

Bridging the Gap between Visual Analytics and Digital Humanities: Beyond the Data-Users-Tasks Design Triangle

Victor Schetinger*
TU Wien

Kathrin Raminger†
Musik und Kunst Privatuniversität
Susana Zapke‡
Musik und Kunst Privatuniversität

Velitchko Filipov‡
TU Wien
Silvia Miksch‡
TU Wien

Nathalie Soursos§
Musik und Kunst Privatuniversität

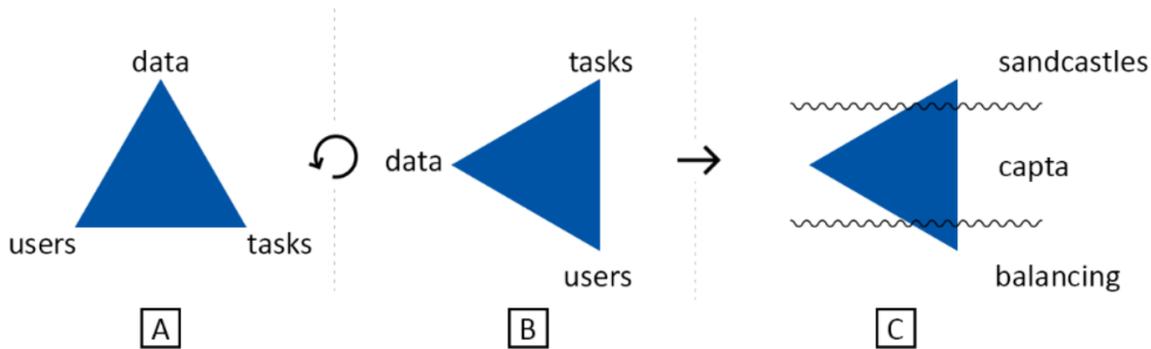


Figure 1: A) The Data-Users-Tasks design triangle. B) A counter-clockwise rotation breaks the isonomy of the triangle and establishes a vertical structure. C) Each part of the triangle stands within a particular realm of problematics.

ABSTRACT

This paper is an exercise in bridge building through re-thinking the common methodological foundations of Visual Analytics (VA) from a Digital Humanities (DH) perspective. We introduce the Data-Users-Tasks design triangle and re-purpose it to expose the complexities underlying interdisciplinary DH projects, tying it to undergoing discussion in the literature. We describe our experiences interacting with researchers from various disciplines, including musicology and (art) history, and conclude with the lessons that we learned along the way.

Keywords: Digital Humanities, Visual Analytics, Visualization, Interaction Design

1 INTRODUCTION

The goal of Visual Analytics (VA) is to support the discovery of knowledge and development of hypothesis from large volumes of data [8]. It can be defined as the “[...] science of analytical reasoning facilitated by interactive visual interfaces.” [17]. VA is important for fields dealing with large amounts of information that needs to be analyzed by humans, such as climate, medicine, and commerce. Recently it has gained momentum within the Digital Humanities

(DH) community [2, 4, 12, 19], where a large amount of different disciplines are brought together pursuing a common research question or goal.

VA projects can be understood as a dialectic between a team of visualization, interaction, analytic, perception, and cognition specialists and an interested party. When VA couples with DH, the interested parties will generally consist of a team of academics from a wide range of domains, such as linguistics, archeology, musicology, (art) history, or geography. What cohesively unites them as an interested party in this definition is their need for a visual analytical solution. The characterization of what a solution is or can be, however, may vary wildly according to each field’s research methodologies, goals, values, and individual experiences. This means that a team of VA specialists must be willing to immerse itself in the other domains’ perspectives, performing an epistemological drift [5], while still anchored to the pragmatics of data visualization.

Such interdisciplinary dynamics have been the subject of discussion both in the visualization and humanities communities. Among the challenges identified by the literature are: the nature of the data [3, 9], varying from textual sources that can be digital or physical, to archeological objects that allude digitization; determining the value of mixed contributions for each respective field [5, 7]; the faithfulness of visualizations to represent the subtleties of subjects [11, 16]; and the actual process of collaboration.

In this paper, we aim at contributing to this discussion by characterizing such challenges from a VA perspective. We introduce the Data-Users-Tasks design triangle [13] and re-purpose it to expose the complexities underlying interdisciplinary DH projects. We describe our experiences and discussions with researchers from various disciplines, including musicology and (art) history, and conclude with the lessons that we learned along the way.

