt)¹⁰⁰
- Zur Bildung, Vernetzung und Alignierung von (eigenen) Ontologien: <http://ontome.dataforhistory.org/> [ontoME], betrieben durch das Data for History Konsortium.

5.2 Linked Data als *game changer*

Ein immenser Vorteil von Linked Data, der bislang nur in wenigen geisteswissenschaftlichen Bereichen genutzt wird, ist die Möglichkeit des *machine reasoning*. Da die explizierten Ontologien von Linked Open Data maschinenlesbare Informationen enthalten und die Daten in Wissensnetzwerke eingebunden sind, können digitale Systeme sinnvolle Zusammenhänge erkennen und wiedergeben. So können weitere Zugänge eröffnet werden, die zwar logisch vorhanden, aber aufgrund der Datenfülle nicht ohne Weiteres sichtbar sind. Als ein simples Beispiel lässt sich etwa mittels *machine reasoning* die Verwandtschaft von Personen eruieren. Im Sinne von, wenn die Entität X ein Kindelement von Entität Y und die Entität Z eine Geschwisterentität von Y ist, so besteht eine Verwandtschaft zweiten Grades. Falls die entsprechenden Informationen hinterlegt sind, lässt sich dies für einen Computer problemlos auf beliebige weitere Grade erhöhen.

Potente Suchmaschinen wie Google, machen sich solche sogenannten *knowledge graphs* und die Funktion des *machine reasoning* zunutze, um Suchresultate auffindbar zu machen, die nicht nur einen eingegebenen Suchstring auf einer Website enthalten, sondern dieselben oder ähnliche Konzepte verwenden. In Bezug auf den *knowledge graph* sind Begriffe als bedeutungsgleich oder als ähnlich hinterlegt, sodass bei einer Abfrage das Gemeinte und nicht das Gesuchte als hochgerankte Treffer angezeigt werden.

Die Erstellung solcher Graphen bedingt jedoch neben der Modellierung von Ontologien auch das Vorhandensein von Sprachmodellen, die automatisierte

99 <https://lod-cloud.net/>

100 <https://www.nie-ine.ch/>

Auswertungen überhaupt erst ermöglichen. Damit wird die Welt des maschinellen Lernens betreten. Unter *machine learning* oder genauer *deep learning* verstehen wir Ansätze, die ähnliche Versprechen bezüglich Zugänglichkeit machen wie *knowledge graphs*, jedoch einen grundlegend anderen technologischen Hintergrund aufweisen.

5.3 Maschinelle Lernverfahren

Machine learning basiert auf der Annahme, dass Computer eine Vielzahl von Vergleichen innert kürzester Zeit berechnen können. Entsprechend kann diese Leistung genutzt werden, um selbstgesteuerte Optimierungen vorzunehmen oder um Muster innerhalb grosser Datenmengen zu erkennen. Auf einer sehr allgemein gehaltenen Ebene führt beispielsweise Andreas Schneider das Konzept mit Blick auf die Archive ein, geht jedoch auf die unterschiedlichen Formen des *machine learning*-Ansatzes nur marginal ein.¹⁰¹ Eine zentrale Neuerung in den vergangenen Jahren war die Einführung von *deep learning*. Das Verfahren basiert auf neuronalen Netzen, die Optimierungen der Ausgabe hin zu vorgegebenen Inputs vornehmen (sog. *supervised learning*). Die Algorithmen versuchen entsprechend durch Trainings ein Modell zu erstellen, das in der Lage ist, erwartete Bewertungen für ähnliches Material zu erzeugen.

Aus Gründen einer besseren Verständlichkeit verstehen wir hier maschinelles Lernen in einem breiten Sinn und decken somit unterschiedliche Technologien bzw. Vorgänge unter dem selben Oberbegriff ab, somit können wir anhand von Beispielen die Vor- und Nachteile kurz streifen.

Insgesamt lohnt sich die Anwendung von maschinellen Lernverfahren, wenn grosse Datenkorpora bearbeitet werden sollen und Zeitreserven für Experimente vorhanden sind. Aktuell gibt es nur sehr wenige Verfahren, die direkt und ohne Reibungsverlust in Erinnerungsinstitutionen eingesetzt werden können.

101 Schneider, Die Datenflut mit intelligenten Algorithmen bändigen.

Die Stärken von *machine learning* liegen in den zwei Aufgabenbereichen Clustering und Annotation, die hier kurz skizziert werden.

