

Cooperative Digital Humanities: A Methodology

Mohammad Alharbi¹, Tom Cheesman¹, and Robert S. Laramee^{1,2}

¹ Swansea University, UK

{m.alharbi.508205, t.cheesman, r.s.laramee@swansea.ac.uk}

² University of Nottingham, UK

{robert.laramee@nottingham.ac.uk}

Abstract. Researchers in the digital humanities use visualization with increasing popularity to address their challenges. However, interdisciplinary and collaborative projects with visualization researchers are associated with various and common research challenges, such as collaborative communications and methodological differences. Different strategies have been proposed to guide and steer general cooperative projects to realize the common team objectives. In this paper, we propose a methodological workflow for interdisciplinary digital humanities and visualization research based on our previous work and experience. Our methodological workflow consists of three spaces, three channels, and three criteria. The three spaces feature the main collaborative entities: problem, task, and solution spaces. The three channels illustrate the connections between spaces and include communication, pre-visualization, and evaluation channels. The three quality criteria include expressiveness, purposefulness, and trustfulness. These three criteria are included to ensure useful outcomes from each space. In each section of the workflow, we draw from our previous collaborative work to demonstrate the effectiveness of the workflow.

Keywords: First keyword · Second keyword · Another keyword.

1 Introduction and Motivation

Digital humanities scholars increasingly adopt visualization approaches to enrich their research. They find that visual interfaces create new modes of knowledge generation and facilitate more effective discovery of new observations [1, 13, 15]. Hinrichs and Forlini [13] claim that visualization should be considered not just a means to an end but as a research process in its own right, which has led to the development of multiple interdisciplinary collaborations between the digital humanities and visualization communities. These collaborations have also been studied and discussed in both communities in order to identify means to enhance the collaborations and discuss the challenges encountered [11, 15, 26].

We believe that cooperation ideally follows a conceptual workflow that considers all aspects of collaboration if possible. Without this, collaboration may develop in undesirable directions due to the different perspectives each stakeholder brings to the project.

In this paper, based on our previous collaborative work, we contribute a methodological workflow of the collaboration between interdisciplinary projects. The goal of this workflow is to guide the collaborative work between digital humanities and visualization. The workflow also aims to foster more effective interdisciplinary research which integrates all of the discipline involved in a balanced manner. The workflow via the quality criteria ensures the usefulness of the results obtained in each space.

The rest of this paper is organized as follows: In Section 2, we present previous research related to digital humanities collaborative research. Section 3 introduces our proposed conceptual workflow based on our collaborative experience. Finally, Section 4 concludes our paper and points our future directions.

2 Related Work

Collaboration between the visualization team and digital humanities for interdisciplinary visualization projects has been the subject of significant discussion. Recent developments in interdisciplinary research highlight challenges in digital humanities projects and encourage research to propose a collaborative framework to address these challenges [10, 12, 29]. Munzner [21] proposes a general nested model that guides the process of design and evaluation of visualization projects, while Kath et al. [17] propose a methodological framework supporting knowledge generation of collaborative projects using visualizations. Simon et al. [27] suggest the liaison role shares knowledge and language with both domains to foster collaborative communication. El-Assady et al. [11] present a conceptual workflow of the problem-solving process and collaboration in digital humanities projects with visual text analytics. Jänicke et al. [15] discuss collaboration themes, including the initial start of projects, development iterations, and evaluation methods. Roberts et al. [23] discuss a similar process on the collaboration between academia and industry in visualisation projects, discussing the nature of such projects and how knowledge transfers between the two parties throughout an interview study. Most recently, Schetinger et al. [25] introduce a re-purposed framework of the Data-Users-Tasks triangle [20] to overcome limitations in the context of digital humanities.

In this paper, we provide a methodological workflow based on our previous collaboration with digital humanities. The approach combines the three most important aspects: domain, tasks, and design spaces. It also integrates quality criteria to ensure useful outcomes.

3 Collaborative Workflow

The proposed workflow is a conceptual workflow (Figure 1) to inform the collaboration and design of interdisciplinary visualization projects. It features three spaces, three channels, and three criteria. The following section discusses the workflow components and how they complement one another.