*e-mail: victor.schetinger@tuwien.ac.at

†e-mail: k.raminger@extern.muk.ac.at

‡e-mail: velitchko.filipov@tuwien.ac.at

§e-mail: n.soursos@extern.muk.ac.at

¶e-mail: s.zapke@muk.ac.at

||e-mail: miksch@ifs.tuwien.ac.at

2 THE DATA-USERS-TASKS DESIGN TRIANGLE

The Data-Users-Tasks triangle (Fig. 3 A) is a methodological framework to aid the design process of VA projects [13], addressing three factors:

- **Data:** What kind of data are the users working with?
- **Users:** Who are the users of the VA solution?
- **Tasks:** What are the tasks of the use?

Its usefulness stems from its simplicity in characterizing both abstract and practical design requirements in three concepts. In the original formulation, these concepts are isonomic in their codependency: the user analyzes data to perform a task. Without a task, there is no reason to analyze the data. Without the user, there is no one to perform the analysis. Without data, there is nothing to be analyzed. This defines an interdisciplinary VA project through a subject-object-verb syntax, and the resulting sentence represents the shared vision, or contract between both teams.

In the sentiment of agile practices [1] the triangle can be resourcefully applied as a tool in many phases of conceptualization, design, and development. It structures discussion around three core topics and produces a minimalist object as output. It can be used to aid project initialization, facilitating communication between stakeholders [14], or to enforce scope. For example, the EVA project [10] uses VA to assist in the identification of fraudulent bank transactions. The data is a large pool of anonymized bank transactions, the users are financial experts, and the main task is to detect irregular events and patterns that could hide fraud.

The starting point of a project could be the desire to explore newly discovered data, which would put it at the forefront of discussion, followed either by who (users) would use it, how or why (tasks). It could also arise from the needs of a user, such as physician wanting to improve a diagnosis of a particular condition, and the data requirements come as a consequence of that. Generally, this process is much simpler in domains with objective and well-defined tasks and evaluation metrics, that can be straightforwardly adapted to prototypes and tested. This is not true for most DH projects, which deal with complex, openly defined problems and multi-faceted data with no ground truths [18]. Furthermore, the simplicity of the design triangle subtly hides a chicken-and-egg problem: what comes first, the user, the task, or the data? Each project has different priorities, and they might not be crisply clear from the beginning. While the output triangle can be thought of as isonomic, its definition process must have happened in an ordered process, for such is the nature of time.

The limitations of the design triangle for the context of digital humanities became clear to us while working in the Interacting Music Mapping Vienna (IMMV) project. While instances of the triangle with corresponding expected data, users, and tasks were present from the inception of the project, they failed to capture the vision and goals of the humanists. This paper is based on our experiences in trying to understand and overcome this failure.

Interactive Music Mapping Vienna

The starting point of the joint FWF-Peek-project Interactive Music Mapping Vienna: Exploring a city. 1945 up to the present day was the question of how music acts in the urban context of Vienna as a social identification instrument and how music is functionalized to urban symbolic politics. Vienna is associated to its music more than any other European city and has a strongly defined and longstanding identity constructed by means of music. IMMV aims to unravel the narrative of the Music City of Vienna (Musikstadt), by analyzing the basic types of music utilization (city tourism, city politics, image creation) [20], with a focus on public festivities.

In this interdisciplinary project we explore how VA approaches can be used to embed large amounts of historical and musicological data in space and time. VA provides a means to navigate and interact with multi-modal data in different granularities and levels of abstraction. The core of this project is the interplay between music as an artefact and urban identity and political symbolism. With VA we are making this interconnection accessible to a wide audience through interactive media and visual means.

From this description an initial setup for the triangle follows: data about public festivities is used to allow expert (musicologists, historians) and novice (students, the general public) users to explore narratives within the subject of the Music City of Vienna (task). This condenses the previous two paragraphs in a compact summary of the project, prioritizing the pragmatics of its output in terms of development. When contrasted with the triangle example for the EVA project [10], however, it is a more vague characterization. From the EVA triangle one can already idealize a solution, imagine how it could look like, and sketch it together with the stakeholders. This materialization of intention is an important step in development, advancing the dialectic. What could follow then is the implementation of this shared vision, with its own set of challenges and risks. In Munzner's nested model [15] methodology this would correspond to advancing from the domain characterization phase.

