

Christopher Pollin

# Datenvisualisierung

Handbuch Repositorienmanagement, Hg. v. Blumesberger et al., 2024, S. 503–520  
<https://doi.org/10.25364/978390337423227>



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz,  
ausgenommen von dieser Lizenz sind Abbildungen, Screenshots und Logos.

Christopher Pollin, Universität Graz, Zentrum für Informationsmodellierung, christopher.pollin@uni-graz.at |  
ORCID iD: 0000-0002-4879-129X

## Zusammenfassung

Daten- bzw. Informationsvisualisierung verfolgt den Zweck, visuelle Darstellungen von Datensätzen zu generieren, die dabei helfen sollen, Aufgaben effizienter zu erledigen. Dieser Artikel gibt einen grundlegenden Überblick über dieses Themenfeld. Dies beinhaltet das Referenzmodell der Informationsvisualisierung nach Card et al., die unterschiedlichen Funktionen und Klassifikationen von Visualisierungen, sowie praxisnahe Beispiele und Werkzeuge, die Potenziale und Herausforderungen des Themengebietes veranschaulichen. Eines dieser Beispiele ist das Visualisierungsprojekt *Poppy Field* von Valentina D'Efilippo. Abschließend wird das Thema im Kontext von Repositorien eingeordnet.

**Schlagwörter:** Datenvisualisierung; Informationsvisualisierung

## Abstract

### Data Visualization

The purpose of data or information visualization is to create visual representations of data sets to help to carry out tasks more effectively. This contribution gives a basic overview of this topic. It includes the reference model of information visualization according to Card et al., the different features and classifications of visualizations, as well as practical examples and tools that illustrate the potentials and challenges of the domain. One of these examples is the visualization project *Poppy Field* by Valentina D'Efilippo. Finally, the topic is put into the context of repositories.

**Keywords:** Data visualization; information visualization

## 1. Einleitung

Datenrepositorien speichern große Datenmengen und stellen sie den Fachgemeinschaften zur Verfügung. Zu viele und zu dichte Informationsmengen führen zu einem *Cognitive Overload*, also einer kognitiven Überlastung von Benutzer:innen. Eine Möglichkeit, sich in größeren Datenmengen dennoch zurecht zu finden, ist die Daten- bzw. Informationsvisualisierung.

Eine Metauntersuchung von Alejandro Benito und Roberto Therón<sup>1</sup> liefert einen empirischen Überblick über die Schnittmenge in der Forschung rund um Informationsvisualisierung und digitale Geisteswissenschaften. Es zeigt die vielfältigen Anwendungsbereiche von Visualisierung für dieses Fach, die sich von Mensch-Maschinen-Interaktion, Text-, Bild- und Netzwerkanalyse bis hin zu Semantic Web und Datenbanken, sowie dem Management von Forschung erstrecken. Natürlich lassen sich vergleichbare Überschneidungen zwischen der Disziplin der Informationsvisualisierung mit anderen Disziplinen aus den Natur-, Rechts und Wirtschaftswissenschaften festmachen. Überall dort, wo Daten im größeren Ausmaß generiert werden und es den Bedarf gibt, damit arbeiten zu wollen, entsteht fast eine Notwendigkeit visueller Darstellungsformen dieser Daten, um die Kognition von Benutzer:innen zu unterstützen. Da Datenrepositorien einen Mittler zwischen Forschungsdaten und Forschung darstellen, ist es nicht unerheblich, Datenvizualisierung im Zusammenhang mit Datenrepositorien zu diskutieren.

Tamara Munzner definiert Visualisierung im Kontext von Informationssystemen, wie etwa Datenrepositorien, als “visual representations of datasets designed to help people carry out tasks more effectively.”<sup>2</sup> Ausgehend von dieser Definition stellt sie drei zentrale Fragen bei der Umsetzung von Visualisierungen:

- Warum wird eine Visualisierung benötigt („Why“)?
- Welche Daten werden dargestellt („What“)?
- Und wie werden sie dargestellt („How“)?

Der vorliegende Text bietet einen praxisorientierten Einstieg in das Thema der Datenvizualisierung und beginnt mit der Darstellung eines Beispiels, das sehr gut geeignet ist, um Grundbegriffe und Best Practices der Datenvizualisierung zu erörtern.<sup>3</sup> Am Beispiel der *Poppy Field – Visualising War Fatalities*, einer interaktiven Datenvizualisierung, werden die grundlegenden theoretischen Rahmenbedingungen, sowie die Begrifflichkeiten des Themenkomplexes erörtert und definiert.

