https://doi.org/10.3986/9789610511038_9

EVA MAYR

https://orcid.org/0000-0001-8402-5990
University of Continuing Education Krems, Research Group Digital Cultures and Digital Humanities, Department for Arts and Cultural Studies, Austria.

FLORIAN WINDHAGER

https://orcid.org/0000-0002-5170-2243
University of Continuing Education Krems, Research Group Digital Cultures and Digital Humanities, Department for Arts and Cultural Studies, Austria.

PROJECT INTAVIA – OR HOW TO DIGITALLY AND VISUALLY REFRAME THE HISTORY OF ALL (WO)MEN'S LIFE AND WORK[1]

Eva Mayr and Florian Windhager

Biographical texts on the lives of notable men and (much less frequently) women have a long and multifaceted history, deeply rooted in the culture and literature of our present-day societies. Precursors for biographical texts can be found in inscriptions on Egyptian tombs and monuments, such as the Pyramid Texts, to glorify the lives and deeds of noble individuals. In Greek literature, Herodotus' *Histories*, Homer's *Odyssey*, and Xenophon's *Anabasis* continued this work by interweaving biographical and mythological narrative perspectives. Looking beyond individual biographies, Plutarch (ca. 46–120 AD) wrote the famous *Parallel Lives*, which compared the lives of famous Greeks and Romans.

The rediscovery of ancient texts and the interest in the potential of individual human beings fostered the creation of the first early modern biographical texts in the European Renaissance. As a collection of artists' biographies, Giorgio Vasari's *Le Vite de' più eccellenti pittori, scultori, e architettori* [The Lives of the Most Excellent Painters, Sculptors, and Architects] is known as the foundational work for this genre. By focusing on groups of people sharing common professional or social characteristics, he also laid the foundations for the development of prosopographical studies.

[1] The project InTaVia has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004825.

The collection and curation of "national biographies" (i.e., as compilations and lexica) led to a significant advancement in the organization, documentation and dissemination of biographical information on a national scale. These collections compiled standardized biographical texts about notable individuals from a specific country, making them available for historians, educators, and the general public. This idea of systematically compiling biographical information on prominent individuals within a nation began to take shape in the 18th and 19th century and resulted in printed editions of biographical dictionaries and encyclopedias.

The transition from printed dictionaries to digital databases began in the late 20th century with the advent of computing technology and its global interconnection. Databases and related software tools enabled historians to handle ever larger amounts of information, making it easier to search and analyze biographical data and identify patterns and relationships within. The integration of digital humanities tools and methods has further enhanced the scope and benefits of national biographical databases. Techniques such as data visualization, network analysis, and geographic information systems (GIS) allow researchers to explore distributions, connections and trends within the data in innovative ways. Projects like Mapping the Republic of Letters, which traces the correspondences of Enlightenment thinkers, demonstrated how cultural objects (such as letters) can help to redraw the history-making interactions between individuals and ideas and thus show the potential of analyzing work and lives from a synoptic perspective.

In this context, the H2020 project InTaVia was conceived as a Research and Innovation Action funded by the European commission from 2020 to 2023. In this chapter, we will introduce this collaborative research project (section 1), before we provide an overview of the underlying data and its translation into the InTaVia knowledge graph (section 2). For the curation, analysis, and communication of these data, a tripartite frontend has been developed (section 3). We conclude this chapter with the discussion of challenges and insights generated in the InTaVia context (section 4).

1. IN/TANGIBLE EUROPEAN HERITAGE: VISUAL ANALYSIS, CURATION, AND COMMUNICATION

In modern societies, cultural information is available in a great variety of forms and formats, from tangible objects, collected by galleries, museums and libraries, to documents on "intangible" cultural heritage, such as the (hi)stories of important cultural actors. In contrast to material objects, biographical documents are commonly collected by archives or research institutions in written form. After digitizing material object collections, this data is mostly stored in type-specific, local databases, created and hosted by individual GLAM institutions. Even though projects like Europeana started to aggregate the content of individual object databases and allow for search across all of them, the objects within remain largely unconnected. Related cultural information, such as biographical texts about cultural actors and object creators (e.g., artists, authors, or inventors) remain mostly absent, as they have been commonly digitized by other heritage institutions and stored in different databases. This lack of links between biographical and object information is aggravated by missing standards between local data formats and database projects and a lack of (machine) readability and visibility of local data collections. Thus, they are difficult to use for experts and even less suited for offering and conveying a better understanding of cultural-historical topics to the interested public.

Against this background, the H2020 project InTaVia[2] pursued the goal of linking cultural object data and biographical information projects to generate a more synoptic picture of the lives and works of European cultural actors – and to make them available for further curation, analysis and communication. To this end, the consortium harmonized and integrated

[2] INTAVIA: In/ Tangible European Heritage: Visual Analysis, Curation and Communication, accessed June 26, 2024, https://intavia.eu.

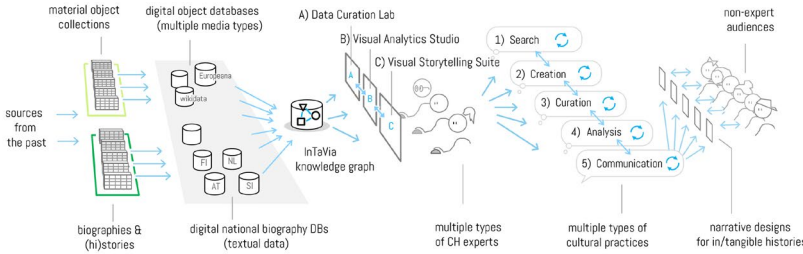

Figure 9.1. InTaVia project architecture: digitized object and person data are integrated into the InTaVia knowledge graph, which cultural heritage experts and researchers can use in a visualization-based interface for searching, curating, analysis, and communication. *Source*: authors.

national biographical databases from Slovenia,[3] Austria,[4] Finland,[5] and the Netherlands[6] and linked them with related cultural objects from Europeana and Wikidata in a comprehensive *cultural knowledge graph*. To enable access to this rich resource collection, the project developed an *information portal for the visual analysis, curation and communication* of this cultural data at multiple levels of aggregation. This overall project architecture is depicted in Figure 1.