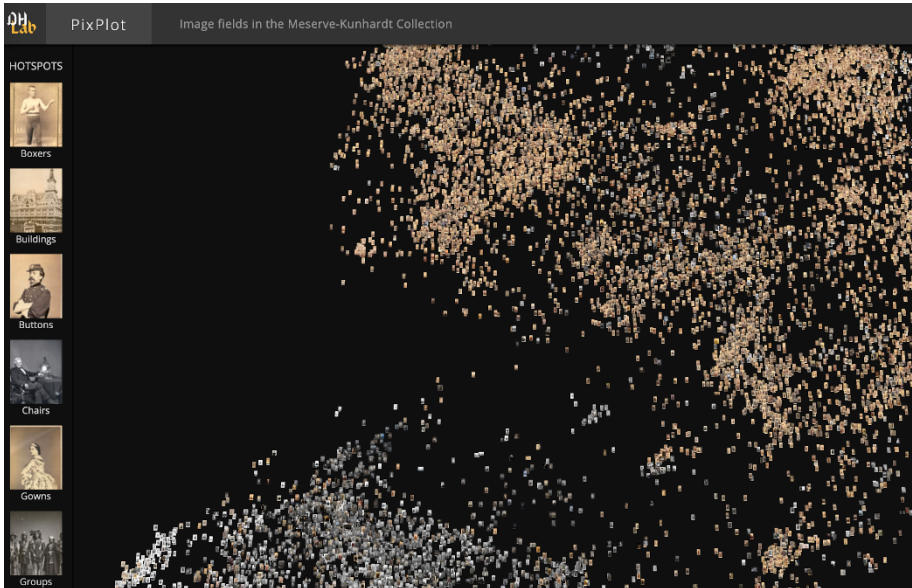


Abb. 45: DHLab, Universität Yale, PixPlot, Visualisierung der Meserve-Kunhardt Collection mit manuell annotierten Hotspots (links), Screenshot, <https://s3-us-west-2.amazonaws.com/lab-apps/pix-plot/index.html>

Das Finden von Strukturen und das Zusammenführen von ähnlichen Dokumenten, Bildern oder anderen Formaten ist eine grosse Stärke algorithmischer Lösungen, die auf *unsupervised* Lernverfahren beruhen. Häufig handelt es sich hierbei um «herkömmliches» *machine learning*, das Vektorisierungen und/oder Vergleiche vornimmt. Im Textbereich wurde oben bereits Topic Modeling (wie oben wieder im Sinne von LDA) beschrieben, ein Lernverfahren, das als ein Resultat ein **Dokumenten-Clustering** ausgibt. Umgemünzt auf Bilder können diese als Vektoren umgerechnet werden, um Ähnlichkeiten darzustellen. Als Beispiel kann hier *imgs.ai* hingewiesen werden, ein Projekt von Fabian Offert, Olga Harlamov und Peter Bell, in dem Bilder aufgrund von Ähnlichkeiten auf unterschiedlicher Basis (Pixel, Pose, etc.) verglichen werden können.¹⁰² Auf einer ähnlichen Ausgangslage beruht *PixPlot* des DHLab der Universität Yale. Die Bilder aus einer Sammlung werden in Form von Miniaturansichten aufgrund von Ähnlichkeiten in Clustern visualisiert und durch Zoomen erforschbar.¹⁰³

102 Online: <https://imgs.ai/>. Einführend zu *imgs.ai*: Oliveira, *imgs.ai*.

103 PixPlot bedient sich dem k-nearest-neighbor Grösse bei der Visualisierung. Zum Projekt: <https://dhlabs.yale.edu/projects/pixplot/>.

Der bereits oben vorgestellte VIKUS-Viewer verfügt über eine ähnliche Funktionalität, wobei t-SNE Cluster vorerzeugt und visuell umgesetzt werden.¹⁰⁴

Im Gegensatz zum Cluster handelt es sich bei der **Annotation** meistens um eine Zuordnung aus dem *supervised learning*, die entsprechend auf Trainingsdaten basiert. Das Ziel des Vorgangs ist es, zusätzliche Informationen zu einem bestimmten Objekt oder Wort zur Verfügung zu stellen. Ein klassisches Beispiel einer Annotation sind Zuschreibungen zu Wortarten (sog. *part-of-speech*) oder die Identifikation als benannte Entität (*Named Entities*, wie oben beschrieben). Auch die Identifikation von Bildinhalten bzw. Inhalten von Bildteilen wird über Annotationen hergestellt.

Die Aufgabe wird häufig und ausgesprochen erfolgreich mittels neuronalen Netzen erfüllt, die auf grossen Mengen an Trainingsdaten beruhen und bei textuellen Daten auf vortrainierte spezifische Sprachmodelle angewiesen sind. Insbesondere Sprachmodelle führen dazu, dass Aufgaben schneller erlernt werden können. Bislang stehen vor allem für moderne Sprachen mit Millionen von Sprechende entsprechende Modelle zur Verfügung, wobei die neueste Generation, die auf sogenannten Transformatoren beruht, selbst in der Lage ist, Texte zu verfassen.¹⁰⁵ Für historische Sprachformen mit weit weniger Trainingsdaten ist die Lage komplexer, da die Sprachmodelle häufig selbst erstellt werden müssen und nicht dieselbe Leistungsfähigkeit entwickeln wie grosse Sprachmodelle.

5.4 Crowdsourcing & Citizen Science: Annotation und Kuratation

Das Wissen, der Enthusiasmus und die Arbeitszeit Freiwilliger zur Erhebung, Anreicherung und Erweiterung von Informationen wird nicht nur in den Geisteswissenschaften seit Jahrzehnten genutzt. Dialektologische Erhebungen wären ohne die Mitarbeit von unzähligen Freiwilligen undenkbar und auch die Biologie setzt seit Langem darauf. Mit dem *digital turn* und der Möglichkeit, mit benutzerfreundlichen Plattformen oder gar spezifischen Apps den potenziellen Pool an Mitmachenden enorm zu steigern, versuchen sich unterschiedliche Institutionen an der Nutzung dieser Resource.¹⁰⁶

Wie bereits für maschinelle Lernverfahren und *linked data* Technologien ist Crowdsourcing und Citizen Science zur Erweiterung von Zugangsformen nur indirekt von Interesse. Der Einbezug von Freiwilligen ist jedoch aus zwei Gründen sinnvoll: Erstens ist es möglich, Expertenwissen anzuzapfen, wie es beispielsweise die Bibliothek der ETH Zürich gemacht hat, die ehemalige Mitarbeitende der Swissair eingebunden hat, um Flugzeugtypen und Flugpersonal auf Fotografien zu verschlagworten.¹⁰⁷ Zweitens können Aufgaben mit einem hohen Zeitaufwand, der von

104 Siehe bspw.: <https://vikusviewer.fh-potsdam.de/vangogh/> (Ansicht «TSNE»). Zu t-SNE siehe insbesondere: Maaten/Hinton, *Visualizing Data Using t-SNE*.