---

1 Benito, A.; Therón, R. (2020), S. 45-57.

2 Munzner, T. (2014), S. 1.

3 Reiterer, H.; Jetter, H.-C. (2013), S. 295-306.

## **2. Poppy Field – Visualising War Fatalities: Ein Beispiel**

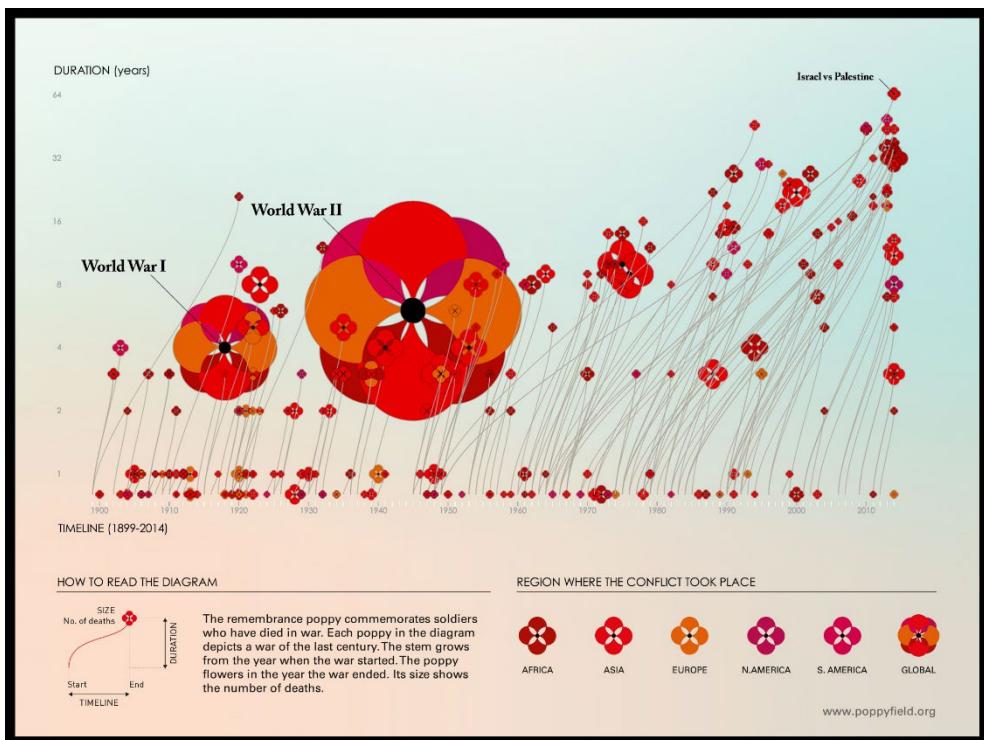
Im Projekt “Poppy Field - Visualising War Fatalities”<sup>4</sup> werden Daten des Polynational War Memorial visualisiert. Diese umfassen die Dauer, Anzahl von Opfern, beteiligte Kontinente und Nationen militärischer Konflikte im Zeitraum von 1900 bis 2010. Die Notwendigkeit dieser Visualisierung entsteht nicht nur, weil die visuelle Repräsentation der Daten die Arbeit damit erleichtert, also die Datenexploration ermöglicht, sondern auch, weil damit Geschichten erzählt werden können. Unter “Data Storytelling” wird ein strukturierter Ansatz zur Vermittlung von Erkenntnissen aus Daten verstanden, der aus der Kombination dreier Schlüsselemente besteht: Daten („What“), visuelle Darstellungen („How“) und Erzählungen („Why“).<sup>5</sup> Für Valentina D’Efilippo<sup>6</sup> ist im Projekt “Poppy Field” eben ein Zweck der Datenvisualisierung, neben der Exploration der Daten, das Erzählen einer Geschichte. Das Ergebnis des Projektes ist in Abbildung 1 ersichtlich. Mohnblumen, seit dem Ersten Weltkrieg ein Symbol des Gedenkens, repräsentieren jeweils einen militärischen Konflikt. Je größer eine Mohnblume, desto höher ist die Zahl der Opfer, wobei die unterschiedlichen Blütenfarben die Beteiligung der Kontinente aufzeigen. Aus dem Startpunkt auf der Zeitleiste und dem Endpunkt, definiert jeweils als Punkt auf einem Feld aus Datum (x-Achse) und Kriegsdauer (y-Achse), ergibt sich ein gebogener Stängel. Das Ergebnis ist ein Informationsraum in Form eines schematischen Mohnblumenfelds, der eine Vielzahl an Datenpunkten unterschiedlichen Typs kodiert und miteinander in Verbindung setzt. Etwas, das mit einer reinen Auflistung der Daten in einer Liste oder Tabelle die kognitiven Fähigkeiten des Menschen überfordern würde.

---

<sup>4</sup> D’Efilippo, V.; Pigelet, N. (2018)

<sup>5</sup> Weber, W. (2020), S. 295-311.

<sup>6</sup> D’Efilippo, V.; Ball, J. (2013), S. 100f.



**Abbildung 1:** Valentina D'Efilippo, Nicolas Pigelet: *Poppy Field – Visualising War Fatalities*, URL: <https://www.poppyfield.org/fig00/PoppyField.jpg>.

Beim Betrachten der Abbildung sind zwei Dinge sofort erkennbar: zwei dominierende Mohnblumen in der ersten Hälfte – die beiden Weltkriege – und eine Häufung von langstängeligen, kleineren Blumen in der zweiten Hälfte, die eine Vielzahl langandauernder, aber eben kleinerer Konflikte zeigt, wie etwa den israelisch-palästinensischen Konflikt rechts oben. Sowohl die Datenexploration als auch eine ästhetische und dem Thema angemessene Vermittlung, die besonders dem “Data Storytelling” geschuldet ist, sind so gegeben. Auch eine weitere wichtige Dimension der Datenvisualisierung kann damit gezeigt werden: die Interaktion. Erst durch die interaktiven Möglichkeiten einer Webanwendung kann eine Überflutung an Information, die beispielsweise durch Beschriftung jedes einzelnen Konfliktes anfallen würde, vermieden werden. Vorbildlich wird das “Information Visualisation Man-

tra” von Ben Shneiderman umgesetzt: „Overview, zoom and filter, details on demand“.<sup>7</sup> So gibt das Mohnblumenfeld zu Beginn einen Überblick über den gesamten Datenbestand, aus dem sich Verteilung, Entwicklung oder Ausreißer auf einen Blick erfassen lassen. Über die Regler in der Timeline kann in bestimmte Zeiträume hineingezoomt werden, um etwa nur Konflikte zwischen 1945 und 1950 untersuchen zu können. Dies lässt sich mit der Filterfunktion nach der Quantität der Opfer kombinieren, sodass User:innen die für sie relevanten Informationsräume aus dem Datenbestand heraus erzeugen können. Nun haben User:innen die Möglichkeit, in dem so selektierten Datenbestand zusätzliche Details, wie etwa eine Liste der beteiligten Nationen, zu erhalten, indem sie auf die einzelnen Mohnblumen klicken.

### **3. Begriffe, Definitionen und Klassifikationen**

Der Themenkomplex Visualisierung geht mit einer Fülle von Begrifflichkeiten einher und ist stark gekoppelt an die Begriffstrias Daten-Information-Wissen.<sup>8</sup> Auf oberster Ebene ist die Datenvisualisierung – oft als Synonym für Informationsvisualisierung verwendet – abgrenzbar zur Wissensvisualisierung zu verstehen. Ziel der Wissensvisualisierung, vielmehr ein Teilgebiet der Wissensmodellierung, ist es, die Verwendung visueller Darstellungen zur Verbesserung der Übertragung von Wissen zwischen mindestens zwei Personen zu erforschen.<sup>9</sup> Datenvisualisierung aber umfasst die Anwendung computerbasierter, interaktiver, visueller, externer Repräsentationen von abstrakten Daten mit dem Ziel, menschliche Kognition zu erweitern. Die Datenvisualisierung eröffnet einen Kommunikationsweg für abstrakte Daten, wohingegen die Wissensvisualisierung die Übertragung von Wissen fokussiert.