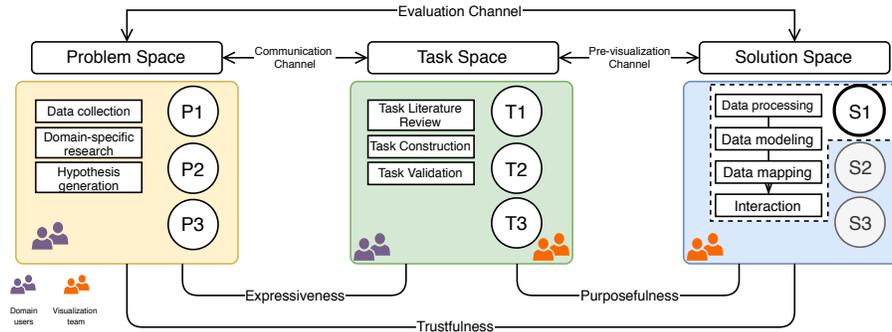


Fig. 1. Our proposed methodological and interdisciplinary workflow. The workflow consists of three main components: the domain, task, and solution spaces. The tasks are informed by a communication channel between the two users’ groups. A pre-visualization channel is attempted between the task and solution spaces to prepare for implementation. In the solution space, one or multiple solutions are implemented to address the predefined tasks. The terms expressiveness, purposefulness, and trustfulness indicate the quality criteria that need to be fulfilled to obtain useful outcomes (Section 3.2).

3.1 Three Spaces and Three Channels

This section consists of a discussion of our workflow. The word “channel” is used to illustrate the connective phases between spaces as they usually involve communication between the two users in the workflow.

Problem Space: The problem space is the starting point of the workflow. It essentially resembles the domain users, the data, and more likely a set of challenges. For example, in the TransVis project [4], a collection of German translations of Shakespeare’s *Othello* was curated by the domain expert in order to be analyzed and visualized. Exploring and examining the collection without computational and visual aids is a laborious and challenging process for digital humanities scholars. The domain users are usually interested in studying how existing approaches can solve their problems, and they generate hypotheses to be confirmed and evaluated based on their data. In this space, the domain problems are clearly identified. Each problem statement needs to be unambiguous, focused, concise, complex, and arguable [11].

Communication Channel: This channel plays a vital role when collaborating on interdisciplinary projects. In our collaboration, we think of this communication as an educational experience for both domain scholars and visualization teams. The domain scholars strive to understand computational and visual tools as much as possible. This understanding and awareness increase the chances of designing helpful solutions. This means that the visualization team explains the fundamentals and does not assume that they are self-explanatory. The visualization team also strives to understand the domain data and problems, and is

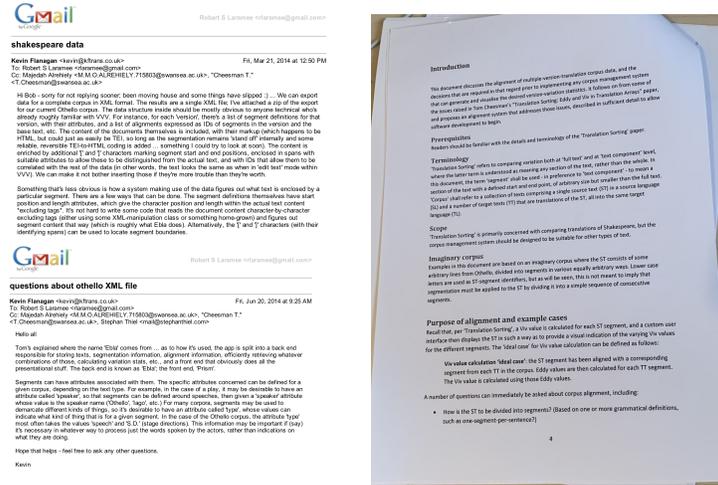


Fig. 2. On the left, email that was exchanged in the early stages of the collaboration to enable sufficient understanding of domain-specific terminology. On the right, a screenshot of a formal document that explains the dataset components and terminology.