105 Siehe bspw. GPT-3, und dazu: Brown u. a., *Language Models are Few-Shot Learners*.

106 Innerhalb der Schweizer Dialektologie ist dazu die DialäktApp von Adrian Leemann et al. zu nennen: <http://www.dialaektaepp.ch>.

107 Siehe bspw. Graf, *Die Swissair reiste nach Stockholm-Bromma*.

Mitarbeitenden nicht erbracht werden könnte, an eine mehr oder weniger grosse Menge von Freiwilligen übertragen werden, wie es etwa im Editionsprojekt *Transcribe Bentham* zur Transkription der nicht ganz einfach zu lesende Handschrift von Jeremy Bentham erfolgt.¹⁰⁸

Wissenschaftlich steht die Erforschung von Crowdsourcing-Aktivitäten noch am Anfang. Obwohl der Bereich durch Forschungsförderung (siehe für DE: <https://www.buergerschaffenwissen.de/>) und Universitäten (bspw. Kompetenzzentrum der Universität und ETH Zürich: <https://citizenscience.ch/de/>) stark gefördert wird, ist die Reflexion und die empirische Auswertung von Chancen und Risiken noch nicht soweit gediehen. Die meisten Publikationen formulieren aktuell daher noch vorwiegend Versprechen und präsentieren mehr oder minder erfolgreich durchgeführte Einsatzszenarien.¹⁰⁹

Um die Zugänglichkeit zu verbessern, bieten sich innerhalb des Crowdsourcing zwei Richtungen an, die hier nur ganz kurz angesprochen werden:

9. Annotation: Darunter ist die Anreicherung mit Informationen oder Beschreibungen zu verstehen, von Transkription bis zur Verschlagwortung. Die Interessierten werden entsprechend motiviert an einer spezifischen Aufgabe mitzuarbeiten, die mehr oder weniger Fachwissen verlangt. Das Interesse einer Community kann dabei durch die Möglichkeit geweckt werden, selbst Funde zu machen (etwa bei der Aufbereitung von Texten) oder das eigene Fachwissen einzubringen (bspw. bei der Identifikation von Fahrzeugtypen).
10. Kuration: Darunter wird die Zusammenstellung und Würdigung von Beiträgen verstanden. Im Gegensatz zur Annotation wird Kuration seltener eingesetzt, da bei der Kuration eine (irgendwie gelagerte) Diskussion auf Augenhöhe stattfinden muss. Kuration steht damit auch auf einer Stufe mit Co-Design oder Participatory Design, worunter ausserdem die (Weiter-)Entwicklung von (Forschungs-)Fragestellungen verstanden wird.¹¹⁰

Ein nur unter Vorbehalten angesprochenes Thema ist die Frage nach dem sogenannten *community building*, also kurzgesagt die Rekrutierung von Freiwilligen und das Aufrechterhalten der Beziehung. Causer und Wallace haben dazu für *Transcribe Bentham* eine der wenigen Untersuchungen durchgeführt und zeigen ein nicht unerwartetes Bild: Die Teilnehmenden haben häufig einen akademischen Hintergrund (MA-Abschluss oder PhD) und gehören zur Alterskategorie zwischen 40 und 60+.¹¹¹ Nicht direkt adressiert wird der Beitrag der Freiwilligen. In *Transcribe Bentham* und in diversen anderen Projekten, ist eine kleine Gruppe von Personen für den Grossteil aller

108 *Transcribe Bentham* gilt als eines der ältesten und vor allem als erfolgreichstes Crowdsourcing Projekt im Bereich der Citizen Science: Riley, *Transcription Update – 10 December 2020* | UCL *Transcribe Bentham*.

109 Siehe bspw. Bemme/Munke, *Offene Daten und die Zukunft der Bürgerforschung in Wissenschaftlichen Bibliotheken*.

110 Zu Participatory Design siehe Pullin, *Participatory Design and the Open Source Voice*.

111 Causer/Wallace, *Building A Volunteer Community: Results and Findings from Transcribe Bentham*.

Beiträge verantwortlich. Dies führt wiederum dazu, dass diese Personen besonders «gepflegt» werden müssen und ein regelrechtes (und häufig auch zeitaufwändiges) *community building* betrieben werden muss, um ein Crowdsourcing-Projekt erfolgreich voranzubringen.