Die Datenvisualisierung kann nun weiter aufgeteilt werden in Scientific Visualization“ (SciVis) und Information Visualization (InfoVis)<sup>10</sup> und teilweise auch in eine eigene dritte Kategorie der Geographic Visualization (GeoVis).<sup>11</sup> Diese drei Kategorien unterscheiden sich in der Wahl der räumlichen Repräsentation der Daten – der Projektion. Für eine SciVis ist die räumliche Position mit dem Datensatz gegeben. Das heißt beispielsweise, dass ein menschliches Gehirn den Raum darstellt, auf dem bestimmte Gehirnareale eingefärbt werden, um gemessene Gehirnaktivitäten – also Daten – abzubilden. In der InfoVis bzw. in der GeoVis wird der Raum zur

---

7 Shneiderman, B. (1996), S. 336-343.

8 Eine ausführliche Diskussion dieser Begriffe im Kontext der Informationswissenschaft findet sich bei Favre-Bulle, B. (2001) und bei Rowley, J. E. (2007), S. 163-180.

9 Meyer, R. (2010), S. 23.

10 Nazemi, K. et al. (2021), S. 477f.

11 Andrews, K. (2020)

Datenrepräsentation erst geschaffen. Hier muss erst ein geeigneter Raum generiert werden, um anhand von Daten eine bestimmte Aufgabe effektiv darstellen und bearbeiten zu können. Für InfoVis heißt das, dass diagrammatische Darstellungen gewählt und Daten etwa in einem Liniendiagramm dargestellt werden. Ein Beispiel hierfür ist auch das Projekt *Poppy Field*, da es sich einer Komposition von visuellen Strukturen bedient. Für GeoVis ist das ähnlich, wobei der Raum durch geografische Karten, wie etwa Datensätze auf einer Weltkugel oder Landkarte, gegeben ist.

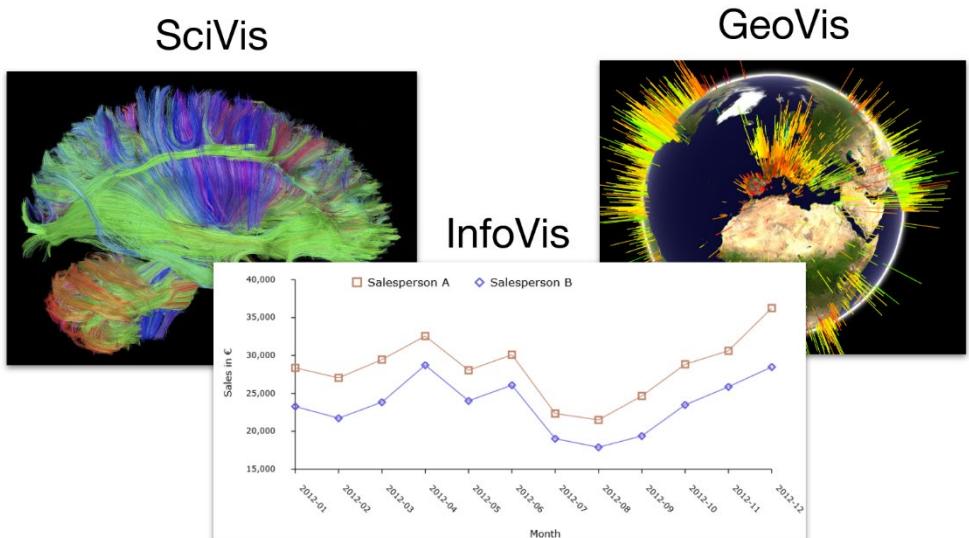


Abbildung 2: Eigene Darstellung: SciVis, InfoVis, GeoVis

### 3.1. Klassifikation von Datenvisualisierungen

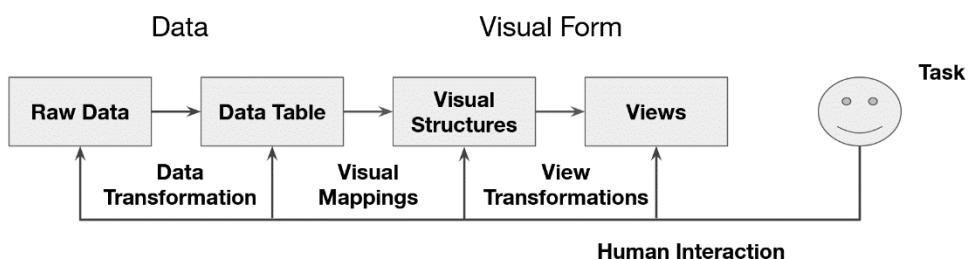
Klassifiziert werden Datenvisualisierungen nach ihrer Lesart und ihrer Komplexität. Die Lesart einer Visualisierung kann author-driven oder reader-driven sein. Als author-driven werden Visualisierungen bezeichnet, die einen angeleiteten, das heißt einen erklärenden Charakter, haben. Ziel ist es, eine Aussage sichtbar und nachvollziehbar für Leser:innen zu machen. Das Data Storytelling hat genau diese Funktion und wird auch als „Explanatory-Data Visualisation“ oder auch „Narrative-Information Visualisation“ bezeichnet. Dem gegenüber steht die Gruppe der „Exploratory-Data Visualisation“, also der Visualisierungen, die reader-driven sind. Diese Klasse an Visualisierungen ist nicht angeleitet und folgt einem explorativen Paradigma, in dem Leser:innen über Funktionalitäten unterschiedliche Views – also Perspektiven auf Daten – selbstständig erzeugen können. Indem über die Filterfunktionen im

*Poppy Field Project* der Informationsraum angepasst werden kann, werden von User:innen neue Views generiert und der Datenbestand exploriert. Das *Poppy Field Project* kann aber durchaus auch als hybride Lösung mit explorativen, aber auch erklärenden Elementen verstanden werden.<sup>12</sup>

Komplexität als *Klassifikationskriterium* bezieht sich auf die Anzahl der Datendimensionen. Darunter versteht man die Anzahl der diskreten Informationsarten, die in einem Diagramm visuell kodiert sind. Im anfangs angeführten Beispiel der Mohnblumen gibt es fünf Datendimensionen: die Zahl der Opfer, Start- und Endzeitpunkt, sowie Kontinent und Nation. Bei einem Liniendiagramm, das die Entwicklung eines Preises über die Zeit hinweg zeigt, reichen zwei Dimensionen für Preis und Datum. Die Komplexität nimmt also mit der Anzahl der Dimensionen zu, nicht aber mit der Anzahl der Datensätze. Eine Visualisierung wird also nicht komplexer, wenn statt 100 Konflikten 1.000 dargestellt werden. Mit höherer Komplexität wird es auch herausfordernder, die Daten und ihre Abhängigkeiten effizient für User:innen nachvollziehbar zu machen, da auch die Anzahl der visuellen Kodierungen zunimmt.<sup>13</sup>

### 3.2. Referenzmodell der Informationsvisualisierung

Zentrale Elemente der Informationsvisualisierung sind Daten, die Wahl der visuellen Kodierung bzw. visuellen Strukturen und der Views. Diese Begriffe, die zum Teil schon erwähnt wurden, lassen sich gut anhand des *Referenzmodells der Informationsvisualisierung* nach Card et al.<sup>14</sup> definieren.