In the case of the IMMV triangle there is not enough resolution for a vision to be sketched. This is normal, as many projects start with a wide scope around a research subject, refining it over time. One way to go about resolving this is discussing each corner of the triangle individually, while assuming the other two fixed. For instance, assuming the same data and users, how could we improve the definition of the tasks? This exercise revealed to us over time that each concept was not isonomic at all. In our project tasks were elusive to define because of the chimeric nature of our data. Information about public festivities in Vienna over the decades were spread out in many shapes and places, both physical and digital sources, in public and private archives. Any projection of tasks assumed the filling of gaps in our data, which then revealed more inaccessible sources in dynamics reminiscent of Zenos' paradoxes: to move any distance forward one must first travel half of that distance, and to travel half distance one must first travel one quarter of the distance, and so on. Is there then any movement at all?

At the same time our data was holding the project back by hindering the definition of tasks, it was also the most valuable scientific product of efforts. By continuously expanding our database through the research and contextualization of sources it became an interesting artifact on its own. We realized that this process was a fundamental task by itself, that could not be separated from the development of interactive visualizations that draw on this data. Our definition of users, on the other hand, was almost orthogonal to this plane. The fact that both students and scholars were expected users of the solution seemed an almost independent factor considering the interplay between data and tasks. This led us to recognize the limitations of the data-users-tasks triangles for DH, and adapt it.

3 RE-PURPOSING THE DESIGN TRIANGLE

Before understanding the re-purposed triangle from the point of view of a VA scholar, one should abandon the methodological comfort zone it provides. It is rooted in a pragmatic, result-oriented approach that can get easily frustrating for both teams in a VA project due to the nature of problems involved. Here, data, users, and tasks take on a different ontological role so one can better understand the dynamics of DH projects. The goal is not to cut through steps or accelerate the development of a final product, but to provide grounding and perspective.

The transformation is a two-step process, where the first step is to break the isonomy of its elements (Fig. 1 B): all parts are fundamental, but some are more fundamental than others. With just

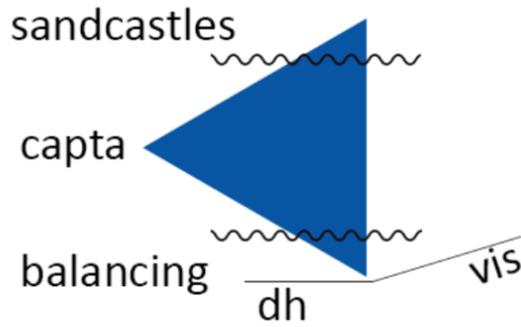


Figure 2: The three triangle divided in its three realms of problematics, with the balancing act at the bottom. Each team holds one end of the rope, and struggles to balance the triangle to the users tip.

a slight counter-clockwise rotation, a vertical hierarchy is formed, with the tasks at the top and the users at the base, boldly supporting the structure against the weight of data. The next step, then, is to re-contextualize each concept, relating them to the challenges within DH (see Fig. 2). This exposes three different realms within a project: on the basis, the realm of balancing; in the middle, the realm of capta, and the realm of sandcastles at the top. Each of them is related to a vertex in the triangle, having the purpose of expanding its meaning and highlight its core aspect.

3.1 Users and the realm of balancing

The realm of balancing is the bottom tip of the triangle, and relates to the concept of users. It is inspired in the metaphor of the balancing act [7]. One must imagine a rope with the team of VA specialists holding one side, and digital humanists holding the other: on top of it, the triangle is being balanced by the users tip (Fig. 2). This represents the fact that all decisions in a project are evaluated through the projections of its participants about who are the end users and what is important for them. A developer with lack of flexibility or experience might project his or herself as an end user, and pull the project towards his/her own personal vision and biases. This could result in an excessive focus on implementing certain features (s)he finds important, or using technology within his or her comfort zone.

The balancing act can easily turn into a tug-of-war between teams, especially if conflicts of interest arise in academic goals. As Jänicke highlights, fields have different criteria for what is considered a contribution or valid research output: papers, software, books. Furthermore, while this bipartisan view of a project captures many defining aspects, the influence of individuals, their interests and experiences is what really matters [5]. Humans are holding the ends of the rope.

