

Leben, Werke und Datensilos Zur Verknüpfung und Visualisierung von im/materiellen Komponenten des kulturellen Erbes

Mayr, Eva

eva.mayr@donau-uni.ac.at
Universität für Weiterbildung Krems, Austria

Liem, Johannes

johannes.liem@donau-uni.ac.at
Universität für Weiterbildung Krems, Austria

High-Steskal, Nicole

nicole.high-steskal@donau-uni.ac.at
Universität für Weiterbildung Krems, Austria

Grebe, Anja

anja.grebe@donau-uni.ac.at
Universität für Weiterbildung Krems, Austria

Windhager, Florian

florian.windhager@donau-uni.ac.at
Universität für Weiterbildung Krems, Austria

Hintergrund

In den letzten Jahren wurde die Digitalisierung der Objektsammlungen von zahlreichen Kulturerbe-Institutionen vorangetrieben. Materielle Kulturgüter aus europäischen Museen, Archiven und Bibliotheken sind als Digitalisate in großem Umfang auf transnationalen Plattformen wie Europeana.eu einer breiten Öffentlichkeit und der Wissenschaft zugänglich gemacht worden. Gleichzeitig, aber davon unabhängig, wurde immaterielles Kulturerbe – wie biografisches Wissen über bedeutende nationale Persönlichkeiten – digital erfasst und als strukturierte und verknüpfte Aggregate in Biographie-Datenbanken verfügbar gemacht. Diese Entwicklungen bieten eine gute Basis für eine digital vermittelte Rezeption, Analyse und Kommunikation von historischen Beständen zum kulturellen Erbe. Jedoch verhindern fehlende Verknüpfungen (zwischen den biographischen Datenbanken und Europeana) und Standardisierungen (zwischen den biographischen Datenbanken der verschiedenen Nationen) sowie mangelnde (Maschinen-)Lesbarkeit und Sichtbarkeit lokaler Datensammlungen oft eine optimale Nutzung – für die wissenschaftliche Analyse durch Expert*innen ebenso wie für ein besseres Verständnis von kulturhistorischen Themen für die interessierten Öffentlichkeit und eine leichtere Exploration von materiellem gemeinsam mit immateriellem Kulturerbe.

Das H2020-Projekt InTaVia (*In/Tangible Cultural Heritage: Visual Analysis, Communication and Curation*, <https://intavia.eu>)

will solche Barrieren reduzieren, materielles und immaterielles Kulturerbe digital zusammenführen und eine synoptische Betrachtung von Leben und Werken der europäischen Kulturgeschichte ermöglichen. Das Konsortium harmonisiert zu diesem Zweck nationale Kulturdatenbestände und entwickelt ein prototypisches Informationsportal für die visuelle Analyse, Kuratierung und Kommunikation von hybriden (i.e. im/materiellen) Kulturdaten auf multiplen Ebenen der Aggregation.

Konzept

Visualisierungen erlauben die Gewinnung von Überblicken und Einsichten in voluminöse und komplexe Datenbestände, indem sie z.B. enthaltene zeitliche, geographische, relationale oder kategoriale Muster, Verteilungen und Zusammenhänge sichtbar machen. Dadurch ergänzen Visualisierungen zur Option der direkten Kontemplation und Interpretation einzelner Objekte distante Blicktechniken auf Sammlungen des kulturellen Erbes und ermöglichen deren offene Exploration – für wissenschaftliche Expert*innen genauso wie für die interessierte Öffentlichkeit (Windhager, Federico et al. 2018).

Der Fokus von Visualisierungen des kulturellen Erbes lag dabei bisher meist auf **materiellen Objekten** in den Feldern von Bildern und Skulpturen (z.B. Glinka et al. 2016), Texten (Jänicke et al. 2017), oder von Dokumenten der performativen Künste (z.B. Khulusi et al. 2020). Durch die Öffnung von musealen Objektdatenbanken und Verknüpfung mit komplementären kulturellen Datenbeständen kann jedoch ein reichhaltigeres Verständnis von Objekten in diversen interpretativen und kontextuellen Horizonten ermöglicht werden (Mayr & Windhager 2020). So kann etwa biographisches Wissen über das Leben von Künstler*innen dabei helfen Muster in einer Visualisierung ihrer Objektsammlungen besser zu verstehen (Mayr et al. 2019).