**Abbildung 3: Eigene Darstellung: Referenzmodell der Informationsvisualisierung nach Card et al.**

Den Ausgangspunkt bilden rohe Daten. Im Prozess der Data Transformation werden rohe Daten in strukturierte Daten, die Data Tables, überführt. Das heißt, dass

<sup>12</sup> Weber, W. (2020), S. 299-301.

<sup>13</sup> Iliinsky N.; Steele, J. (2011)

<sup>14</sup> Card, S. et al. (1999), S. 12-25.

Zahlenwerte und Text ihre Bedeutung durch ein Modell bzw. ein Schema erhalten. Die auf rein syntaktischer Ebene befindlichen Daten werden kontextualisiert und so zu Information. Die Zahl „1945“ ist nicht mehr eine Zeichenkette, sondern das Enddatum eines Konfliktes. Auf dem Weg hin zur Visualisierung müssen die strukturierten Daten nun auf visuelle Strukturen gemappt werden. Die Aufgabe des *Visual Mappings* ist es, die Daten, die in Data Tables strukturiert sind, den verschiedenen visuellen Ausprägungen der verfügbaren Visual Structures zuzuordnen. Diese visuellen Strukturen sind die Archetypen der eigentlichen Visualisierungen und definieren die grundsätzliche Art der Informationsdarstellung: Balkendiagramme, Netzwerke oder Treemaps. Das Visual Mapping bestimmt, welche Datenwerte welchen visuellen Variablen der visuellen Strukturen zugeordnet werden. Visuelle Variablen sind beispielsweise Position, Höhe, Breite, Farbe oder Form. Im Beispiel aus Abbildung 1 wird der quantitative Wert der Opferzahl in Form des Durchmessers der Mohnblume kodiert. Die Dauer des Konflikts wird als geschwungene Linie zwischen zwei Punkten in einem Koordinatensystem kodiert und Farben repräsentieren die unterschiedlichen Kontinente. Die Visual Structure entspricht so einem Punktdiagramm, das den Anforderungen der Daten entsprechend adaptiert wurde.

Wie bereits erwähnt, entstehen die Views durch die Interaktivität der User:innen, die in der Lage sind, den Informationsraum ihren Anforderungen entsprechend anzupassen. Das Hineinzoomen in die Jahre 1945 bis 1960 ist ein solcher View, wie auch das gefilterte Ergebnis nach allen Konflikten, die am asiatischen und afrikanischen Kontinent stattfanden. Genau diese Anforderungen sind wichtig und auch im Referenzmodell mit der Human Interaction und der Task der User:innen abgebildet. An sämtlichen Prozessen, angefangen vom Informationsbedürfnis, über die Datenerzeugung und -modellierung, bis zum Design und der Implementierung sind Akteure beteiligt. Tamara Munzner fasst dies als „Human in the Loop“<sup>15</sup> zusammen. Viele Probleme und Möglichkeiten werden erst klar, wenn der Mensch sich in iterativen Abläufen damit beschäftigt. Das ist notwendig, um in der großen Menge an möglichen Visualisierungen, die aus einem Datenbestand heraus erzeugt werden können, auch die zu finden, die für die Anwender:innen einen Mehrwert generieren.

---

15 Munzner, T. (2014), S. 2-4.

### 3.3. Beschränkungen

Hier gilt es, drei Beschränkungen in der Entwicklung von Visualisierungen zu berücksichtigen: die menschliche Wahrnehmung und Kognition, das Medium zur Darstellung und die algorithmische Verarbeitung durch Computer.

Der Mensch ist gut darin, Muster zu erkennen. Innerhalb einer Menge von gleichen Objekten erkennen Menschen sofort Ausreißer, wenn sie eine andere Farbe, Form, Position oder Größe haben. Im Gegensatz dazu ist es schwieriger, sich wahrgenommene Bilder zu merken und aus der Erinnerung mit anderen zu vergleichen. Ebenso stellt die potenzielle kognitive Überladung ein Problem dar. Diese ist dann gegeben, wenn zu viel Information gleichzeitig auf die User:innen einwirkt, also die Informationsdichte zu hoch ist.

Ein zu viel an Information entsteht ab einem gewissen Punkt, da auch der verfügbare Raum auf einem Bildschirm oder Blatt Papier, also dem Medium der Visualisierung, begrenzt ist. Es können nicht alle Perspektiven auf einen Datensatz gleichzeitig abgebildet werden. Durch Interaktivität, beispielsweise durch eine Weboberfläche, kann das Prinzip der Multiperspektivität umgesetzt werden. Es gilt die Faustregel, pro Visualisierung nur eine Aussage zu kommunizieren, es können aber im Sinne der Multiple Views<sup>16</sup> durch Interaktivität unterschiedliche Visualisierungstypen genutzt werden, um andere Perspektiven auf Daten zu erzeugen.

Die dritte Beschränkung ist schlichtweg dadurch gegeben, dass aus Daten Visualisierungen berechnet werden müssen. Je komplexer die Berechnungen sind, um eine Visualisierung aus einem Datenbestand zu erzeugen, desto länger kann sie dauern oder gar nicht in sinnvoller Zeit generiert werden.<sup>17</sup>

## 4. Data und Data Tables

Deskriptive Statistik und Datenvisualisierung sind eng miteinander verknüpft. Wo die deskriptive Statistik versucht, empirische Daten quantitativ zu beschreiben, ist es die Datenvisualisierung, die diese Ergebnisse darstellt.<sup>18</sup> Grundlage sind aber stets, wie aus dem Referenzmodell von Card et. al hervorgeht, strukturierte Daten. Abbildung 4 zeigt exemplarische Datensätze, sowie alle Dimensionen (Spalten) aus dem “Poppy Field Project”.<sup>19</sup>

---

<sup>16</sup> Windhager, F. (2019), S. 4-6.

<sup>17</sup> Munzner, T. (2014), S. 9-16.

<sup>18</sup> Shardt, Y.; Weiß, H. (2021), S. 4-30.