2. THE INTAVIA KNOWLEDGE GRAPH

The InTaVia Knowledge Graph (IKG) was conceived and set up as a central knowledge base, in which the data from the various biography and object

[3] *Slovenska biografija*, see Tomaž Erjavec, Joh Dokler, and Petra V. Ogrin, "Slovenian Biography," in *Biographical Data in a Digital World*, ed. Antske Fokkens et al. (Linz: CEUR, 2017), 16–21, accessed June 26, 2024, https://ceur-ws.org/Vol-2119/paper3.pdf.

[4] Austrian Prosopographical Information System (APIS), see Matthias Schlögl and Katalin Lejtovicz, "A Prosopographical Information System (APIS)," in *Biographical Data in a Digital World*, ed. Antske Fokkens et al. (Linz: CEUR, 2017), 53–58, accessed June 26, 2024, https://ceur-ws.org/Vol-2119/paper9.pdf.

[5] BiographySampo, see Eero Hyvönen et al., "BiographySampo–Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research," in *European Semantic Web Conference* (Cham: Springer International Publishing, 2019), 574–589.

[6] BiographyNet, see Antske Fokkens et al., "BiographyNet: Extracting Relations between People and Events," in *Europa baut auf Biographien: Aspekte, Bausteine, Normen und Standards für eine europäische Biographik*, ed. Ágoston Z. Bernád, Christine Gruber, and Matthias Schlögl (Wien: new academic press, 2017), 193–224, accessed June 26, 2024, https://arxiv.org/pdf/1801.07073.pdf.

databases were merged and linked. This knowledge graph enabled the querying, further enrichment, harmonization and linking of data, while preserving relevant provenance information.

While a Blazegraph triplestore provided a flexible and performative basis, we developed the InTaVia Data Model IDM-RDF[7] as an ontology, based on CIDOC CRM[8] (version 7.1.1) and the BioCRM extension.[9] Furthermore, the graph implemented an adapted version of the proxy model from the Object Reuse and Exchange Ontology[10] for linking individual entities. The IKG also provides a RestAPI for simplified access to the data.[11]

For the enrichment, harmonization and linking of the data, the IKG relies on ETL pipelines that run in a Kubernetes cluster. Various Prefect v1[12] pipelines import and update person entities, convert them to IDM-RDF and enrich them with object entities. To improve this workflow, a SHACL validation of the data has been added, before the data is loaded into the triplestore. These pipelines allow for the ingestion of further biographical databases in the future and their enrichment with related cultural objects from Wikidata and Europeana.

At the end of the project, the IKG contained 24,588,310 triples[13] describing and connecting more than 165,960 persons,[14] 24,446 places[15] and 230,068

---

[7] For documentation, see Carla Ebel et al., *InTaVia Data Model (IDM-RDF)*. Deliverable within the H2020 Project InTaVia, 2021; see "idm-rdf: Intavia Data Model for RDF data," InTaVia, GitHub repository, last modified December 7, 2023, https://github.com/InTaVia/idm-rdf.

[8] Martin Doerr, "The CIDOC Conceptual Reference Module – An Ontological Approach to Semantic Interoperability of Metadata," *AI Magazine* 24, no. 3 (2003): 75, accessed July 30, 2025, https://doi.org/10.1609/aimag.v24i3.1720.

[9] Jouni Tuominen, Eero Hyvönen, and Petri Leskinen, "Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research," in *Biographical Data in a Digital World*, ed. Antske Fokkens et al. (Linz: CEUR, 2017), 59–66, accessed June 26, 2024, https://ceur-ws.org/Vol-2119/paper10.pdf.

[10] Lagoze, Carl, Herbert Van de Sompel, Pete Johnston, Michael Nelson, Robert Sanderson, and Simeon Warner, eds, "ORE Specification – Abstract Data Model," *Open Archives Initiative*, accessed June 26, 2024, http://www.openarchives.org/ore/1.0/datamodel.

[11] . "InTaVia IDM-Json Backend." InTaVia, API Documentation. accessed June 26, 2024, https://intavia-backend.acdh-dev.oeaw.ac.at/v2/docs.

[12] "Prefect Docs," *Prefect Technologies*, *Prefect 1*, accessed June 26, 2024, https://docs-v1.prefect.io/.

[13] Triples are structured statements connecting two entities, for example, "Carla was born in Amsterdam" or "Oskar created Carla's portrait in 1897."

[14] These are represented by approximately 230,000 person proxies, as for some persons multiple biographies exist.

[15] These are again represented by approximately 36,000 place proxies, representing different wordings of the same place.

cultural heritage objects from Europeana, as well as 160,239 from Wikidata (see Figure 2). The majority of documented biographical events (including events of object creations) stretch out through the 19th and 20th centuries. A critical evaluation of this knowledge graph uncovered several development challenges:[16]

1. The number of inter-entity relations turned out to be quite unbalanced. Two main clusters can be seen in figure 2, while several unrelated actors remain at the periphery of the graph.

2. In addition, the person data is clearly skewed: The IKG covers not only fewer women than men; but also reproduces the well-known elite bias of national biographies and their hagiographic tendencies.

3. Finally, many biographical entities in the InTaVia knowledge consist of a disappointingly low number of biographical events: oftentimes, no information is available beyond the birth and death of a person, or the creation of the most famous objects.

The critical evaluation of the knowledge graph thus clearly showed the need for the further enrichment of the original data collection. Within the InTaVia project, we explored several approaches, how the density of events could be increased:[17]

1. NLP-pipelines are available for all of the languages within the InTaVia project. To improve their performance, we developed approaches to resolve abbreviations, which are frequently used in biographical texts.