Auf der **immateriellen** Seite des Kulturerbes fokussiert InTaVia auf **biographische Narrative**, die als lexikalische Texte erfasst und in den letzten Jahren mithilfe von NLP Verfahren in strukturierte Ereignis-Daten überführt und als prosopographische Datenbanken aggregiert wurden (Fokkens et al. 2014; Schlögl & Lejtovicz 2017). Diese Transformation von biographischen Texten in maschinenlesbare Formate eröffnet auch neue Möglichkeiten für deren visuelle Repräsentation (Windhager, Schlögl et al. 2018).

Über Personenreferenzen (z.B. die gemeinsamen Normdatei GND) lassen sich materielle Objektdaten (z.B. aus Europeana, aber auch aus lokalen Datenbanken einzelner Kultureinrichtungen) mit dem Wissen über “immaterielle” biographische Ereignisketten und die darüber verknüpften Akteure (z.B. Ersteller, Auftraggeber, Besitzer, oder assoziierte Organisationen) verknüpfen. Die Entwicklung von synoptischen Visualisierungsmethoden, mit denen Datenkonstellationen von “Leben und Werk” neu analysiert und kommuniziert werden können, bildet ein Hauptziel des Projekts InTaVia. Dass hiermit die klassische Grundidee der “biographischen Kritik” und Interpretation (vgl. Vasari 1550/2008; oder Johnson 1781/1868) relativ spät ihrer ersten digitalen und visuellen Remedialisierung zugeführt wird, schmälert leider keine der assoziierten informationstechnischen (Figure 1) und gestalterischen Herausforderungen (Windhager 2020), insbesondere unter den Vorzeichen einer benutzerzentrierten Designansatzes (Mayr et al. 2018).

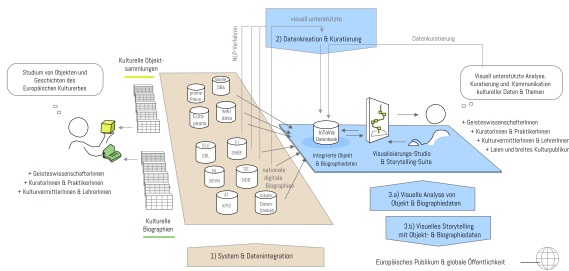


Abb. 1: Architektur der InTaVia-Plattform

Partizipative und prototypische Entwicklungen

In zwei partizipativen Design-Workshops mit 41 internationalen Expert*innen aus verschiedenen Feldern des Kulturerbes – mit wissenschaftlichem, genauso wie mit praktischem Hintergrund – wurde die konzeptuelle Architektur der geplanten InTaVia Informationsplattform diskutiert und auf ihre Passung mit den Bedürfnissen und Prozessen der potenziellen Nutzer*innen hin evaluiert. Aus den reichhaltigen Ergebnissen dieses Prozesses, die die Grundlage für die weiteren technischen Entwicklungen in InTaVia bilden, waren drei von besonderer Relevanz.

Trotz der bereits initial geplanten reichhaltigen Datenbasis – welche die über 50 Mio. Objekte der Europeana mit Beständen von vier nationalen Biographiedatenbanken aus Österreich, Slowenien, den Niederlanden und Finnland verknüpft – formulierten viele Workshopeteilnehmer*innen den Bedarf nach Optionen der Einbringung und Einbindung von eigenen Objekt- oder Biographie-Datenbeständen. Ebenso wurden Befürchtungen dokumentiert, dass Fehler in den probabilistisch extrahierten Daten ebenso wie widersprüchliche Einträge in verschiedenen Datenbanken (z.B. weil eine Person in mehreren Biographieportalen mit unterschiedlichen Daten zu finden ist) zu Verzerrungen in der Datenbasis und deren Visualisierung führen könnten.

Daraus wurde die Notwendigkeit einer benutzerseitigen **Datenkuratierung** abgeleitet, die den künftigen Benutzer*innen die Möglichkeit gibt, eigene Daten einzuspeisen und mittels NLP zu strukturieren, die Ergebnisse von NLP-Prozessen zu korrigieren, fehlende Verknüpfungen zwischen Daten herzustellen oder fehlerhafte Verknüpfungen zu korrigieren und bestehende Quellenkonflikte aufzulösen oder sie als solche auch im Rahmen der Visualisierung explizit sichtbar zu machen.