<sup>19</sup> Die verwendeten Daten für die Webanwendung des “Poppy Field Project” im CSV-Format : <https://www.poppyfield.org/data/PoppyDataCSV.csv>

wars	from	to	duration notes	participation	number_who	where	fatalities	source
War in Eastern Ukraine	2014	2014	0 This armed conflict in Ukraine	2 Europe, Asia	Europe	4035	War in Donbas	
Sectarian Conflict	2012	2014	2 End year and from Central African Republic	1 Africa	Africa	2099	Sectarian Conflict	
South Sudan Civil War	2011	2014	3 End year and from South Sudan	1 Africa	Africa	15000	South Sudan Civil War	
Syrian Civil War	2011	2014	3 End year and from Syria	1 Asia	Asia	260215	Syrian civil war	
Yemen vs Al-Qaida	2009	2014	5 This conflict is between Yemen, United States and Al-Qaida	2 Asia, North America	Asia	4270	Yemen vs Al-Qaida	
Nigerian Govt v Boko Haram	2009	2014	5 This conflict is between Nigeria and Boko Haram	1 Africa	Africa	4627	Nigerian Govt v Boko Haram	
Waziristan conflict	2007	2014	7 This conflict is between Pakistan, Afghanistan and Taliban	2 Asia	Asia	23494	Waziristan conflict	
Kivu Conflict	2006	2013	7 Democratic Republic of Congo and FARDC	2 Africa	Africa	10105	Kivu Conflict	
Iraqi Insurgency	2003	2014	11 Start year, ended in Iraq, United States	5 Asia, North America	Asia	184512	Iraqi Insurgency	

**Abbildung 4: Eigene Darstellung: Einträge des Datensatzes aus dem Poppy Field Project**

Diese „Attribute“, wie sie auch genannt werden, können unterschiedlich konfiguriert sein. Die Dimensionen „wars“, „notes“, „participation“, „who“, „where“ und „source“ sind Text, wohingegen „from“, „to“, „duration“, „number\_participants“ und „fatalities“ Zahlen beinhalten. Diese vorerst banale Erkenntnis führt aber dazu, dass nicht jeder Wert für eine Datavisualisierung gleich verwendet werden kann. Während die Zahlen aus den „fatalities“ sinnvollerweise aufaddiert werden können, bleibt dies bei „from“ und „to“ – zwei Datumsangaben – sinnbefreit. Auch bei den textuellen Dimensionen erkennt man sofort, dass sich „where“ von „wars“ unterscheidet. Erstes gibt jedem Konflikt seinen eindeutigen Namen, zweiteres beinhaltet dieselben Begriffe mehrmals. Folglich gibt es unterschiedliche Typen von Attributen. Im Allgemeinen werden nominale bzw. kategoriale, ordinale und quantitative Attribute unterschieden.<sup>20</sup> Ein Beispiel eines quantitativen Attributs ist „fatalities“, eines nominalen bzw. kategorialen ist „where“ und die Datumsangaben können als ordinale Attribute verstanden werden. Sie sind durch folgende Charakteristika unterscheidbar: mit quantitativen Attributen können arithmetische und statistische Operationen durchgeführt werden. Ordinale und kategoriale bzw. nominale Attribute haben einen einordnenden bzw. gruppierenden Charakter. Erstere lassen sich nach dem Prinzip „etwas ist größer als etwas anderes“ sortieren: Das Jahr 1945 ist vor dem Jahr 1950. Mit Zweiteren kann zwar sinnvoll gruppiert werden, aber ohne, dass es eine Struktur oder Ordnung impliziert. Eine Ordnung im Sinne von „Afrika ist höherwertiger als Europa“ ist nicht möglich.<sup>21</sup>

20 Munzner, T. (2014), S. 31-33; Ware, C. (2019)

21 Nazemi, K. et al. (2021), S. 480-482.

## 5. Visual Mapping

Der Prozess des Visual Mappings regelt ausgehend von strukturierten Daten, welche Attribute an welcher Stelle positioniert werden sollen, um eine Visualisierung zur Bearbeitung einer Frage zu erzielen.<sup>22</sup> Eine kommerzielle Softwarelösung zur Generierung von Datenvisualisierungen, die sehr gut diesen Mappingprozess in einem User Interface darstellt, ist Tableau.<sup>23</sup> Ein weiteres und einfacheres, aber freies Tool ist RAWGraphs.<sup>24</sup> In beiden Softwarelösungen kann die Tabelle aus den Poppy Fields als CSV oder Excel geladen und Datendimensionen können auf bestimmte visuelle Strukturen der unterschiedlichen Diagrammtypen (Charts) gemappt werden. Aus der Fülle von unterschiedlichen Visualisierungen – der *The Data Visualisation Catalogue*<sup>25</sup> oder das *Data Viz Project*<sup>26</sup> geben einen guten Überblick – werden ein Säulendiagramm (Bar Chart), eine Kastengrafik (Box Plot) und ein Streudiagramm (Scatter Plot) kurz vorgestellt. Eine weitere Möglichkeit, aus Daten Visualisierungen zu erzeugen, sind Programmiersprachen. Sehr häufig werden Python<sup>27</sup> und R<sup>28</sup> verwendet, um Daten auszuwerten und zu visualisieren, oder Bibliotheken wie D3.js<sup>29</sup>, um interaktive Visualisierungen im Web zu implementieren. Letzteres fand auch Verwendung im *Poppy Field Project*.

Zur Erzeugung eines Bar Chart reicht es, zwei Attribute miteinander zu verknüpfen, die auf der x- und y-Achse projiziert werden. Die Zeilen werden nach einem kategorialen Attribut gruppiert, wie etwa „where“, und die Summe der „fatalities“ jeder Gruppe wird ermittelt. Die Kategorien aus der Dimension „where“ kommen auf die x-Achse und die Summen der „fatalities“ auf die y-Achse. Das Bar Chart ist gut geeignet, um einen Überblick über oder einen Vergleich von Daten zu ermöglichen, und es gilt als eine der einfachsten Visualisierungsformen.