[16] Matthias Schlögl et al., *The InTaVia Knowledge Graph – Publishing European National Biographical and Cultural Heritage Object Data,* submitted to *Semantic Web Journal*, 2025, accessed February 25, 2026, https://semantic-web-journal.net/system/files/swj3851.pdf.

[17] Angel Daza, Antske Fokkens, and Tomaž Erjavec, "Dealing with Abbreviations in the Slovenian Biographical Lexicon," in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, ed. Yoav Goldberg et al. (Abu Dhabi: Association for Computational Linguistics, 2022), 8715–8720, accessed July 30, 2025, https://doi.org/10.18653/v1/2022.emnlp-main.596.

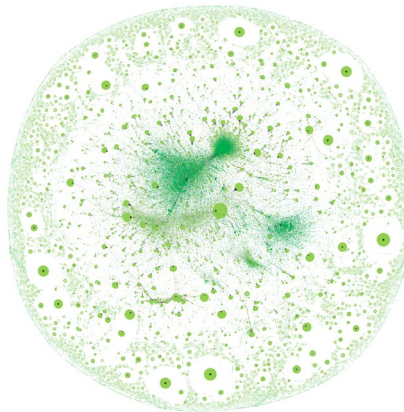

Figure 9.2. Network visualization of the InTaVia knowledge graph generated with Cosmograph.[18] Nodes in dark green correspond to person entities, nodes in light green to object entities. *Source*: authors.

2. Recent developments in Artificial Intelligence also offer novel opportunities for data enrichment[19] and the identification of novel relations.[20]
3. Another option for data enrichment is the extraction of further resources. For example, a novel NLP pipeline has been developed to extract events from Wikipedia texts on entities.[21]
4. Finally, we decided to support the upload of further datasets and the manual curation of object and biography data within the InTaVia frontend.

---

[18] Nikita Rokotyan, Olya Stukova, D. Kolmakova, and D. Ovsyannikov, *Cosmograph: GPU-accelerated Force Graph Layout and Rendering* [Computer software] (2022), accessed June 24, 2024, https://cosmograph.app/.

[19] Lea Krause et al., "Confidently Wrong: Exploring the Calibration and Expression of (Un) Certainty of Large Language Models in a Multilingual Setting," in *Proceedings of the Workshop on Multimodal, Multilingual Natural Language Generation and Multilingual WebNLG Challenge (MM-NLG 2023)*, ed. Albert Gatt et al. (Prague: Association for Computational Linguistics, 2023), 1–9.

[20] Heikki Rantala, Petri Leskinen, Lilli Peura, and Eero Hyvönen, *Searching Interesting Relations in Cultural Heritage Knowledge Graphs* (Manuscript under review, 2024), accessed June 24, 2024, https://seco.cs.aalto.fi/publications/2024/rantala-et-al-searching-interesting-relations-2024.pdf.

[21] Daza Arévalo and José Angel, "Generating Structured Data from Wikipedia Biographies," in *There's History in All Men's Lives* (Ljubljana: Zenodo, September 26, 2023), https://doi.org/10.5281/zenodo.10040678.

3. THE INTAVIA FRONTEND

The InTaVia frontend can be reached at https://intavia.acdh-dev.oeaw.ac.at/ (accessed June 26, 2024). It consists of three interconnected main modules for (1) search and curation (Data Curation Lab), (2) visualization and analysis (Visual Analytics Studio), and (3) communication and storytelling (Storytelling Suite).

To ensure the platform's usefulness for digital humanities scholars and cultural heritage experts as well as its usability, the development of the InTaVia platform followed a user-centered design approach. More than 40 cultural heritage experts and GLAM practitioners were involved in three workshops to identify relevant user practices and define user requirements at the outset of the project.[22] During the development, expert users gave feedback on the state of the platform development during several workshops or individual test sessions in three iterations.

3.1. DATA CURATION LAB

The Data Curation Lab is the user's entry point to the InTaVia Knowledge Graph. It provides the means to search, inspect, collect, and curate cultural information on persons, objects, institutions, and places (see figure 3).

Users can access the knowledge graph through *queries*, which can be customized based on specific parameters like text labels, entity types, or relations to specific entities. This allows for targeted searches and retrieval of relevant cultural entities. Query results are first shown in a list (Figure 3b) where each entry provides a summary of the respective entity, such as its label, type, and number of related events. To manage retrieved entities and reuse them across all platform modules, users have to add selected entities or the whole query result set to collections (Figure 3c). *Collections* are used throughout the InTaVia frontend as user-defined sets of entities for visualization, analysis and communication.

For close reading, all information available on an entity is presented on a *detail page* that includes network, map, and timeline views on their

[22] Eva Mayr et al., "The Multiple Faces of Cultural Heritage: Towards an Integrated Visualization Platform for Tangible and Intangible Cultural Assets," in *2022 IEEE 7th Workshop on Visualization for the Digital Humanities (VIS4DH)* (Oklahoma City: IEEE, 2022), 13–18, https://doi.org/10.1109/VIS4DH57440.2022.00008.