In den präsentierten Möglichkeiten der visuellen Analyse und Kommunikation sehen die Teilnehmer*innen ein hohes Potenzial für ihre berufliche Arbeit mit kulturellen Objekten und Akteuren. Betont wurde auch der Bedarf nach einem flexiblen Tool, das sich an aktuelle Daten, Themen und Fragestellungen anpassen lässt, und in dem Nutzer*innen verschiedene Ebenen und Informationen nach Bedarf ein- und ausblenden können, um etwa die thematische und visuelle Komplexität an diverse Zielgruppen der Kulturvermittlung anpassen zu können.

Visualisierungen werden im Projekt an mehreren Stellen eine Rolle spielen: (1) Zunächst sollen Benutzer*innen ohne technisches Vorwissen beim Verständnis der NLP Prozesse und bei der Datenkuratierung durch Visualisierungen unterstützt werden (z.B. indem unterschiedliche extrahierte Entitäten mit ihren Wahr-

scheinlichkeiten visualisiert werden, anstatt ausschließlich der wahrscheinlichsten Ergebnisse). (2) Interaktive Visualisierung erlauben die Exploration und Analyse der Daten in verschiedenen Kombinationen (Objekte / Objekte und Akteure / Akteure) und Aggregationszuständen (von einzelnen Individuen bis hin zu verknüpften Datenkonstellationen wie Institutionen oder Regionen) entlang verschiedener Datendimensionen (wie Zeit, Raum, Relationen oder Mengen, vgl. Figure. 2).

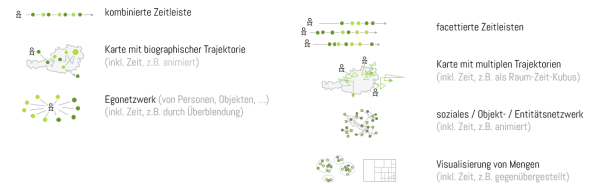


Abb. 2: Verschiedene Optionen, um Objekte (hellgrün) synoptisch mit Biographiedaten (dunkelgrün) zu visualisieren: von Einzelakteuren (links) hin zu Personengruppen (rechts)

Für jede dieser Datendimensionen sind in der Folge diverse Aggregationen denkbar. So können Informationen von einzelnen Individuen und Objekten ausgehend zu signifikant größeren Datenkonstellationen aggregiert werden – wie zum Beispiel Gruppen, kulturellen Institutionen oder raum-zeitlichen Regionen (Figure 3, unten). Der essentiellen Bedeutung der zeitlichen Orientierung dieser Daten wird zudem durch multiple Enkodierungen Rechnung getragen, um Nutzer*innen verschiedene analytische Perspektiven mit komplementären Stärken und Schwächen zu offerieren.

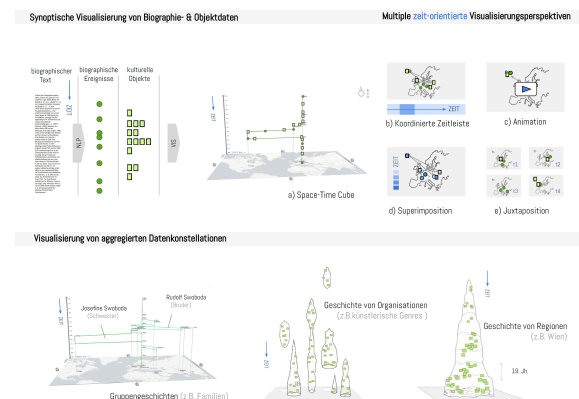


Abb. 3: Verschiedene Optionen der Visualisierung von Zeit (oben) und der Aggregation von Objekt- und Biographiedaten in der geographischen Raumzeit (unten).

(3) Schließlich können auch *narrative Visualisierungen* die Kommunikation der Ergebnisse von solchen visuellen Analysen unterstützen: Dazu werden in einem Editor narrative Techniken mit Visualisierungen kombiniert und mit audiovisuellem Material und Texten angereichert. Durch die so ermöglichte Kombination von narrativ-sequentieller Präsentation der Informationen und freier Exploration der Visualisierung, können diese narrativen Visualisierungen auch der interessierten Öffentlichkeit einen Zugang zu der geplanten Datenbasis eröffnen.