---

22 Nazemi, K. et al. (2021), S. 484-487.

23 Tableau <https://www.tableau.com>. Für den akademischen Bereich gibt es eine einjährige Testversion.

24 RAWGraphs <https://rawgraphs.io>

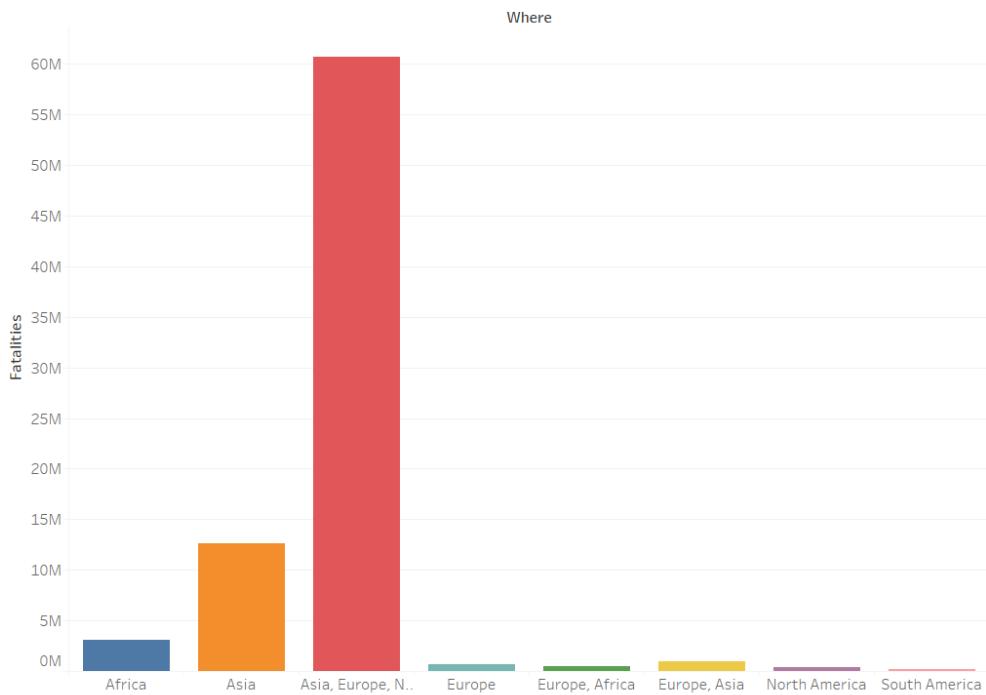
25 The Data Visualisation Catalogue <https://datavizcatalogue.com>

26 Data Viz Project <https://datavizproject.com>

27 Zu Libraries in diesem Zusammenhang: Siehe Tanner, G. (2019).

28 Kabacoff, R. (2020): Data Visualization with R <https://rkabacoff.github.io/datavis>

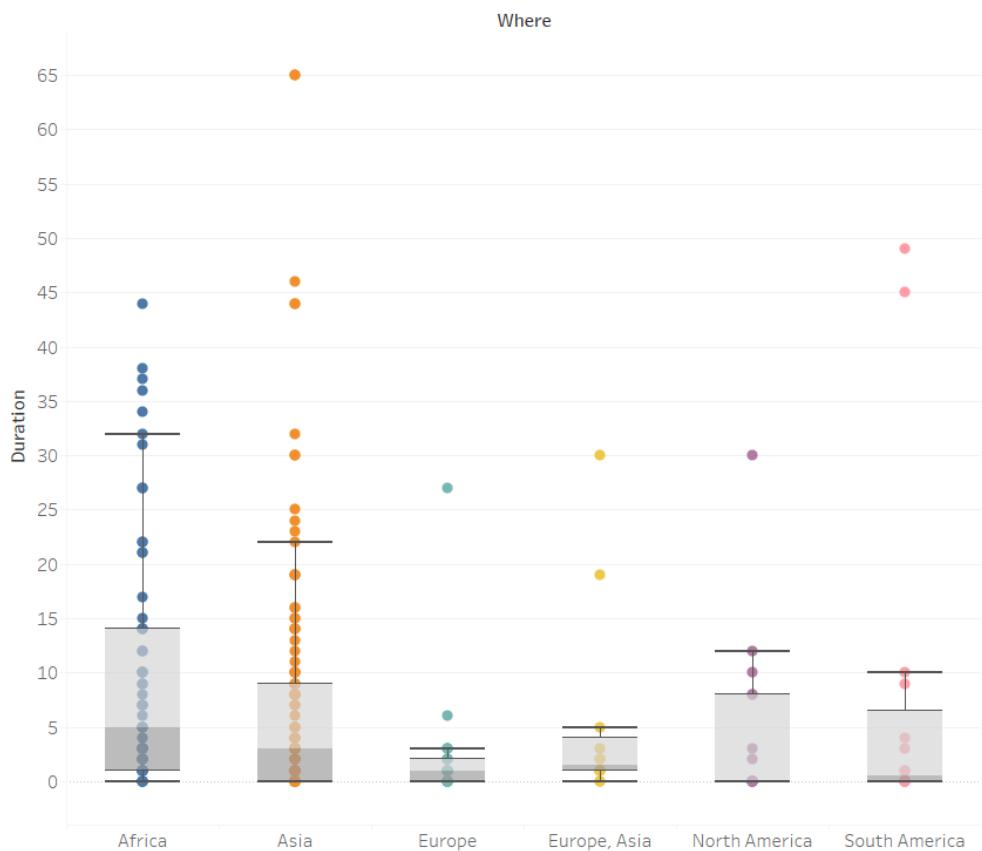
29 Bostock, M. (n. d.): Data-Driven Documents <https://d3js.org>



**Abbildung 5: Eigene Darstellung. Bar Chart der Poppy Field-Daten umgesetzt mit Tableau**

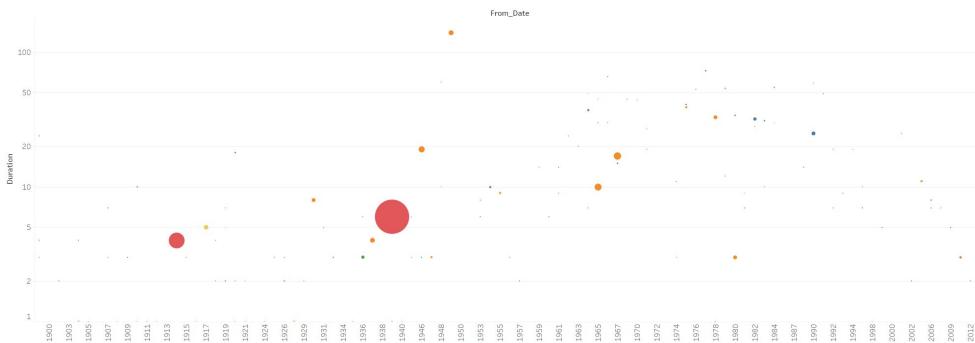
Ein Box Plot veranschaulicht aggregierte statistische Werte eines quantitativen Attributes, wie etwa Median, Quartil oder Ausreißer<sup>30</sup> und ist gut geeignet, um die Verteilung von Daten darzustellen. In Abbildung 5 wird die Dauer der Konflikte auf die y-Achse gelegt und die x-Achse beschreibt das Attribut „where“. Weiters werden die vorher genannten statistischen Mittel auch sichtbar gemacht. Jeder Punkt repräsentiert einen Konflikt und desto höher der Wert auf der y-Achse ist, desto länger ist die Dauer des Konflikts. Gleichzeitig erkennen wir, dass es in Afrika eine Häufung von Konflikten gibt, die zwischen 1 und 14 Jahren dauern, sowie einige Ausreißer nach oben hin, d. h. deren Dauer länger ist.