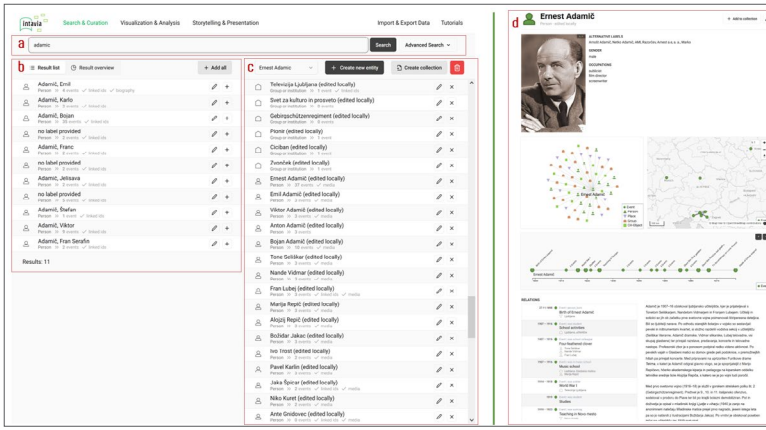

Figure 9.3. Overview of the Data Curation Lab containing search options (a), a result list (b) and a personal data collection (c). Selecting an entity opens the detail view (d) with related sources and IDs, a network, map, and time visualization, an event list, and a biographical text. *Source*: authors.

related entities and events, in addition to media files and a biographical text (Figure 3d).

To go beyond the data within the InTaVia Knowledge Graph (and also counteract existing imbalances and gaps within the data), the Data Curation Lab allows the local *import* of a user's own local data from different sources like Excel sheets or JSON files. It is important to note that these data are not shared with the IKG, nor is the IKG updated when entities are locally edited.

Furthermore, users can create and *curate* data within the Data Curation Lab. This functionality allows manually correcting and enriching the data from the knowledge graph or merging it with locally imported data. In addition to curating basic entity information like labels, linked URLs, and occupations, users can also enrich an entity's events and relations to other entities. Additionally, they can create new entities not represented in the knowledge graph. These rich curation features give users complete control over their data used for analysis and storytelling in the subsequent steps.

3.2. VISUAL ANALYTICS STUDIO

The data on cultural objects and biographies stored in the IKG encompass a variety of facets and dimensions that can be of interest to historians and cultural scientists and can be explored in new ways using visual-analytical

means:[23] the geographical position of biographical or artistic events, various events or cultural entities (objects or people), relationships between people and/or objects and chronological sequences of events. These aspects can be relevant for various questions at different levels of aggregation – from historical individuals to diverse groupings.

Consequently, the Visual Analytics Studio module of the InTaVia platform provides temporal, spatial, and relational perspectives on cultural heritage information[24] (see Figure 4). Users can create coordinated visualizations in flexible workspaces with adaptable multi-panel layouts that provide synoptic views with several perspectives on the selected data simultaneously. The three visualization types supported by the VA Studio are *maps* (including a space-time cube option), *network graphs*, and *timelines*. Visualizations can be customized to fit the current information-seeking needs. Multiple options for color encoding (e.g., by entity type or by time) enable different perspectives on the data. Various visual aggregation methods, such as donut charts or bee swarm clusters, prevent occlusion and visual clutter in analysis scenarios with many events in close spatial or temporal proximity. In addition, features like visualizing life paths with trajectories on maps or in space-time cubes support historians' analysis tasks. These visual analytics tools help digital humanities scholars and cultural heritage experts understand the connections and patterns within their data collections and develop and consolidate insights.

Figure 4 shows the multi-perspective interface for the visual analysis of life and work using an example of Ernest Adamič.[25] It contains a list of events (a), a timeline visualization which compares his life with the life of Bojan Adamič (b), a visualization of their network showing related persons, places, groups and cultural objects (c), and a map showing important places

---

[23] Florian Windhager, Eva Mayr, Matthias Schlögl, and Maximilian Kaiser, "Visuelle Analyse und Kuratierung von Biographiedaten," in *Digital History: Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft*, ed. Karoline D. Döring et al. (Walter de Gruyter GmbH & Co KG, 2022), 137–150, https://doi.org/10.1515/9783110757101-008; Florian Windhager et al., "Visualization of cultural heritage collection data: State of the art and future challenges," *IEEE transactions on visualization and computer graphics* 25 (2018): 2311–2330, https://doi.org/10.1109/TVCG.2018.2830759.

[24] Eva Mayr et al., "The multiple faces," 13–18.

[25] Ahac Meden, "Adamič, Ernest (1898–1977)," *Slovenska biografija*, accessed June 23, 2024, http://www.slovenska-biografija.si/oseba/sbi1000070/#novi-slovenski-biografski-leksikon.

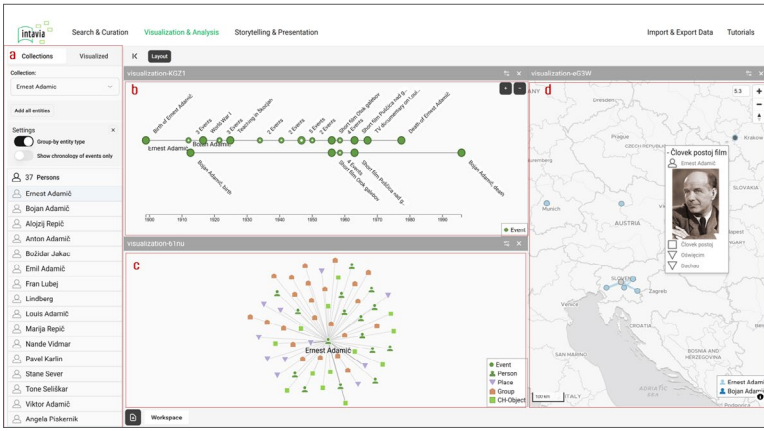

Figure 9.4. Visual Analytics Studio with an overview on the entities in the collection (a), which can be added to timelines (b), network visualizations (c), and maps (d). *Source*: authors.

in their lives (d). If a user moves the mouse over an event, it is also highlighted in the other views and a detailed preview is displayed. Clicking on the object opens a detailed view with further information and visualizations.

For communication purposes, the visualizations created in the VA Studio can be re-used in the final stage of the workflow by incorporating them into visual stories.