5.5 Standardisierte Publikation von Bildern via IIIF und Viewer

Das International Image Interoperability Framework (IIIF)¹¹² bietet einen Zugang zu Digitalisaten, wobei Bildgrösse, Ausrichtung und Farbschema fast beliebig manipuliert werden können. IIIF ist ein open source Framework, das sich zunehmend zum Standard für die Präsentation von Digitalisaten aus Bibliotheken, Archiven und Museen entwickelt und es Nutzenden ermöglicht, die digitalen Bilder von Objektsammlungsübergreifend zusammenzubringen. Handschriftenabbildungen (bspw. aus e-codices¹¹³) können so in bestehenden Viewern (häufig demonstriert anhand des Mirador-Viewers, daneben gibt es aber weitere IIIF-Viewer wie den OpenSeadragon oder den Universal Viewer¹¹⁴) mit Werken aus Sammlungen von Museen zusammengebracht werden.¹¹⁵ Wie bereits oben im Zusammenhang mit LOD beschrieben, bedient sich IIIF APIs, die zur Übermittlung der Bilddaten, aber auch von Metadaten oder Objektinformationen genutzt werden. Der zentrale Vorteil des Frameworks liegt im Abruf über die URL, wodurch mit wenig Kenntnissen (eigtl. wird nur eine Base-URL benötigt) Bilder manipuliert und an unterschiedlichen Orten eingebunden werden können. Auf diese Weise können Sammlungen von Erinnerungsinstitutionen, wie von Gasser gefordert, mehr bieten, als dass über eine Stichwortsuche und gebunden an die Plattform der Datensatz zu einzelnen Objekten einsehbar ist.¹¹⁶

Zur Präsentation der Vorteile von IIIF wurde als Instrument der Mirador-Viewer geschaffen. Dieser wird parallel zum IIIF-Standard weiterentwickelt und kann direkt in Websites eingebettet werden. Mit den Funktionsweisen von IIIF in Verbindung mit den IIIF-Viewern Universal Viewer, Mirador und OpenSeadragon hat sich Julien Raemy ausführlich befasst.¹¹⁷

6. Fazit

Mit der vorliegenden Studie wird aufgezeigt, dass für Sammlungen und Dokumente eine Vielzahl von visuellen und semantischen Zugangsformen bestehen, die neue und

112 <https://iiif.io/>

113 <https://www.e-codices.unifr.ch/de>

114 Ausführliche Aufstellung von IIIF Viewern: <https://iiif.io/apps-demos/#awesome-iiif>

115 Die Liste an Institutionen, die IIIF unterstützen, wächst stetig. Für eine aktuelle Übersicht: <https://iiif.io/community/#participating-institutions>

116 Gasser, Sonja: Das Digitalisat als Objekt der Begierde: Anforderungen an digitale Sammlungen für Forschung in der Digitalen Kunstgeschichte, S. 261.

117 Raemy, Julien Antoine: The International Image Interoperability Framework (IIIF): raising awareness of the user benefits for scholarly editions, Bachelor thesis, Haute école de gestion de Genève, 2017. Online: <<https://doc.rero.ch/record/306498>>, hier v. a. S. 34f. Stand: 14.12.2020.

andere Formen der Erkundung von grossen, mittleren und kleinen Korpora ermöglichen. Dabei sollte mitbedacht werden, dass die Zugangsform die Nutzung der damit verbundenen Dokumente mitprägt und somit teilweise auch einschränkt. Es lohnt sich dementsprechend, bei der Schaffung von Zugängen auf einer Webplattform eine breit angelegte Auslegeordnung zu machen und zu diskutieren, welche Recherchewege mit welcher Absicht angeboten werden sollen. Dabei gilt es zu beachten, dass unterschiedliche mediale Ausprägungen sich je nach Art besser oder schlechter für spezifische Zugangsformen eignen. Ein weiteres ausschlaggebendes Kriterium für Archive bei der Entscheidung für oder gegen einen Zugang ist die Menge an Dokumenten, Dossiers oder Instanzen, die damit verfügbar gemacht werden sollen. Denn gewisse Zugangsformen können besser mit grösseren Mengen an Objekten umgehen, während sich andere eher für kleinere Dokumentenkorpora eignen.

Die vorliegende Auslegeordnung deutet somit in zwei Richtungen: Einerseits lässt sich der Zugang auf diverses Material bereits zum heutigen Zeitpunkt mit visuellen und textuellen Lösungen stark erweitern und um Formen ergänzen, die «andere» Zugriffe auf das Material geben. So wird ein Publikum angesprochen, das bislang nur beschränkt mit Archivmaterialien in Berührung kam. Als zweiter Aspekt wird andererseits durch die Zugänge immer auch eine erste Interpretation des vorliegenden Materials mitgeliefert. Analog zur Archivtektonik, die historisch gewachsen ist und mehr über einen Verwaltungsapparat im Wandel berichtet, als über die darin enthaltenen Dokumente,¹¹⁸ deuten realisierte Zugänge auf antizipierte Inhalte hin, die gefunden werden sollen.

Diese Studie ist vor dem Hintergrund der Zugänglichmachung von Archivalien entstanden. Archivalien decken typischerweise ein ausgesprochen breites Spektrum an Themen ab und können in unterschiedlichen Formen vorliegen. Entsprechend empfiehlt sich eine Segmentierung der Inhalte nach (möglichst) kohärenten Themenblöcken, die gemäss spezifischer User-Stories durchforstet werden sollen und aufgrund des Zusammenhangs des Materials auch sinnvoll durchforstet werden können.¹¹⁹

Der vorliegende Beitrag hat eine Vielzahl von visuellen und textuellen Zugängen zusammengestellt, die aufzeigen, was aktuell in Form von Prototypen, proof-of-concept, Projektkooperationen und teilweise schon im grösseren Rahmen produktiv eingesetzt wird. Insgesamt zeugt die Vielheit einerseits von den diversen Möglichkeiten, andererseits muss aufgrund der mit den Zugängen verbundenen Komplexität die Userführung bewusst gewählt und abgewogen werden. Gleichzeitig müssen blinde Flecken, die aufgrund der Zugänge entstehen, wie etwa Ungenauigkeiten aufgrund der Nutzung von automatisierten Verfahren oder Auslassungen in visuellen Darstellungen, benannt und offen diskutiert werden.