<sup>30</sup> Munzner, T. (2014), S. 308-310.



**Abbildung 6: Eigene Darstellung. Box Plot der *Poppy Field*-Daten umgesetzt mit Tableau**

Ein Scatter Plot ähnelt am stärksten dem *Poppy Field*. Das Koordinatensystem wird durch das Beginndatum und die Dauer des Konfliktes gebildet. Jeder Punkt repräsentiert einen Konflikt, wobei die Größe der Kreise die Anzahl der Opfer wiedergibt und die farbliche Kodierung die regionale Verortung. Was aber auch ersichtlich wird, sind die Grenzen der Darstellung auf nur einer Seite. Sowohl die Anzahl der Opfer, als auch die Dauer der Konflikte liefert eine recht große Bandbreite und sorgt dafür, dass Konflikte nur als kleine Punkte im Vergleich der beiden Weltkriege dargestellt werden können. Selbiges gilt auch für die, die das Koordinatensystem nach oben hinstreckt. Eine Möglichkeit ist, statt einer linearen eine logarithmische Darstellung zu wählen, um so die y-Achse zu stauchen, aber dennoch die Verhältnismäßigkeit beizubehalten.



**Abbildung 7: Eigene Darstellung. Scatter Plot der *Poppy Field*-Daten umgesetzt mit Tableau**

## 6. Zusammenfassung

Clifford Wulfman fasst entsprechend zusammen: “It is important for creators and readers of these depictions to remember that they are not ‘data’ but readings, interpretations of data meditated by programmed algorithms and hermeneutic desires.”<sup>31</sup>

Wo Datenbestände in ihrer Zusammenschau die kognitiven Fähigkeiten des Menschen überfordern, dort gibt es Potenzial für Datenvisualisierungen. Diese Verfahren der visuellen Analyse und der visuellen Kommunikation erlauben neue Wege zur Analyse und Exploration von Daten- und Quellenmaterial sowie deren Nutzung, um Erkenntnisse weiterzuvermitteln.<sup>32</sup> Datenvisualisierungen sind, so Clifford Wulfman im Zitat zu Beginn der Zusammenfassung, nicht ein fertiges Produkt, sondern eine computerbasierte Form der Kommunikation, die es erst zu interpretieren gilt. Gerade in der geisteswissenschaftlichen Forschung ist multiperspektivischer Diskurs zu einem Forschungsfeld die Norm. Nicht eine Visualisierung ist die einzige korrekte, sondern unterschiedliche Visualisierungsformen können für die Arbeit der Geisteswissenschaftler:innen sinnstiftend sein. Die Visualisierung an sich ist kein Erkenntnisgewinn, sondern erst die Arbeit damit und die Kritik daran. Genauso wichtig in diesem Zusammenhang ist auch, die nicht unmittelbar sichtbaren

31 Wulfman, C. E. (2014), S. 94-109.

32 Windhager, F. (2019), S. 1-2.

Ebenen solcher Diskurse zu berücksichtigen<sup>33</sup>: Wie können beispielsweise Unschärfe, Fragmentierung und Unsicherheiten historischer Information in Visualisierungen berücksichtigt werden? Florian Windhager veranschaulicht dies im Kontext biographischer und historischer Daten. Er sieht einen besonderen Mehrwert in der Darstellung räumlich-zeitlicher und kontextsensitiver Information in einem PolyCube, um die Bewegung von historischen Akteur:innen in Raum und Zeit darzustellen.<sup>34</sup> Weitere Projekte, die neue Darstellungsformen erproben, werden von der interdisziplinären Forschungsgruppe Urban Complexity Lab (UCLAB) der Fachhochschule Potsdam umgesetzt.<sup>35</sup> So wird im Projekt *Reading Traces Visualizing Fontane's Reference Library* erforscht, welche explorativen Wege der Wissensdarstellung im Digitalen für einen Bibliotheksbestand möglich sind.<sup>36</sup>

Visualisierungen sind kein Selbstzweck, sondern haben funktionalen Charakter. Die Darstellungen werden nicht (nur) der schönen Bilder wegen gemacht, wie auch die Generierung davon kein triviales Unterfangen ist. Herausforderungen erstrecken sich ausgehend von den Daten und den Modellen („What“), den Anforderungen, warum man sie benötigt („Why“), und der Frage, wie Daten auf visuelle Strukturen gemappt werden können („How“). Das „How“ beinhaltet auch die Paradigmen der Multiple Views und Interaktivität, aus denen heraus unterschiedliche Views von User:innen erzeugt werden.

Die FAIR Kriterien werden zu Recht im Kontext von Datenrepositorien hochgehalten. Daten müssen gefunden werden können und zugänglich, interoperabel und wiederverwendbar sein. Datenvisualisierungen können bei den ersten zwei Kriterien helfen. Nämlich dann, wenn das Finden von gesuchten Objekten aufgrund der schieren Masse nicht mehr möglich ist. Ein gutes Beispiel dafür ist der Wikidata Query Service, der es erlaubt, mittels SPARQL den gesamten Wikidata-Datenbestand abzufragen. Neben einer tabellarischen Darstellung der Suchtreffer können, wenn es die gefundenen Attribute erlauben, unterschiedliche Visualisierungen, wie Bar Chart oder Treemap selektiert werden, um Datenbankabfragen zu visualisieren. Datenvisualisierungen helfen also, Trends, Ausreißer, Beziehungen oder Muster in Daten sichtbar zu machen. Sie helfen dort, wo die kognitiven Fähigkeiten der Anwender:innen überfordert sind. Sie sind keine Erkenntnis per se, sondern helfen dabei, neue Erkenntnisse zu gewinnen, neue Perspektiven einzunehmen, oder eben auch neue Geschichten zu erzählen.