3.3. STORYTELLING SUITE

Narrative techniques of visualization-based storytelling open up a new, lower-threshold access to cultural heritage data – especially for interested laypersons and the general public.[26] However, the creation of such data – and visualization-based stories is very time-consuming without suitable tools. The InTaVia platform therefore offers the possibility of enriching selected data from the InTaVia Knowledge Graph and visualizations from the Visual Analytics Studio with explanatory texts, media, and interactive elements and linking them narratively.[27]

---

[26] Jakob Kusnick, et al., *Report on narrative visualization techniques for OPDB data*, deliverable, InTaVia project, 2021, accessed July 10, 2023, https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5e47d9524&appId=PPGMS.

[27] Jakob Kusnick et al., "Every Thing Can Be a Hero! Narrative Visualization of Person, Object, and Other Biographies," *Informatics* 11, no. 2 (2024): 26, https://doi.org/10.3390/

The *Story Creator*, as the first component of the Visual Storytelling Suite, integrates several features to intuitively create stories on cultural actors (see Figure 5). Users can create and organize their content visually in slides. It allows combining novel or pre-existing visualizations from the VA Studio with other media into unique layouts (see Figure 5b). When designing a story, experts decide which visualizations to use, which event data to highlight and which representation to offer for interactive exploration by the target audience. Transitions between individual stages of the story on maps or timelines are animated later on and thus made intuitively comprehensible. All visualizations can be enriched and combined with various multimedia content, such as images, videos, 3D models, text blocks, and HTML (see Figure 5c). Interactive elements such as quiz games, external media, and 3D renderings can further raise narrative engagement and user experience.

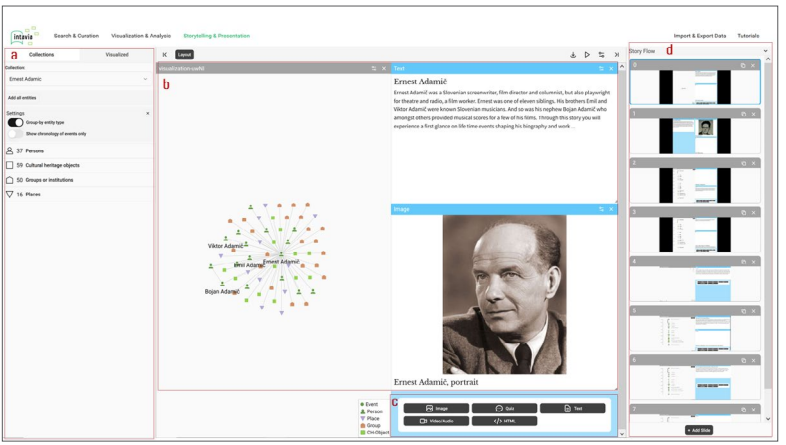

Figure 9.5. Story Creator using the collection panel for entity selection (a), a workspace for the creation of slides (b) with visualizations and further media elements (c), as well as an overview panel for all slides (d). *Source*: authors.

The *Story Viewer*, as the second major component of the Storytelling Suite, provides the missing link to story audiences and fosters the interactive reception of stories developed with the Story Creator, as well as sharing them

informatics11020026; Johannes Liem et al., "A Workflow Approach to Visualization-Based Storytelling with Cultural Heritage Data," in *2023 IEEE 8th Workshop on Visualization for the Digital Humanities (VIS4DH)* (Melbourne: IEEE, 2023), 13–17, https://doi.org/10.1109/VIS4DH60378.2023.00008.

via unique URLs. It brings stories to life through dynamic visual elements, ensures smooth transitions between slides, and delivers high-quality visual renderings in a responsive manner to support mobile devices (see Figure 6). Additionally, it incorporates interactive features that aim to enhance user engagement and provide an immersive storytelling experience. Its design focuses on maintaining user interest and involvement, thus setting a new standard for visualization-based storytelling technology.

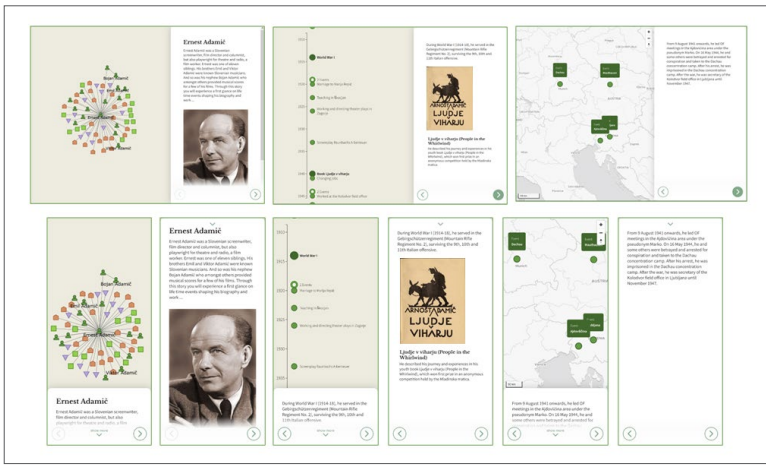

Figure 9.6. Three selected slides from a story on the life of Ernest Adamič[28] for desktop (top) and mobile devices (bottom), which combine network, time, or map visualizations with text and other media. *Source*: authors.

4. GOING BEYOND DATA SILOS – CHALLENGES AND OPPORTUNITIES

Based on our experiences and discussions within InTaVia, but also with scholars working on related questions,[29] we want to summarize our expe-

[28] Ahac Meden, "Adamič, Ernest (1898–1977)," *Slovenska biografija*, accessed June 23, 2024, http://www.slovenska-biografija.si/oseba/sbi1000070/#novi-slovenski-biografski-leksikon.