118 Stoler, *Along the Archival Grain*.

119 Analog zur Zugriffsweise im *Impresso*-Projekt, das die Zeitungen nach Herkunftsland und nach Sprache zugänglich macht: <https://impresso-project.ch/app/>. Siehe auch die sogenannten quickaccess-Zugänge, bspw. für das Staatsarchiv Zürich: <https://archives-quickaccess.ch/stazh>.

Allen analysierten und positiv bewerteten Zugängen ist gemein, dass sie durch eine konsequente Nutzerführung überzeugen, die sich an digitalen Sehgewohnheiten orientieren. Dazu gehört, dass etwa Schriftarten gewählt werden, die im Netz gut lesbar sind und dass die Navigation über linke Spalten bzw. spezifische Menüs geschieht. Eine (evtl. reduzierte) Zugriffsmöglichkeit sollte auch über kleinere Bildschirme gewährleistet sein (getreu dem Motto «*mobile first*»). Bei der Analyse lässt sich auch ein Mix feststellen aus der einfachen Handhabung der Lösungen und der Möglichkeit, sich in Resultaten und Suchen zu vertiefen. Neben der zielgerichteten Suche wird somit auch das zufällige Auffinden von interessanten Objekten ermöglicht. Dadurch ergibt sich im besten Fall ein spielerischer Zugang zum Material, der gleichzeitig seriöse Recherchen ermöglicht.

Eine weitere Erkenntnis ist ein Wandel in den Ansprüchen an Zugangsformen. Obwohl die Studie ein spezielles Augenmerk auf innovative und neuartige Zugänge legt, war die Vielzahl an Formen und die bereits diversen produktiven Einsatzformen überraschend. Während bis um zirka 2010 das Finden entlang etablierter und antizipierter Suchzugänge ein primäres Ziel darstellte, sind in den vergangenen zehn Jahren die Ansprüche gestiegen und vor allem diverser geworden. Es wird entsprechend nach visuellen Zugangsformen verlangt, die zwar auch, aber nicht ausschliesslich, auf textuellen Informationen beruhen können, da visuelle Informationen ebenso weitergegeben werden sollen. Dazu kommt ein erhöhter Bedarf an Vernetzung mit weiteren Informationen, innerhalb, aber auch ausserhalb der Bestände einer Erinnerungsinstitution.

Aus allen diesen Punkten ergibt sich der bereits erwähnte Anspruch, das Unerwartete und das nicht explizit Gesuchte zu finden und damit die Serendipität als Überraschungsmoment in Suche und Erkundung zu ermöglichen. So kann den Nutzenden nicht nur die Breite der Überlieferung, sondern mehr noch die Konstruiertheit eines Archivbestandes – im wahrsten Sinne des Wortes – vor Augen geführt werden.

Die weitere Auseinandersetzung mit Zugangsformen wird daher in den kommenden Jahren und aufgrund der raschen Entwicklung hinsichtlich technischer Möglichkeiten, aber auch wegen der sich im Wandel befindenden Erwartungen unterschiedlicher Nutzungsgruppen, ein zentrales Thema in Archiven und allgemein dem Ökosystem der Erinnerungsinstitutionen bleiben.

Literaturverzeichnis

- Bemme, Jens; Munke, Martin: Offene Daten und die Zukunft der Bürgerforschung in Wissenschaftlichen Bibliotheken. In: Bartoschek, Thomas/Nüst, Daniel/Pesch, Mario (Hg.): Forum Citizen Science 2019. 2019, S. 27–39. DOI: 10.17605/OSF.IO/QHRC4.
- Bertin, Jacques: Graphische Semiologie. Diagramme, Netze, Karten. Berlin 1974.
- Blei, David M.: Introduction to probabilistic topic models. In: Communication of the ACM (2011).
- Blei, David M.; Ng, Andrew Y.; Jordan, Michael I.: Latent Dirichlet Allocation. In: J. Mach. Learn. Res. 3 (2003), S. 993–1022. <http://dl.acm.org/citation.cfm?id=944919.944937> (29.03.2014).