encouraged to participate in domain readings and discussions which can help discover relevant mutual problems [1]. The development of such common knowledge can be complex, and we suggest constructive regular meetings at the early stages of collaboration to bridge the differences between the two domains. Simon et al. [27] propose a liaison role in the workflow who shares sufficient knowledge in both disciplines to foster more effective interdisciplinary communication and contributes to the project by capturing the problem complexity or mental model. We also suggest documenting a glossary of terminology that define the key terms in the domain area. In the early stages of our collaboration, various email was exchanged and sessions were held to create sufficient understanding of the dataset element and its associated terminology (Figure 2).

In this communication phase, flexibility is an essential skill as the discussions strive for balance between the two disciplines. If the visualization team focuses more on the implementation and computational side, it might result in failure to deliver useful solutions. Additionally, what each discipline considers a contribution may vary and this could take the project in an undesired direction if the initial communications and discussions are not balanced [25].

Task Space: In this space, the tasks are formulated based on the research problems (gaps) and interdisciplinary discussions, and are expressed differently between domains. One problem could result in one or more tasks to be solved. The domain scholar might have broad, high-level tasks, such as close or distant reading, while the visualization team is responsible for transferring these tasks into more technical, well-expressed tasks. The tasks are complete, discriminative, objective, and measurable [30]. Although this is not always achievable, it

	TransVis [4]	AlignVis [2]	VNLP [3]
Data processing	data cleaning, integration, tokenization, normalization, feature extractions		
Data modeling	Eddy and Viv analysis	Similarity computation	Embedding analysis, similarity computation
Data mapping	Segments colors	Segments and edges colors	Histogram, bar charts, etc
Interaction	multiple sorting options, Filtering and selection	Confidence threshold, Filtering and selection	Overview similarity results, Customizable pipeline items

Table 1. Example representatives of the results of the implementation stages in the solution space that correspond to our contributions, TransVis [4], AlignVis [2], and VNLP [3].

is nevertheless attempted. In a previous collaboration [4], we adapted the detailed Brehmer and Munzner [7] typology of visualization tasks, which can be communicated to the domain scholar in order to abstract user tasks.

Pre-visualization Channel: This channel is usually where the visualization team parts ways with the domain scholars. The main goal of this channel is to study user tasks and data and begin implementing visual solutions. Here, the visualization team surveys the design space of existing approaches in order to explore potential design solutions and carefully study their advantages and limitations [30]. It is also beneficial to study the domain specific tools because the use of visualization is becoming an essential element of research [13]. The main properties of the activities in this channel are that they involve iterative sketching and trials, and it is crucial to communicate the results to the domain scholar and validate appropriateness against the tasks specified.

Solution Space: Implementing a visual solution starts with data transformation. Often, the data that comes from the domain suffers from a number of problems and may come from a variety of sources with different formats or conventions. Therefore, the data must be preprocessed in order to be cleaned, integrated, and prepared for the next stage. In data modeling, the data is analyzed and interesting meta-data derived, such as Eddy and Viv [4], alignment detection [2], and embeddings generation (VNLP). In data mapping, the abstracted data is mapped to visual encodings. Lastly, user interaction is implemented to aid exploratory analysis. Table 1 shows example representatives of the results of the implementation stages that correspond to our contributions, TransVis [4], AlignVis [2], and VNLP [3].

Based on our previous collaborations, we recommend implementing prototypes iteratively with a subset of the data and presenting the results to the domain scholars. Such frequent presentations and discussions help satisfy the user tasks, obtain intuitive results, and increase the domain scholar’s engagement [16].

Evaluation Channel: Evaluating the efficacy and usability of the visual solution is an essential goal of any interdisciplinary project. However, many visualization approaches lack an in-depth, effective quantitative or qualitative evaluation [5].