[29] In two events, we discussed our results with a community of interested researchers and scholars: Next to the conference *There is History in All Men's Lives*, we organized the symposium *Smashing the Silos! The Future of Cultural Heritage Information and Visualization* as a pre-event to DH2023, see "Smashing the Silos! The Future of Cultural Heritage Information and Visualization," InTaVia, accessed June 23, 2024, www.intavia.eu/dh2023.

riences with ten theses for the work with cultural data, as a unique type of digital information:

1. **Most cultural data is scattered and siloed**. Currently, digitized information on the lives of cultural actors and on cultural objects are scattered and siloed. By collecting and digitizing them in an unrelated fashion across separated institutions, historically relevant relations have been either broken or never made explicit or visible. InTaVia made a first attempt to reconnect cultural actors and objects and, thereby, enable the contextual analysis of cultural objects with a synoptic historic dimension. >> *To overcome fragmented perspectives, it is relevant to link cultural databases!*

2. **Most cultural data is not human readable**. Data formats like CIDOC CRM, EDM, TEI, but also IDM RDF cannot be easily understood by non-data-experts, nor can they be readily used. Novel interfaces are required, which build upon more intuitive forms of representation and human cognition. InTaVia developed visualization-based interfaces which make use of human visual faculties, and combined them with a narrative presentation, a very intuitive form of human communication and thinking. >> *To foster general utilization and understanding, it is relevant to develop intuitive interfaces to cultural data!*

3. **Cultural data is potentially big and rich in detail**. Over the course of centuries, a lot of cultural heritage information has been collected – and recently also digitized. Therefore, it has become difficult or even impossible to gain high-level overviews on the available information with conventional means. In InTaVia, we aimed for the creation of both: tools to gain an overview on cultural big data (via the Visual Analytics Studio) as well as tools for detailed inspection of individual entities (via detail pages). >> *In the cultural heritage field, the development of scalable reading techniques is an omnipresent challenge, to intertwine the exploration and analysis of big data sets with the attentive study and interpretation of details – and all possible scales in between!*

4. **Cultural data is oftentimes scarily scarce**. When it comes to structured cultural data, we frequently experienced their structure to be worryingly scarce and sparse. Most entities within the InTaVia knowledge graph have only few events attached: Birth, death, and creation events are most frequent. But also the unstructured texts of national biographies often lack details. Here, existing biases also become obvious: Actors and objects which are regarded as more important have more detailed descriptions than the majority of entries. In InTaVia, we developed several mitigation strategies to enrich such data stumps. On the digital side, NLP and AI allow the retrieval of structured data from other sources (like Wikipedia or correspondence networks). On the expert's side, the interface allows for manual data input, curation, and upload of additional local data. >> *It is relevant to develop novel solutions for automatic and manual data enrichment.*

5. **Cultural data is often uncertain and contested**. Historical sources are rarely conclusive, complete, or uncontested. Frequently, integrating multiple data sources also results in inconsistencies as manifestations of different historiographical conclusions or assumptions. Furthermore, more ancient cultural topics and entities are less well documented than recent eras. Such uncertainties and ambiguities need to be made transparent – together with the available information on the provenance of the data. Therefore, the InTaVia knowledge graph contains multiple proxies for contested entities or events, representing different information sources for persons, places, or objects. In the frontend, these information pieces are represented next to each other, while the sources are made transparent to the user. Visual representation of incomplete or uncertain information is an important challenge for the visualization of cultural information.[30] >> *It is relevant to preserve and document uncertainties and ambiguities*

[30] Windhager et al., "Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges," *IEEE Transactions on Visualization and Computer Graphics* 25 (2018): 2311–2330, https://doi.org/10.1109/TVCG.2018.2830759; Florian Windhager, Saminu Salisu, and Eva Mayr, "Exhibiting Uncertainty: Visualizing Data Quality Indicators for Cultural Collections," *Informatics* 6, no. 3 (2019): 29, https://doi.org/10.1109/TVCG.2018.2830759.

in cultural data, to make the oftentimes contested bigger pictures of historical research transparent.

6. **Cultural data is frequently biased**. Cultural databases often contain various biases. National biographies focus more frequently and more extensively on men than women, cultural objects of minority cultures or classes are less often represented in collections. In addition, the digitization processes – as an ongoing activity over a longer time period – did not start with a random collection of cultural information, but with those, which are regarded most important. As such, the InTaVia knowledge graph is an accumulation of biased data sources. Still, by integrating different data sources, it contains multiple (biased) perspectives, which can complement one another. Also, we developed open ingestion and enrichment pipelines, which allow to integrate further databases in the future and to update the data within the InTaVia Knowledge Graph, as digitization of cultural information progresses. >> *It is relevant to be aware of existing biases in cultural data and to strive for their compensation by integrating multiple sources.*

7. **Cultural data is autotelic**. By definition, cultural objects are known to be collected due to their high "aesthetic, historic, scientific, or social value,"[31] which contributes to their inherent attraction power. As such, an interface to these objects has to match, but also freely show and properly frame their beauty and appeal. In InTaVia, we included URLs to related media representations in the knowledge graph, wherever possible, i.e., for all entities. On the frontend side, we showed these related media materials in several instances: on the detail page for each entity, as a small preview upon mouse-over in the visualizations for each event, and as a preview for each entity in the collection panel of the Visual Analytics Studio and Story Editor. Thereby, we can also increase the user experience of the InTaVia platform and attract the interest of users for specific cultural actors or objects with related media representations. >> *It is relevant to*

[31] International Council on Monuments and Sites (ICOMOS), *Historic gardens (The Florence charter)* (Florence, 1982).

integrate close-up views of cultural objects saliently in an interface, but also to build attractive interfaces with high user experience.