- Bludau, Mark-Jan; Brüggemann, Viktoria; Busch, Anna; Dörk, Marian: Reading Traces: Scalable Exploration in Elastic Visualizations of Cultural Heritage Data. In: *Computer Graphics Forum* 39/3 (2020), S. 77–87. DOI: 10.1111/cgf.13964.
- Brown, Tom B.; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared; Dhariwal, Prafulla; Neelakantan, Arvind; Shyam, Pranav; Sastry, Girish; Askell, Amanda; u. a.: Language Models are Few-Shot Learners. In: arXiv:2005.14165 [cs] (2020). <http://arxiv.org/abs/2005.14165> (11.09.2020).
- Brüggemann, Viktoria; Dittrich, Katja; Dörk, Marian; Herseni, Johannes; Rauenbusch, Jens; Thomet, Fidel: Die bibliografischen Daten der Deutschen Nationalbibliothek entfalten. In: *Dialog mit Bibliotheken* 30/1 (2018), S. 11–16. <https://d-nb.info/1154315215/34> (13.07.2021).
- Brüggemann, Viktoria; Bludau, Mark-Jan; Dörk, Marian: Zwischen Distanz und Nähe: Formen der Betrachtung und Bewegung in (digitalen) Sammlungen. In: *Das digitale Objekt – Zwischen Depot und Internet* 7 (2020), S. 115–123. <https://www.deutsches-museum.de/assets/Verlag/Download/Studies/studies-7-download.pdf> (13.07.2021).
- Burrell, Andrew; Hendery, Rachel; Thieberger, Nick: Glossopticon: Visualising Archival Data. In: 2019 23rd International Conference in Information Visualization – Part II. Adelaide, Australia 2019, S. 100–103. DOI: 10.1109/IV-2.2019.00029.
- Causser, Tim; Wallace, Valerie: Building A Volunteer Community: Results and Findings from Transcribe Bentham. In: *Digital Humanities Quarterly* 006/2 (2012).
- Dörk, Marian; Feng, Patrick; Collins, Christopher; Carpendale, Sheelagh: Critical InfoVis: Exploring the Politics of Visualization. In: CHI '13 Extended Abstracts on Human Factors in Computing Systems. Paris 2013, S. 2189–2198. DOI: 10.1145/2468356.2468739.
- Dörk, Marian; Pietsch, Christopher; Credico, Gabriel: One View is Not Enough: High-Level Visualizations of a Large Cultural Collection. In: *Information Design Journal* 23/1 (2017), S. 39–47. DOI: 10.1075/idj.23.1.06dor.
- Drucker, Johanna: Graphical Approaches to the Digital Humanities. In: Schreibman, Susan/Siemens, Ray/Unsworth, John (Hg.): *A New Companion to Digital Humanities*. Chichester, UK 2015, S. 238–250. DOI: 10.1002/9781118680605.ch17.
- Drucker, Johanna: *Visualization and Interpretation: Humanistic Approaches to Display*. Cambridge, Massachusetts 2020.
- Ehrmann, Maud; Romanello, Matteo; Bircher, Stefan; Clematide, Simon: Introducing the CLEF 2020 HIPE Shared Task: Named Entity Recognition and Linking on Historical Newspapers. In: Jose, Joemon M. u. a. (Hg.): *Advances in Information Retrieval*. Cham 2020, S. 524–532. DOI: 10.1007/978-3-030-45442-5_68.
- Few, Stephen: *Information Dashboard Design: The Effective Visual Communication of Data*. Beijing; Cambridge MA 2006.
- Fitzpatrick, Kathleen: Giving It Away: Sharing and the Future of Scholarly Communication. In: *Journal of Scholarly Publishing* 43/4 (2012), S. 347–362. DOI: 10.3138/jsp.43.4.347.
- Fitzpatrick, Kathleen: Scholarly Publishing in the Digital Age. In: *Between humanities and the digital*. 2015, S. 457–466.
- Fornaro, Peter: Zugang zum digitalen Archiv. In: *Arbido* 3 (2019). <https://arbido.ch/de/ausgaben-artikel/2019/zugang-acc%C3%A8s/zugang-zum-digitalen-archiv> (13.07.2021).
- Gasser, Sonja: Das Digitalisat als Objekt der Begierde: Anforderungen an digitale Sammlungen für Forschung in der Digitalen Kunstgeschichte. In: Andraschke, Udo/Wagner, Sarah (Hg.): *Objekte im Netz. Wissenschaftliche Sammlungen im digitalen Wandel*. Bielefeld 2020, S. 261–276. DOI: 10.14361/9783839455715-019.