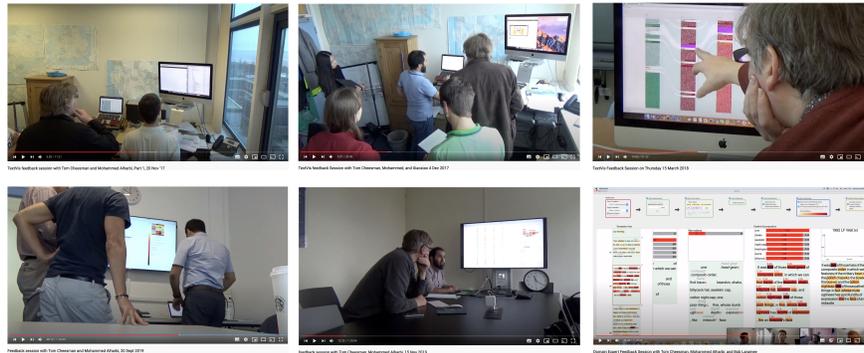


Fig. 3. Samples of different recordings of our domain expert feedback sessions.

Furthermore, humanities scholars tend to doubt and question computational, qualitative evaluation. A lack of ground truth is one of the most common challenges in digital humanities [29]. Jänicke et al. [15] report that there are more visual approaches for text analysis tasks published in digital humanities than in the visualization communities due to the usual demand of quantitative evaluations which are challenging to incorporate as a result of the limited number of collaborators from the humanities. Munzner [21] provides guidance on evaluation methods for different design choices. Lam et al. [18] provide a scenario-based method to study evaluation for information visualization. They introduce seven scenarios derived through an extensive literature review of over 800 visualization publications. There has also been work on evaluating visualization which guides users on how to carry out an evaluation for information visualization [9,19,22,28].

In our collaboration with the domain scholar, we evaluate our project usability obtaining domain expert feedback and conducting use cases. The domain expert feedback is based on regular sessions to demonstrate the design features. All of the sessions are video-recorded for post-analysis and archiving. Figure 3 shows a selection of feedback session recordings of our collaboration with the domain expert. Semi-structured interview questions are planned and guided by Hogan et al [14]. The early sessions usually consist of mock-ups, sketches, or software demonstrations to guide the development of features, and gradually become active hands-on use of the software by the domain expert. During the sessions, the software evolves due to feature demands. During the face-to-face feedback sessions, patterns can be observed, such as the discovery of software bugs and data-level errors.

3.2 Quality Criteria

The design triangle (data–users–tasks)(Figure 4) methodological approach to inform the design of interactive visualizations suggests three quality criteria that need to be fulfilled in order to obtain useful outcomes [20]. Expressiveness refers

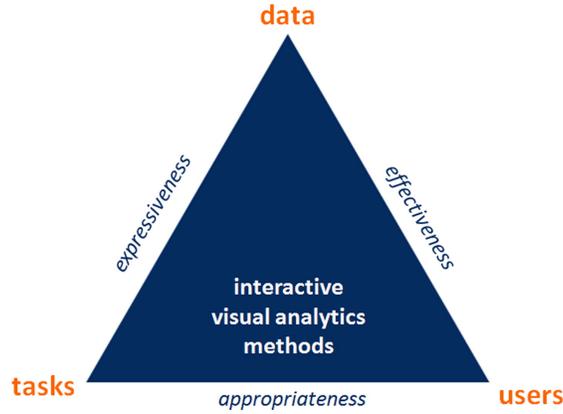


Fig. 4. The design triangle by Miksch and Aigner [20]. They include factors to be considered during the design and implementation of interactive visualizations. Image courtesy of Miksch and Aigner [20].

to the requirement of conveying the information contained in the data, effectiveness concerns the degree to which the visualization addresses the cognitive capabilities of the human visual system and the context of the user, and appropriateness quantifies the cost-value ratio of the benefit of the visualization process with respect to achieving the intended task. Schetinger et al. [25] repurpose the design triangle and propose three quality criteria that take into consideration the context of digital humanities. Trustfulness reflects the degree to which it can provide guarantees of faithfulness within the epistemological framework of its domain, purposefulness is similar to appropriateness and represents the relation between users and tasks, and meaningfulness expresses the potential value of the custom-made visualization software in terms of generating new insight from the data.