33 Graham, E. (2017), S. 449f.

34 Windhager, F. (2019), S. 1-2.

35 Projekte des Urban Complexity Lab (UCLAB) <https://uclab.fh-potsdam.de/projects>

36 Bludau, M.-J. et al. (2020), S. 77-87 und Reading traces Visualizing Fontane's reference library <https://uclab.fh-potsdam.de/projects/reading-traces>

## Bibliografie

- Andrews, Keith (2020): Information Visualisation. Course Notes. Graz: University of Technology Graz. <https://courses.isds.tugraz.at/ivis/ivis.pdf> (abgerufen am 16.08.2021)
- Benito, Alejandro; Therón, Roberto (2020): A Data-Driven Introduction to Authors, Readings, and Techniques in Visualization for the Digital Humanities. In: IEEE Computer Graphics and Applications 40, pp. 45-57. <https://doi.org/10.1109/MCG.2020.2973945>
- Bludau, Mark-Jan et al. (2020): Reading Traces. Scalable Exploration in Elastic Visualizations of Cultural Heritage Data. In: Computer Graphics Forum 39 (3), pp. 77-87.
- Card, Stuard; Mackinlay, Jock; Shneiderman, Ben (1999): Readings in Information Visualization. Using Vision to Think. San Francisco: Morgan Kaufmann.
- D'Efilippo, Valentina; Ball, James (2013): The Infographic History of the World. Richmond Hill: Firefly Books.
- D'Efilippo, Valentina; Pigelet, Nicolas (2018): Poppy Field – Visualising War Fatalities. <https://www.poppyfield.org> (abgerufen am 24.09.2021)
- Drucker, Johanna (2020): Visualization and Interpretation. Humanistic Approaches to Display. Cambridge: Massachusetts Institute of Technology.
- Engebretsen, Martin; Kennedy, Helen (eds.) (2020): Data Visualization in Society. Amsterdam: University Press Amsterdam. <https://doi.org/10.1515/9789048543137>
- Favre-Bulle, Bernard (2001): Information und Zusammenhang. Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Wien; New York: Springer.
- Graham, Elyse (2017): Introduction: Data Visualisation and the Humanities. In: English Studies 98 (5), pp. 449-458. <https://doi.org/10.1080/0013838X.2017.1332021>
- Iliinsky, Noah; Steele, Julie (2011): Designing Data Visualizations. Representing Information Relationships. Sebastopol: O'Reilly Media. <https://www.oreilly.com/library/view/designing-data-visualizations/9781449314774/ch01.html> (abgerufen am 24.09.2021)
- Meyer, Robert (2010): Knowledge Visualization. In: Baur, Dominikus et. al. (eds.): Trends in Information Visualization. Hauptseminar Medieninformatik WS 2008/2009. An Overview of Current Trends, Development and Research in Information Visualization. Technical Report. München: LMU Munich. Department of Computer Science. Media Informatics Group, pp. 23-30.
- Munzner, Tamara (2014): Visualization Analysis and Design. New York: A K Peters/CRC Press.
- Nazemi, Kawa; Kaupp, Lukas; Burkhardt, Dirk; Below, Nicola (2021): 5.4 Datavisualisierung. In: Putnings, Markus; Neuroth, Heike; Neumann, Janna (Hg.): Praxishandbuch Forschungsdatenmanagement. Berlin, Boston: De Gruyter Saur, S. 477-502. <https://doi.org/10.1515/9783110657807-026>
- Rehbein, Malte (2017): Informationsvisualisierung. In: Jannidis, Fotis; Kohle, Hubertus; Rehbein, Malte (Hg.): Digital Humanities. Stuttgart: J. B. Metzler, S. 328-342. [https://doi.org/10.1007/978-3-476-05446-3\\_23](https://doi.org/10.1007/978-3-476-05446-3_23)

- Reiterer, Harald; Jetter, Hans-Christian (2013): Informationsvisualisierung. In: Kuhlen, Rainer; Semar, Wolfgang; Strauch, Dietmar (Hg.): Grundlagen der praktischen Information und Dokumentation. Berlin et al.: De Gruyter Saur, S. 192-206.  
<https://doi.org/10.1515/9783110258264.192>
- Rowley, Jennifer E. (2007): The Wisdom Hierarchy. Representations of the DIKW Hierarchy. In: Journal of Information Science 33 (2), pp.163-180.
- Shardt, Yuri; Weiß, Heiko (2021): Methoden der Statistik und Prozessanalyse. Eine anwendungsorientierte Einführung. Berlin, Heidelberg: Springer Vieweg.  
<https://doi.org/10.1007/978-3-662-61626-0>
- Shneiderman, Ben (1996): The Eyes Have It. A Task by Data Type Taxonomy for Information Visualizations. In: Proceedings of the IEEE Symposium on Visual Languages. Washington: IEEE Computer Society Press, pp. 336-343.
- Tanner, G. (2019): Introduction to Data Visualization in Python. How to Make Graphs Using Matplotlib, Pandas and Seaborn. In: Towards Data Science 23.01.2019. <https://towardsdatascience.com/introduction-to-data-visualization-in-python-89a54c97fbed> (abgerufen am 24.09.2021)
- Ware, Colin (2020): Information Visualization. Perception for Design. 4th edition. Burlington: Morgan Kaufmann.
- Weber, Wiebke (2020): Exploring Narrativity in Data Visualization in Journalism. In: Engebretsen, Martin; Kennedy, Helen (eds.): Data Visualization in Society. Amsterdam: Amsterdam University Press, pp. 299-301. <https://doi.org/10.1515/9789048543137-022>
- Windhager, Florian (2019): Choreographien der Existenz. Zur multimodalen Erweiterung biographischer Forschung und Lehre durch Verfahren der visuellen Analyse und Synthese. In: BIOS – Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen 30 (1-2), S. 60-75.
- Wulfman, Clifford E. (2014): The Plot of the Plot. Graphs and Visualizations. In: Journal of Modern Periodical Studies 5 (1) (Special Issue Visualizing Periodical Networks), pp. 94-109. <http://www.jstor.org/stable/10.5325/jmodeperistud.5.1.0094>

**Christopher Pollin** ist wissenschaftlicher Mitarbeiter am Institut Zentrum für Informationsmodellierung der Universität Graz. Er promoviert in Digital Humanities und beschäftigt sich mit semantischen Webtechnologien, digitalen Editionen und Informationsvisualisierung. 2022 gründete er das DH/IT-Unternehmen Digital Humanities Craft.