- Glinka, Katrin; Pietsch, Christopher; Dilba, Carsten; Dörk, Marian: Linking Structure, Texture and Context in a Visualization of Historical Drawings by Frederick William IV (1795–1861). In: *International Journal for Digital Art History* 2 (2016), S. 198–213. DOI: 10.11588/DAH.2016.2.33530.
- Glinka, Katrin; Dörk, Marian: Museum im Display: Visualisierung kultureller Sammlungen (Vikus). In: Bienert, Andreas u. a. (Hg.): *Konferenzband EVA Berlin 2015. Elektronische Medien & Kunst, Kultur und Historie*. 2015, S. 18–23. <https://books.ub.uni-heidelberg.de/arhistoricum/reader/download/157/157-17-7648-1-10-20160826.pdf> (13.07.2021).
- Glinka, Katrin; Dörk, Marian: Zwischen Repräsentation und Rezeption: Visualisierung als Facette von Analyse und Argumentation in der Kunstgeschichte. In: Kuroczyński, Piotr/Bell, Peter/Dieckmann, Lisa (Hg.): *Computing Art Reader: Einführung in die digitale Kunstgeschichte*. 2018, S. 234–250. DOI: 10.11588/ARTHISTORICUM.413.C5825.
- Glinka, Katrin; Meier, Sebastian; Dörk, Marian: Visualising the »Un-seen«: Towards Critical Approaches and Strategies of Inclusion in Digital Cultural Heritage Interfaces. In: Busch, Carsten/Sieck, Jürgen (Hg.): *Kultur und Informatik: Cross Media*. Glückstadt 2015, S. 105–117. http://uclab.fh-potsdam.de/wp/wp-content/uploads/Visualising_the_Unseen_Ku115.pdf (13.07.2021).
- Glinka, Katrin; Pietsch, Christopher; Dörk, Marian: Past Visions and Reconciling Views: Visualizing Time, Texture and Themes in Cultural Collections. In: *Digital Humanities Quarterly* 11/2 (2017). <http://www.digitalhumanities.org/dhq/vol/11/2/000290/000290.html> (27.08.2017).
- Gortana, Flavio; Tenspolde, Franziska von; Guhlmann, Daniela; Dörk, Marian: Off the Grid: Visualizing a Numismatic Collection as Dynamic Piles and Streams. In: *Open Library of Humanities* 4/2 (2018), S. 30. DOI: 10.16995/olh.280.
- Gracy, Karen F.: Archival Description and Linked Data: A Preliminary Study of Opportunities and Implementation Challenges. In: *Archival Science* 15/3 (2015), S. 239–294. DOI: 10.1007/s10502-014-9216-2.
- Graf, Nicole: Die Swissair reiste nach Stockholm-Bromma. In: *ETH-Bibliothek | Crowdsourcing*. 2018. <https://blogs.ethz.ch/crowdsourcing/en/2018/10/19/die-swissair-reiste-nach-stockholm-bromma/> (11.01.2021).
- Graham, Shawn; Weingart, Scott; Milligan, Ian: Getting Started with Topic Modeling and MALLET. In: *Programming Historian* (2012). <https://programminghistorian.org/en/lessons/topic-modeling-and-mallet> (23.05.2020).
- Green, Harriett E.: Publishing Without Walls: Building a Collaboration to Support Digital Publishing at the University of Illinois. In: *Fire!!!* 3/2 (2017), S. 21–36. DOI: 10.5323/fire.3.2.0021.
- Gregory, Ian: Different Places, Different Stories: Infant Mortality Decline in England and Wales, 1851–1911. In: *Annals of the Association of American Geographers* 98/4 (2008), S. 773–794. DOI: 10.1080/00045600802224406.
- Heidmann, Frank: Interaktive Karten und Geovisualisierungen. In: Weber, Wibke/Burmester, Michael/Tille, Ralph (Hg.): *Interaktive Infografiken*. Berlin, Heidelberg 2013, S. 39–69. DOI: 10.1007/978-3-642-15453-9_3.
- Heidmann, Frank; Bludau, Mark-Jan; Dörk, Marian: Relational Perspectives as Situated Visualizations of Art Collections. Utrecht 2019. <https://dev.clariah.nl/files/dh2019/boa/0208.html> (29.12.2020).
- Hochman, Nadav; Manovich, Lev: Zooming into an Instagram City: Reading the Local through Social Media. In: *First Monday* 18/7 (2013). DOI: 10.5210/fm.v18i7.4711.
- IFLA Trend Report: Neue Technologien erweitern den Informationszugang und schränken ihn gleichzeitig ein. In: *Arbido* 3 (2019). <https://arbido.ch/de/ausgaben-artikel/2019/zugang-acc%C3%A8s/neue-technologien-erweitern-den-informationszugang-und-schr%C3%A4nken-ihn-gleichzeitig-ein> (29.12.2020).

- Kräutli, Florian: Visualising Cultural Data: Exploring Digital Collections Through Timeline Visualisations. PhD Thesis. London 2016. <https://researchonline.rca.ac.uk/1774/> (13.07.2021).
- Kuczera, Andreas: Digitale Perspektiven mediävistischer Quellenrecherche. In: *Mittelalter*. 2014. <https://mittelalter.hypotheses.org/3492> (29.12.2020).
- Kuczera, Andreas: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi. In: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*. 2015. <http://mittelalter.hypotheses.org/5995> (11.03.2016).
- Leemann, Adrian; Kolly, Marie-José; Wanitsch, Daniel; Werlen, Iwar; Hasler, Fiona; Eisenblatt, Carina: *DialäktÄpp*. 2010. <http://www.dialaektaepp.ch> (29.01.2021).
- Maaten, Laurens van der; Hinton, Geoffrey: Visualizing Data Using t-SNE. In: *Journal of Machine Learning Research* 9/86 (2008), S. 2579–2605. <http://jmlr.org/papers/v9/vandermaaten08a.html> (28.12.2020).
- Maceli, Monica G.; Yu, Kerry: Usability Evaluation of an Open-Source Environmental Monitoring Data Dashboard for Archivists. In: *Archival Science* 20/4 (2020), S. 347–360. DOI: 10.1007/s10502-020-09340-1.
- Manovich, Lev: How to Compare One Million Images? In: Berry, David M. (Hg.): *Understanding Digital Humanities*. London 2012, S. 249–278. DOI: 10.1057/9780230371934_14.
- Manovich, Lev: Exploring Urban Social Media: Selficity and On Broadway. In: Kitchin, Rob/Perng, Sung-Yueh (Hg.): *Code and the City*. London 2016. <http://manovich.net/index.php/projects/urbansocialmedia> (13.07.2021).
- Marsh, Diana E.; Punzalan, Ricardo L.; Leopold, Robert; Butler, Brian; Petrozzi, Massimo: Stories of Impact: The Role of Narrative in Understanding the Value and Impact of Digital Collections. In: *Archival Science* 16/4 (2016), S. 327–372. DOI: 10.1007/s10502-015-9253-5.
- Mayr, Eva; Windhager, Florian: Vor welchem Hintergrund und mit Bezug auf was? Zur polykontextualen Visualisierung kultureller Sammlungen. In: Andraschke, Udo/Wagner, Sarah (Hg.): *Objekte im Netz. Wissenschaftliche Sammlungen im digitalen Wandel*. Bielefeld 2020, S. 235–245. DOI: 10.14361/9783839455715.
- Merzagli, Michele: Informationen finden und Wissen verlinken: Der Weg der Metadatenstandards vom Archivregal zu den Linked Data. In: *Arbido* 3 (2017). <https://arbido.ch/de/ausgaben-artikel/2017/metadaten-datenqualit%C3%A4t/vom-regal-zum-word-wide-web-die-entwicklung-von-normen-und-standards> (28.12.2020).
- Merzagli, Michele: Neue Standards erarbeiten und Informationsnetze bauen: Verbesserung von Dienstleistungen in Archiven durch Normen und Standards. In: *Arbido* 4 (2019). <https://arbido.ch/de/ausgaben-artikel/2019/servicequalit%C3%A4t-qualit%C3%A4tsservice/neue-standards-erarbeiten-und-informationsnetze-bauen> (13.07.2021).
- Oliveira, Simona: *imgs.ai: Die Suche durch Bilder*. In: *Einblicke in die Digital Humanities*. 2020. <https://einblicke.hypotheses.org/170> (28.12.2020).
- Pullin, Graham: Participatory Design and the Open Source Voice. In: Sterne, Jonathan u. a. (Hg.): *The Participatory Condition in the Digital Age*. Minneapolis, Minnesota ; 2016, S. 101–122.
- Raemy, Julien Antoine: *The International Image Interoperability Framework (IIIF): Raising Awareness of the User Benefits for Scholarly Editions*. Bachelor thesis. Genève 2017. <https://doc.rero.ch/record/306498> (14.12.2020).
- Riley, Chris: *Transcription Update – 10 December 2020 | UCL Transcribe Bentham*. In: *Transcribe Bentham*. 2020. <https://blogs.ucl.ac.uk/transcribe-bentham/2020/12/10/transcription-update-10-december-2020/> (11.01.2021).