Our workflow consists of domain, task and solution spaces. We adopt similar quality criteria that need to be satisfied in order to obtain the most useful results: expressiveness, purposefulness, and trustfulness (Figure 1).

Expressiveness refers to the relation between the problem space and task space. It has general and specific aspects. The general aspect is that interdisciplinary exchange and communication can be challenging [11, 27], so a glossary of terminology can be adopted and all researchers involved are clearly established. In the specific aspect, the tasks must be well-expressed. Different well-established typologies of task abstractions [6, 8] can be utilized to establish a well-defined and expressed task and requirement analysis.

Purposefulness refers to evaluating the visualization against the given tasks. If the requirement and task analysis are optimally defined, this criterion can be quantified. It also important also to consider alternative solutions and how they would achieve the same tasks if they could.

Trustfulness refers to the relation between the solution and the expert user and to what extent they trust the result. Visualizations designed for digital humanities tend to exhibit black-box behavior (not transparent). Rieder and Röhle [24] define transparency as “our ability to understand the method, to see how it works, which assumptions it is built on, to reproduce it, and to criticize it”. Based on this, overcoming the lack of transparency is a challenge. The results of modeling and machine learning algorithms are often difficult to interpret and backtrace. Additionally, visualization tends to reduce informational dimensions to produce a focus that shows certain perspectives or interpretations of the data [29]. Based on our collaboration, the domain users do not appreciate this and struggle to trust such results until they understand how they are derived, which is in most cases very difficult. In our collaboration, we keep a close connection with the domain user in the early stages and validate the visual approach with a subset of the data that they know. The user evaluates the result based on the input data. If this visual approach is deemed faithful to the data and domain knowledge, we test the tool with a larger subset of the data.

4 Conclusion

This paper propose a methodological workflow for collaborative research with digital humanities, introducing three spaces, three channels, and three criteria to guide the collaboration in order to produce visualization solutions. The spaces characterize the domain, task, and solution aspects of the project. The channels illustrate the three communicative means between spaces: the communication channel between the problem space and the task space, the pre-visualization channel between the task space and the solution space, and the evaluation space between the solution space and the problem space. The three criteria (expressiveness, purposefulness and trustfulness) are essential to obtain useful outcomes between each space. For the future work, we would like to to apply our methodological workflow to other real-world interdisciplinary research projects.

References

1. Abdul-Rahman, A., Lein, J., Coles, K., Maguire, E., Meyer, M., Wynne, M., Johnson, C.R., Trefethen, A., Chen, M.: Rule-based visual mappings—with a case study on poetry visualization. In: Proc. of the 15th Eurographics Conference on Visualization (EuroVis). p. 381–390 (2013). <https://doi.org/10.1111/cgf.12125>
2. Alharbi, M., Cheesman, T., Laramée, R.: AlignVis: Semi-automatic Alignment and Visualization of Parallel Translations. In: 24th International Conference on Information Visualisation (IV). pp. 98–108 (2020). <https://doi.org/10.1109/IV51561.2020.00026>
3. Alharbi, M., Cheesman, T., Laramée, R.: VNLP: Visible natural language processing (2020), submitted
4. Alharbi, M., Cheesman, T., Laramée, R.S.: TransVis: Integrated Distant and Close Reading of Othello Translations. IEEE Transactions on Visualization and Computer Graphics pp. 1–1 (jul 2020). <https://doi.org/10.1109/tvcg.2020.3012778>