Von den Berufsprofilen und Bedürfnissen der Teilnehmer*innen wurden in einem iterativen Prozess 10 **Personas** als prototypische Benutzer*innen von DH Expert*innen bis hin zu interessierten Laien abgeleitet, die Richtlinien für die weitere technischen Entwicklung in InTaVia enthalten: Vom fachlichen und technischen Vorwissen über zu erwartende Nutzungs- und Aktivitätsprofile sowie wichtige Einschränkungen (z.B. rechtliche Bedenken, aber auch bestimmte technische Bedürfnisse) sollen diese Personas sicherstellen, dass die zukünftigen Benutzer auch zwischen den geplanten Evaluationszyklen in die Entwicklung der DH-Technologien einfließen. In die Ausarbeitung der entsprechenden Profile flossen neben der Teilnehmer*innen-Analyse auch User Stories der NFDI4Culture-Initiative (2019) ein, ebenso wie Lernerfahrungen aus gescheiterten DH Projekten (Dombrowski 2014, 2019).

Ausblick

Das Projekt InTaVia befindet sich noch in einem frühen Stadium, dennoch erlauben die hier präsentierten Workshopergebnisse bereits die Ableitung von Nutzer*innenbedürfnissen und Anforderungen an DH Technologien, die auch für andere Projekte mit heterogener Nutzer*innen-Basis in Forschungs- und Arbeitsfeldern zu verlinkten Daten von Nutzen sein können. Die entwickelten Personas wurden auf der Projekt-Website zugänglich gemacht und stehen auch anderen Wissenschaftler*innen zur Verfügung.

Die Zusammenführung von Biographie- und Objektdaten in einem synoptischen Wissensgraphen und in einem korrespondierenden Visualisierungssystem, ermöglicht es, auch über alternative Formen der Objektmodellierung und -visualisierung nachzudenken: Bisher wurden digitale kulturelle Objekte oft nur mit statischen Metadaten verknüpft, jedoch lassen sich diese in InTaVia auch als geschichtliche Objekte mit veränderlichen Eigenschaften (z.B. dynamischen Klassifikationen, Standorten, oder Personenrelationen) modellieren und entsprechende Objektbiographien visualisieren. Eine solche Modellierung erhöht zwar die Anforderungen an die vorhandenen Digitalisierungsprozesse in Kulturinstitutionen, andererseits eröffnet sie auch neue Möglichkeiten für die DH-gestützte Forschung beispielsweise zur Provenienz von Objekten.

Ein weiterer Punkt von zentraler Relevanz in einem DH-Projekt ist die Frage der Datenqualität und von Unsicherheiten in den Daten als Herausforderung für die Modellierung und Visualisierung von kulturellen Daten (Windhager et al. 2019). Der Austausch im Rahmen von verschiedenen Workshops innerhalb des Konsortiums und mit den potenziellen Nutzer*innen zeigte, dass die Qualität der Daten, die Vielfalt der Einflüsse auf selbige und die Transparenz von Unsicherheiten in den Daten wichtige Faktoren für das Vertrauen der Benutzer*innen in die entwickelten Technologien sind. So ist es unerlässlich, dass im Projekt, zumindest jene Einflüsse auf die Datenqualität, die den gewählten Prozessen und Methoden geschuldet sind (z.B. Verknüpfung mehrerer Datenbanken, Anwendung von NLP Prozessen) auch transparent kommuniziert und visualisiert werden.

Danksagung. Das Projekt InTaVia wird von der Europäischen Kommission im Rahmen des H2020 Research and Innovation Programme, Grant Agreement No. 101004825 gefördert. Wir möchten uns für die Unterstützung durch das Projektkonsortium (Vrije Universiteit Amsterdam, NL, Research Centre of the Slovenian Academy of Sciences & Arts, SI, Aalto University, FI, University of Southern Denmark, DK, Austrian Academy of Sciences, AT,

University of Stuttgart, DE, Fluxguide, AT, University of Helsinki, FI) und die Teilnehmer*innen der Co-Design-Workshops herzlich bedanken.