Conflicts may arise within the same side of an argument: a historian and an art historian might want to focus on different aspects of the same subject. While the former could be interested on the chronology and political narrative, the latter would give higher importance to aesthetics. Resolving such conflicts and turning them into implementable features would require commitment and imagination from the side of developers. There must be a constant effort from all individuals involved to put themselves in others' shoes.

3.2 Data and the realm of capta

According to humanistic research principles, there is no self-evident truth or observer-independent reality. Humanistic research is based on the general idea that physical, cultural and social phenomena of

our reality are not naturally given but constructed and shaped by certain interests and power relations. As a consequence, humanistic knowledge production is in itself a subjective and interpretative act.

Historical research as in the case of IMMV heavily relies upon heterogeneous physical sources of different kinds of media. Visual knowledge production as performed by VA based on data drawn from historical sources is thereby subject to several challenges. Simply transferring (analog) sources into immaterial data would be the first to name. In our experience, using a database for modeling and structuring historical data according to standard metrics entails certain limitations as it forces historians to reduce complexity, uncertainty, inconsistency and ambiguity inherent not only in historical data but in history itself. This inevitably leads to a decline of reflection and nuance in the representation of History in favor of abstractions, simplifications and generalizations.

Another challenge that must be dealt with properly is missing data (due to deliberate destruction or loss of source material and humanistic knowledge) or inaccessible sources (due to archival restrictions or copyright issues) and therefore incomplete datasets. Further, many DH-projects such as IMMV for several reasons do not even have in mind the production of a complete dataset, but rather aim at the reflection of a certain narrative and the presentation of storytelling. In this case, there are no seemingly objectified, self-evident datasets that can be explored and analyzed, but instead VA has to cope with subjectively curated and interpreted data. Acknowledging these challenges as well as the general principles of humanistic research as outlined above, when re-evaluating the triangle, we decided to use the term capta instead of data. The term has been coined by Johanna Drucker, who in her definition of capta states that: Capta is taken actively while data is assumed to be a given able to be recorded and observed. From this distinction, a world of differences arises. Humanistic inquiry acknowledges the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, taken, not simply given as a natural representation of pre-existing fact. [3].

The realm of capta, then, represents the constant construction of knowledge within a project: the acquisition of information, its interpretation, transformation, and projection in the form of hypotheses. Furthermore, experience acquired through interaction, successes and failures constitutes essential capta for the project. The division of the triangle in three areas (see Fig. 1) intentionally leaves the larger area within the realm of capta, as indeed most work takes place within it. For the sake of clarity of exposition figure 1 maintains the original term data in the triangle, but sets it inside the realm of capta. The final form of the triangle is presented in figure 2.

3.3 Tasks and the realm of sandcastles

Tasks are the most elusive aspect to define in digital humanities projects. From the side of digital humanists VA task taxonomies might seem either unexpressive or too constraining to contribute to the construction of meaning within the project. From the side of VA experts, tasks only become clear after a certain level of maturity and involvement with the problem. A virtual, idealized task at the end of all projects could be thought of as answering the research questions, even if it takes time to define a research question. For this reason it sits at the top of the triangle, aiming at an idealized convergence at the tip.

This is the realm of sandcastles, as we defend Hinrichs' position [6] that visualization should be treated as a research process in its own right. It represents the development of visualization sandcastles, that gradually stack to elevate the project to its apex. Instead of tackling the main research question head-on the team must gradually work its way through the realm of capta by developing smaller visualizations in iterative cycles. This exercise allows researchers from both sides to notice epistemological gaps and breach them. In its time, the team can develop its own language to express tasks in