8. **Cultural data is different for different users**. Different users come with different needs and different levels of prior knowledge – which also influences their needs and wishes for a cultural information portal. Whitelaw[32] coined the term "generous interfaces for cultural collections" to put emphasis on the fact that multiple entry points should serve the requirements of different user groups. In InTaVia, we defined ten personas[33] with specific information needs, prior knowledge, and user practices around cultural information (Figure 7). Based on their characteristic user profiles, we developed an interface, which allows for open exploration of data, for directed analysis, and for storytelling – supported by visualization. In addition, open APIs give access to the backend for re-use with other tools and allow to ingest novel databases, as well as to import own data to the frontend. *>> It is relevant to provide multi-purpose tools (e.g. exploratory, analytic, and narrative) for different types of users, but also open APIs for re-use by cultural heritage experts.*

Figure 9.7. Ten personas with distinct user profiles, which have been defined in InTaVia. *Source*: authors.

[32] Mitchell Whitelaw, "Generous interfaces for digital cultural collections," *Digital humanities quarterly* 9, no. 1 (2015): 1–16.

[33] Eva Mayr, et al., "The multiple faces," 13–18.

9. **Cultural data and tools are non-neutral**. Cultural data and tools are influenced by several factors: Collection strategies and biases, national policies and identities, technological constraints, knowledge and skills of the people involved, and several others. As such, we should not encourage trusting the data and the tools in the first place, but offer avenues for their critical inspection, comparison, and discussion. In InTaVia, we involved experts on cultural heritage and digital humanities in multiple phases to collect their feedback and adjust our developments. In addition, we did inspect and discuss the data and tools within the project team on several occasions to bring in different perspectives. Still, we are aware that the project's data and tools are influenced and constrained by, the project aims, individual and organizational agendas, time constraints, and several other factors. >> *It is relevant to critically reflect on the data and tools in digital humanities in general, but especially for cultural and historical information.*

10. **Cultural data and digital tools are (only) sometimes relevant**. Digital collections, together with digital methods and tools offer new ways for accessing cultural data, for linking and curation and for scalable analysis and communication, but they cannot substitute the traditional, non-digital research on cultural information. They can complement and augment traditional research, but non-digital methods are at least equally important as digital ones. In a postdigital society, digital humanists with their data and tools have to work hand in glove with non-digital cultural heritage experts and the research methods from both sides immensely benefit from productive joint ventures. >> *It is relevant to complement digital and non-digital methods and tools for research and practices around cultural information.*

The InTaVia project explored novel opportunities which digital technologies offer for research and practice around cultural heritage. The resulting knowledge graph and the InTaVia information portal for visual analysis, curation, and communication thus showcase how digital technologies can influence the digital representation and work with cultural heritage. Yes, there is history in all (wo)men's lives – forming them and being formed by them.

We consider digital methods to provide a unique opportunity to strengthen the historiographic perspective in modern times, to link those lives to further (re)sources, widen our analysis, and expand our understanding, as well as our communication activities on the biographies of actors we care about.

BIBLIOGRAPHY

Arévalo, Daza, and José Angel. "Generating Structured Data from Wikipedia Biographies." In *There's History in All Men's Lives*. Ljubljana: Zenodo, September 26, 2023. https://doi.org/10.5281/zenodo.10040678.

Daza, Angel, Antske Fokkens, and Tomaž Erjavec. "Dealing with Abbreviations in the Slovenian Biographical Lexicon." In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, edited by Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, 8715–8720. Abu Dhabi: Association for Computational Linguistics, 2022. https://doi.org/10.18653/v1/2022.emnlp-main.596.

Doerr, Martin. "The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata." *AI Magazine* 24, no. 3 (2003): 75–92. https://doi.org/10.1609/aimag.v24i3.1720.

Ebel, Carla, Jouni Tuominen, Matthias Schlögl, Joonas Kesäniemi, and Florian Windhager. *InTaVia Data Model (IDM-RDF)*. Deliverable within the H2020 Project InTaVia, 2021. Accessed April 18, 2024. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5e67fcfb6&appId=PPGMS.

Erjavec, Tomaž, Joh Dokler, and Petra V. Ogrin. "Slovenian Biography." In *Biographical Data in a Digital World*, edited by Antske Fokkens, Serge ter Braake, Ronald Sluijter, Paul Arthur, and Eveline Wandl-Vogt, 16–21. Linz: CEUR, 2017. Accessed June 26, 2024. https://ceur-ws.org/Vol-2119/paper3.pdf.

Fokkens, Antske, Serge ter Braake, Niels Ockeloen, Piek Vossen, Susan Legêne, Guus Schreiber, and Victor de Boer. "BiographyNet: Extracting Relations between People and Events." In *Europa baut auf Biographien: Aspekte, Bausteine, Normen und Standards für eine europäische Biographik*, edited by Ágoston Zénó Bernád, Christine Gruber, and Matthias Schlögl, 193–224. Wien: New Academic Press, 2017. Accessed June 26, 2024. https://arxiv.org/pdf/1801.07073.pdf.

Hyvönen, Eero, Petri Leskinen, Minna Tamper, Heikki Rantala, Esko Ikkala, Jouni Tuominen, and Kirsi Keravuori. "BiographySampo–Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research." In *European Semantic Web Conference*, 574–589. Cham: Springer International Publishing, 2019.

International Council on Monuments and Sites (ICOMOS). *Historic Gardens (The Florence Charter).* Florence, 1982.

Krause, Lea, Wondimagegnhue Tufa, Selene Baez Santamaria, Angel Daza, Urja Khurana, and Piek Vossen, "Confidently Wrong: Exploring the Calibration and Expression of (Un)Certainty of Large Language Models in a Multilingual Setting." In *Proceedings of the Workshop on Multimodal, Multilingual Natural Language Generation and*

Multilingual WebNLG Challenge (MM-NLG 2023), edited by Albert Gatt, Claire Gardent, Liam Cripwell, Anya Belz, Claudia Borg, Aykut Erdem, and Erkut Erdem, 1–9. Prague: Association for Computational Linguistics, 2023.

Kusnick, Jakob, Eva Mayr, Kasra Seirafi, Samuel Beck, Johannes Liem, and Florian Windhager. "Every Thing Can Be a Hero! Narrative Visualization of Person, Object, and Other Biographies." *Informatics* 11, no. 2 (2024): 26. https://doi.org/10.3390/informatics11020026.