Bibliographie

- Dombrowski, Q.** (2014): "What Ever Happened to Project Bamboo?" in: *Literary & Linguistic Computing* 29: 326–339. <https://doi.org/10.1093/lilc/fqu026>
- Dombrowski, Q.** (2019): *Towards a Taxonomy of Failure*. <http://quinndombrowski.com/?q=blog/2019/01/30/towards-taxonomy-failure>
- Fokkens, A. / Ter Braake, S. / Ockeloen, N. / Vossen, P. / Legêne, S. / Schreiber, G.** (2014): „BiographyNet: Methodological Issues when NLP supports historical research.“ in: *LREC*, 3728–3735.
- Glinka, K. / Pietsch, C. / Dilba, C. / Dörk, M.** (2016): "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). <https://doi.org/10.11588/dah.2016.2.33530>
- Jänicke, S. / Franzini, G. / Cheema, M. F. / Scheuermann, G.** (2017): „Visual text analysis in digital humanities.“ in: *Computer Graphics Forum* 36/6: 226–250.
- Johnson, S.** (1868): *Lives of the Most Eminent English Poets: With Critical Observations on Their Works*. (Neuaufgabe, Original veröffentlicht um 1781). AT Crocker.
- Khulusi, R. / Kusnick, J. / Meinecke, C. / Gillmann, C. / Focht, J. / Jänicke, S.** (2020). A survey on visualizations for musical data. In: *Computer Graphics Forum* 39/6: 82–110.
- Mayr E. / Salisu, S. / Filipov, V. / Schreder, G. / Leite, R. / Miksch, S. / Windhager F.** (2019): „Visualizing biographical trajectories by historical artifacts: A case study based on the photography collection of Charles W. Cushman.“ in: *Workshop on Biographical Data in a Digital World 2019*, Varna, Bulgaria.
- Mayr, E. / Schreder, G. / Windhager, F.** (2018): „Digital Humanities - Eine benutzerzentrierte Perspektive.“ in: *Digital Humanities im deutschsprachigen Raum* (DHd) 2018, Köln.
- Mayr, E. / Windhager, F.** (2020): „Vor welchem Hintergrund und mit Bezug auf was? Zur polykontextualen Visualisierung kultureller Sammlungen.“ In: U. Andraschke / S. Wagner, (Hrsg.): *Objekte im Netz. Wissenschaftliche Sammlungen im digitalen Wandel*, Bielefeld: transcript, 235–245.
- NFDI4Culture** (2019): *NFDI4Culture User Stories*. https://nfdi4culture.de/fileadmin/files/NFDI4Culture_UserStoryAll.pdf
- Schlögl, M. / Lejtovicz, K.** (2017): "APIS-Eine Linked Open Data basierte Datamining-Webapplikation für das Auswerten biographischer Daten." In: *DHd 2017, Book of Abstracts*, Bern. doi: 10.5281/zenodo.3684825
- Vasari, G.** (2008): *The Lives of the Artists* (Neuaufgabe, Original veröffentlicht um 1550). Oxford, UK: Oxford University Press.
- Windhager, F.** (2020): *A Synoptic Visualization Framework for Artwork Collection Data and Artist Biographies*. Doctoral Thesis, Universität Wien. <http://othes.univie.ac.at/65279/1/67771.pdf>
- Windhager, F. / Federico, P. / Schreder, G. / Glinka, K. / Dörk, M. / Miksch, S. / Mayr, E.** (2018): „Visualization of cultural heritage collection data: State of the art and future challenges.“ In: *IEEE Transactions on Visualization & Computer Graphics* 25: 2311–2330.
- Windhager, F. / Salisu, S. / Mayr, E.** (2019): "Exhibiting uncertainty: Visualizing data quality indicators for cultural collec-

tions. in: *Informatics*. Special Issue on Uncertainty in Digital Humanities, 6: <https://doi.org/10.3390/informatics6030029>

Windhager, F. / Schlögl, M. / Kaiser, M. / Bernad, Á. / Sallisu, S. / Mayr, E. (2018). „Beyond one-dimensional portraits: A synoptic approach to the visual analysis of biography data.” in: A. Fokkens / S. ter Braake / R. Sluijter / P. Arthur / E. Wandl-Vogt (Eds.): *Proceedings of BD-2017 - Biographical Data in a Digital World*. Linz: CEUR. <http://ceur-ws.org/Vol-2119/paper11.pdf>