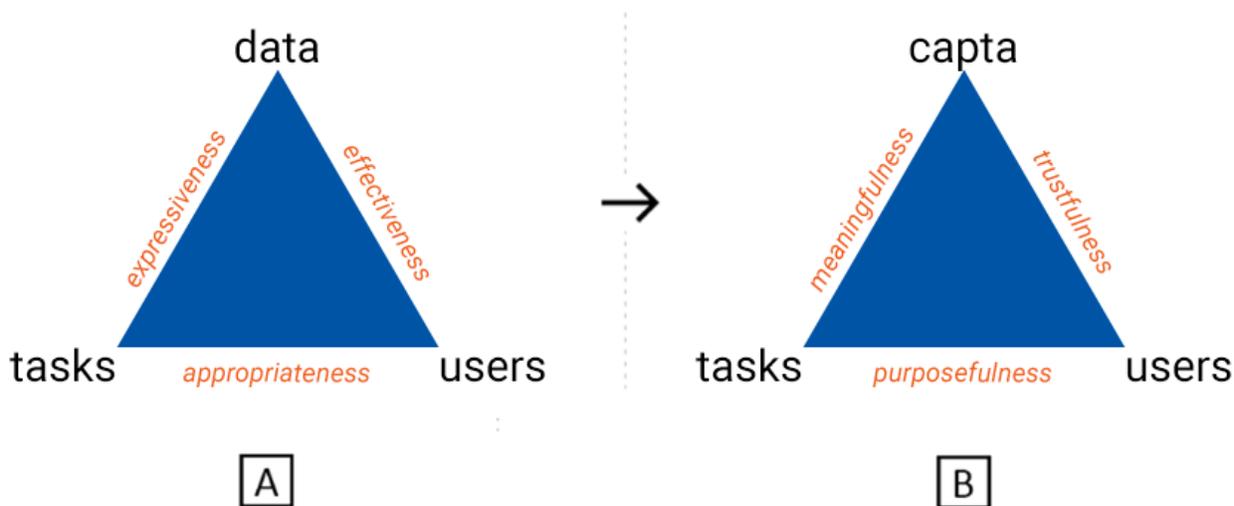


Figure 3: The quality criteria for the original data-users-task design triangle (A), and the re-purposed quality criteria. Data changes to capta, expressiveness changes to meaningfulness, effectiveness changes to trustfulness, and appropriateness changes to purposefulness.

terms of shared experiences and available capta.

3.4 The Quality Criteria

While data, users, and tasks correspond to the vertices of the triangle, the edges represent the relationship between each of them. They complement the methodological purpose of the triangle by their association with important quality criteria for visualization (Fig. 3): Expressiveness refers to the requirement of showing exactly the information contained in the data. Effectiveness concerns to the degree to which the visualization addresses the cognitive capabilities of the human visual system and the context of the user. Appropriateness tries to quantify the cost-value ratio of the benefit of the visualization process with respect to achieving the intended task.

Such criteria reflect a pragmatic and result-oriented approach to the VA design process, and philosophical incompatibilities with the humanities are evident. Expressiveness, for instance, enforces a reductionism of reality to the bounds of the information contained in the data. To fully repurpose the triangle, we propose three different quality criteria that take into consideration the context of digital humanities and the metaphors constructed so far in its exposition: trustfulness, purposefulness, and meaningfulness.

Trustfulness is the relation between capta and users, and reflects the degree to which it can provide guarantees of its faithfulness [18] within the epistemological framework of its domain. In other words, it ties the visualization to the epistemological rigor of the projected users, in whatever field they are. The domain experts have to ask themselves: would my peers trust this visualization?

Purposefulness is the relationship between users and tasks. In the balanced triangle (see Fig. 3) it represents the axis orthogonal to the rope, ascending from the grounding tip to the apex. One could imagine this line as the envisioned plan, what's on paper, such as a written grant proposal. This is the *raison d'être* of a project. In the case of IMMV, it is to unravel the narrative of the Music City of Vienna. As a project develops and shifts focus the whole team must be mindful of the central purpose.

Meaningfulness is the relationship between tasks and capta, and expresses the potential value of the developed visualizations in terms

of construction of meaning. How good and effective are they in generating new insights from capta and, conversely, even more capta.

This completes our re-purposing of the design triangle, where every vertex and every edge is accounted for and re-signified. Furthermore, each realm relates an area of the triangle to an already existing concept within the digital humanities literature.

4 CONCLUSION

In this position paper we have presented a well-known methodological aid from VA and used it to contextualize different challenges involved in digital humanities projects. By providing a perspective on the pragmatics of VA, exposing its problems in interfacing with humanists, and proposing a re-contextualization, we hope to provide valuable insight to the community. We really believe the effort to continually redefine ones epistemological standing is essential to interdisciplinary collaboration. This paper is an exercise in bridge building through re-thinking the common methodological foundations of VA, and we expect to further this discussion in the future.