Kusnick, Jakob, Stefan Jänicke, Carina Doppler, Kasra Seirafi, Johannes Liem, Florian Windhager, and Eva Mayr. *Report on narrative visualization techniques for OPDB data*. Deliverable, InTaVia project, 2021. Accessed July 10, 2023. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5e47d9524&appId=PPGMS.

Liem, Johannes, Jakob Kusnick, Samuel Beck, Florian Windhager, and Eva Mayr. "A Workflow Approach to Visualization-Based Storytelling with Cultural Heritage Data." In *2023 IEEE 8th Workshop on Visualization for the Digital Humanities (VIS4DH)*, 13–17. Melbourne: IEEE, 2023. https://doi.org/10.1109/VIS4DH60378.2023.00008.

Mayr, Eva, Florian Windhager, Johannes Liem, Samuel Beck, Steffen Koch, Jakob Kusnick, and Stefan Jänicke. "The Multiple Faces of Cultural Heritage: Towards an Integrated Visualization Platform for Tangible and Intangible Cultural Assets." In *2022 IEEE 7th Workshop on Visualization for the Digital Humanities (VIS4DH)*, 13–18. Oklahoma City: IEEE, 2022. https://doi.org/10.1109/VIS4DH57440.2022.00008.

Rantala, Heikki, Petri Leskinen, Lilli Peura, and Eero Hyvönen. *Searching Interesting Relations in Cultural Heritage Knowledge Graphs*. Manuscript under review, 2024. Accessed June 24, 2024. https://seco.cs.aalto.fi/publications/2024/rantala-et-al-searching-interesting-relations-2024.pdf.

Schlögl, Matthias, and Katalin Lejtovicz. "A Prosopographical Information System (APIS)." In *Biographical Data in a Digital World*, edited by Antske Fokkens, Serge ter Braake, Ronald Sluijter, Paul Arthur, and Eveline Wandl-Vogt, 53–58. Linz: CEUR, 2017. Accessed June 26, 2024. https://ceur-ws.org/Vol-2119/paper9.pdf.

Schlögl, Matthias, Jouni Tuominen, Joonas Kesäniemi, Petri Leskinen, Go Sugimoto and Victor de Boer, *The InTaVia Knowledge Graph – Publishing European National Biographical and Cultural Heritage Object Data*. Submitted to *Semantic Web Journal*, 2025. Accessed February 25, 2026. https://semantic-web-journal.net/system/files/swj3851.pdf.

Tuominen, Jouni, Eero Hyvönen, and Petri Leskinen. "Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research." In *Biographical Data in a Digital World*, edited by Antske Fokkens, Serge ter Braake, Ronald Sluijter, Paul Arthur, and Eveline Wandl-Vogt,

59–66. Linz: CEUR, 2017. Accessed June 26, 2024. https://ceur-ws.org/Vol-2119/paper10.pdf.

Whitelaw, Mitchell. "Generous Interfaces for Digital Cultural Collections." *Digital Humanities Quarterly* 9, no. 1 (2015): 1–16.

Windhager, Florian, Eva Mayr, Matthias Schlögl, and Maximilian Kaiser. "Visuelle Analyse und Kuratierung von Biographiedaten." In *Digital History: Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft*, edited by Karoline Dominika Döring, Stefan Haas, Mareike König, and Jörg Wettlaufer, 137–150. Walter de Gruyter GmbH & Co KG, 2022. https://doi.org/10.1515/9783110757101-008.

Windhager, Florian, Paolo Federico, Günther Schreder, Katrin Glinka, Marian Dörk, Silvia Miksch, and Eva Mayr. "Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges." *IEEE Transactions on Visualization and Computer Graphics* 25 (2018): 2311–2330. https://doi.org/10.1109/TVCG.2018.2830759.

Windhager, Florian, Saminu Salisu, and Eva Mayr. "Exhibiting Uncertainty: Visualizing Data Quality Indicators for Cultural Collections." *Informatics* 6, no. 3 (2019): 29. https://doi.org/10.3390/informatics6030029.

INTERNET SOURCES

InTaVia. "idm-rdf: Intavia Data Model for RDF data." GitHub repository. Last modified December 7, 2023. https://github.com/InTaVia/idm-rdf.

InTaVia. "InTaVia IDM-Json Backend." API Documentation. Accessed June 26, 2024. https://intavia-backend.acdh-dev.oeaw.ac.at/v2/docs.

InTaVia. "Smashing the Silos! The Future of Cultural Heritage Information and Visualization." Accessed accessed June 23, 2024. https://intavia.eu/activities/smashing-the-silos-the-future-of-cultural-heritage-information-and-visualization/.

INTAVIA: In/ Tangible European Heritage: Visual Analysis, Curation and Communication. Accessed June 26, 2024, https://intavia.eu.

Lagoze, Carl, Herbert Van de Sompel, Pete Johnston, Michael Nelson, Robert Sanderson, and Simeon Warner, eds. "ORE Specification – Abstract Data Model." *Open Archives Initiative*. Accessed June 26, 2024. http://www.openarchives.org/ore/1.0/datamodel.

Meden, Ahac. "Adamič, Ernest (1898–1977)." *Slovenska biografija*. Accessed June 23, 2024. http://www.slovenska-biografija.si/oseba/sbi1000070/#novi-slovenski-biografski-leksikon.

Prefect Technologies. "Prefect Docs." *Prefect 1*. Accessed June 26, 2024. https://docs-v1.prefect.io/.

Rokotyan, Nikita, Olya Stukova, D. Kolmakova, and D. Ovsyannikov (2022). *Cosmograph: GPU-accelerated Force Graph Layout and Rendering* [Computer software]. Accessed June 24, 2024. https://cosmograph.app/.