- Schneider, Andreas: Die Datenflut mit intelligenten Algorithmen bändigen. In: *Arbido 2* (2020). <https://arbido.ch/de/ausgaben-artikel/2020/abbild-und-inszenierung-der-gesellschaft/die-datenflut-mit-intelligenten-algorithmen-b%C3%A4ndigen> (13.07.2021).
- Schöch, Christof; Döhl, Frédéric; Rettinger, Achim; Gius, Evelyn; Trilcke, Peer; Leinen, Peter; Jannidis, Fotis; Hinzmann, Maria; Röpke, Jörg: Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen. In: *Zeitschrift für digitale Geisteswissenschaften 5* (2020). DOI: 10.17175/2020_006.
- Schwan, Hannah; Jacke, Janina; Kleymann, Rabea; Stange, Jan-Erik; Dörk, Marian: Narrelations – Visualizing Narrative Levels and their Correlations with Temporal Phenomena. In: *Digital Humanities Quarterly 13/3* (2019). <http://www.digitalhumanities.org/dhq/vol/13/3/000414/000414.html> (15.01.2021).
- Stoler, Ann Laura: *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*. Princeton, NJ Oxford 2010.
- Vane, Olivia: *Timeline Design for Visualising Cultural Heritage Data*. PhD Thesis. London 2019. <https://researchonline.rca.ac.uk/4325/> (13.07.2021).
- Weingart, Scott: Demystifying Networks, Parts I & II. In: *Journal of Digital Humanities 1/1* (2011). <http://journalofdigitalhumanities.org/1-1/demystifying-networks-by-scott-weingart/> (11.03.2016).
- Weingart, Scott: Topic Modeling for Humanists: A Guided Tour. <http://www.scottbot.net/HIAL/?p=19113> (26.07.2012).
- Whitelaw, Mitchell: Generous Interfaces for Digital Cultural Collections. In: *Digital Humanities Quarterly 9/1* (2015). <http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html> (20.08.2017).
- Wildi, Tobias; Dubois, Alain: Le Matterhorn RDF Data Model: Description archivistique et Linked Open Data. Vers quelle convergence? 2019. https://www.slideshare.net/TobiasWildi/alain-dubois-tobias-wildi-le-matterhorn-rdf-data-model-description-archivistique-et-linked-open-data-vers-quelle-convergence-presentation-au-forum-des-archivistes-5-avril-2019-saintetienne?from_m_app=ios (28.12.2020).
- Windhager, Florian; Glinka, Katrin; Mayr, Eva; Schreder, Günther; Dörk, Marian: Zur Weiterentwicklung des “cognition support”: Sammlungsvisualisierungen als Austragungsort kritisch-kulturwissenschaftlicher Forschung. In: Vogeler, Georg (Hg.): *DHd 2018 Kritik der digitalen Vernunft*. Köln 2018, S. 341–345. DOI: 10.5281/zenodo.3684897.
- Windhager, Florian; Federico, Paolo; Schreder, Günther; Glinka, Katrin; Dörk, Marian; Miksch, Silvia; Mayr, Eva: Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges. In: *IEEE Transactions on Visualization and Computer Graphics 25/6* (2019), S. 2311–2330. DOI: 10.1109/TVCG.2018.2830759.
- ISAD(G): General International Standard Archival Description – Second edition | International Council on Archives. <https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition> (18.12.2020).
- ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, 2nd Edition | International Council on Archives. <https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd> (18.12.2020).