ACKNOWLEDGMENTS

This work was conducted within the framework of the project "Interactive Music Mapping Vienna" (AR384-G24) funded by the Austrian Science Fund (FWF), Programme for Arts-based Research (PEEK) and lead by Susana Zapke (Music and Art, University of the City of Vienna).

REFERENCES

- [1] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. Martin, S. Mellor, K. Schwaber, J. Sutherland, and D. Thomas. Manifesto for agile software development, 2001. Accessed 8th July 2019.
- [2] S. Campbell, Z.-y. Yu, S. Connell, and C. Dunne. Close and Distant Reading via Named Entity Network Visualization: A Case Study of Women Writers Online. p. 5, 2018.
- [3] J. Drucker. Data as capta. *Digital Humanities Quarterly*, 5(1):1–21, 2011.

- [4] M. El-Assady, V. Gold, M. John, T. Ertl, and D. A. Keim. Visual text analytics in context of digital humanities. *Work. Vis. Digit. Humanit. as part IEEE VIS*, pp. 1–10, 2016.
- [5] U. Hinrichs, M. El-Assady, A. J. Bradley, S. Forlini, and C. Collins. Risk the drift! stretching disciplinary boundaries through critical collaborations between the humanities and visualization. 2017.
- [6] U. Hinrichs and S. Forlini. In defense of sandcastles: Research thinking through visualization in dh. In *Proceedings of the conference on Digital Humanities*. International Alliance of Digital Humanities Organizations (ADHO), 2017.
- [7] S. Jänicke. Valuable research for visualization and digital humanities: A balancing act. In *Workshop on Visualization for the Digital Humanities, IEEE VIS 2016, Baltimore, Maryland, USA*, 2016.
- [8] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, eds. *Mastering the information age : solving problems with visual analytics*. Goslar : Eurographics Association, 2010.
- [9] H. Lamqaddam, K. Brosens, F. Truyen, R. J. Beerens, I. De Prekel, and K. Verbert. When the tech kids are running too fast: Data visualisation through the lens of art history research. *IEEE Transactions on Visualization and Computer Graphics*, 2018.
- [10] R. Leite, T. Gschwandtner, S. Miksch, S. Kriglstein, M. Pohl, E. Gstrein, and J. Kuntner. Eva: Visual analytics to identify fraudulent events. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 08 2017.
- [11] E. Mayr, N. Hynek, S. Salisu, and F. Windhager. Trust in information visualization. 06 2019.
- [12] F. McGee, M. During, and M. Ghoniem. Towards visual analytics of multilayer graphs for digital cultural heritage. *Towards Visual Analytics of Multilayer Graphs for Digital Cultural Heritage*, 2016.
- [13] S. Miksch and W. Aigner. A matter of time: Applying a data-users-tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics*, 38:286–290, 2014.
- [14] A. Mosca, S. Robinson, M. Clarke, R. Redelmeier, S. Coates, D. Cashman, and R. Chang. Defining an Analysis: A Study of Client-Facing Data Scientists. In J. Johansson, F. Sadlo, and G. E. Marai, eds., *EuroVis 2019 - Short Papers*, pp. 73–77. The Eurographics Association, 2019.
- [15] T. Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, Nov. 2009.
- [16] J. Thomas and J. Kielman. Challenges for visual analytics. *Information Visualization*, 8:309–314, 12 2009.
- [17] J. J. Thomas and K. A. Cook. A visual analytics agenda. *IEEE Comput. Graph. Appl.*, 26(1):10–13, Jan. 2006.
- [18] H. van den Berg, A. Betti, T. Castermans, R. Koopman, B. Speckmann, K. Verbeek, T. van der Werf, S. Wang, and M. Westenberg. A philosophical perspective on visualization for digital humanities. In *Proc. 3rd Workshop on Visualization for the Digital Humanities (VIS4DH)*, 2018.
- [19] P. van Wierst, S. Hofstede, Y. Oortwijn, T. Castermans, R. Koopman, S. Wang, M. A. Westenberg, and A. Betti. Bolvis: visualization for text-based research in philosophy. In *3rd Workshop on Visualization for the Digital Humanities*, 2018.
- [20] S. Zapke. *Mapping the Music City in the 20th century. Theoretical Approaches*, p. 73. Unravelling Complexities. Understanding Public Space (IRS-Forschungsberichte 44), Wien 2017: Verlag der sterreichischen Akademie der Wissenschaften, 2017.