5. Alharbi, M., Laramee, R.S.: SoS TextVis: An Extended Survey of Surveys on Text Visualization. *Computers* **8**(1), 17–35 (2019)
6. Amar, R., Eagan, J., Stasko, J.: Low-level components of analytic activity in information visualization. In: *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005*. pp. 111–117 (2005). <https://doi.org/10.1109/INFVIS.2005.1532136>
7. Brehmer, M., Munzner, T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics* **19**(12), 2376–2385 (2013)
8. Brehmer, M., Munzner, T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics* **19**(12), 2376–2385 (2013)
9. Carpendale, S.: *Evaluating Information Visualizations*, pp. 19–45. Springer Berlin Heidelberg, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70956-5_2
10. Deegan, M., McCarty, W.: *Collaborative Research in the Digital Humanities: A Volume in Honour of Harold Short, on the Occasion of His 65th Birthday and His Retirement, September 2010*. Ashgate Publishing Company, USA (2012)
11. El-Assady, M., Gold, V., John, M., Ertl, T., Keim, D.A.: Visual text analytics in context of digital humanities. In: *1st IEEE VIS Workshop on Visualization for the Digital Humanities as part of the IEEE VIS 2016* (2016), <https://scibib.dbvis.de/publications/view/686>
12. Gold, M.K.: *Debates in the digital humanities*. University of Minnesota Press (2012)
13. Hinrichs, U., Forlini, S., Moynihan, B.: In defense of sandcastles: Research thinking through visualization in digital humanities. *Digital Scholarship in the Humanities* **34**(1), i80–i99 (10 2018). <https://doi.org/10.1093/llc/fqy051>
14. Hogan, T., Hinrichs, U., Hornecker, E.: The elicitation interview technique: Capturing people’s experiences of data representations. *IEEE Transactions on Visualization and Computer Graphics* **22**(12), 2579–2593 (2016). <https://doi.org/10.1109/TVCG.2015.2511718>
15. Jänicke, S., Franzini, G., Cheema, M.F., Scheuermann, G.: Visual Text Analysis in Digital Humanities. *Computer Graphics Forum* **36**(6), 226–250 (sep 2017). <https://doi.org/10.1111/cgf.12873>, <http://doi.wiley.com/10.1111/cgf.12873>
16. Jänicke, S., Geßner, A., Franzini, G., Terras, M., Mahony, S., Scheuermann, G.: Traviz: A visualization for variant graphs. *Digital Scholarship in the Humanities* **30**(suppl.1), 83–99 (10 2015). <https://doi.org/10.1093/llc/fqv049>, <https://doi.org/10.1093/llc/fqv049>
17. Kath, R., Schaal, G.S., Dumm, S.: New visual hermeneutics. *Zeitschrift für germanistische Linguistik* **43**(1), 27–51 (2015)
18. Lam, H., Bertini, E., Isenberg, P., Plaisant, C., Carpendale, S.: Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics* **18**(9), 1520–1536 (2012). <https://doi.org/10.1109/TVCG.2011.279>
19. Laramee, R.S.: How to write a visualization research paper: A starting point. *Computer Graphics Forum* **29**(8), 2363–2371 (2010). <https://doi.org/https://doi.org/10.1111/j.1467-8659.2010.01748.x>
20. Miksch, S., Aigner, W.: A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics* **38**, 286 – 290 (2014). <https://doi.org/https://doi.org/10.1016/j.cag.2013.11.002>, <http://www.sciencedirect.com/science/article/pii/S0097849313001817>

21. Munzner, T.: A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics* **15**(6), 921–928 (2009). <https://doi.org/10.1109/TVCG.2009.111>
22. Plaisant, C.: The challenge of information visualization evaluation. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. p. 109–116. AVI '04, Association for Computing Machinery, New York, NY, USA (2004). <https://doi.org/10.1145/989863.989880>, <https://doi.org/10.1145/989863.989880>
23. Roberts., R., Laramée., R., Brookes., P., Smith., G.A., D’Cruze., T., Roach., M.J.: A tale of two visions - exploring the dichotomy of interest between academia and industry in visualisation. In: *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: IVAPP*. pp. 319–326. INSTICC, SciTePress (2018). <https://doi.org/10.5220/0006635803190326>
24. Röhle, B.R.T.: *Digital Methods: Five Challenges*, pp. 67–84. Palgrave Macmillan UK, London (2012). https://doi.org/10.1057/9780230371934_4
25. Schetinger, V., Raminger, K., Filipov, V., Soursos, N., Zapke, S., Miksch, S.: Bridging the gap between visual analytics and digital humanities: Beyond the data-users-tasks design triangle (2020), eingeladen; Vortrag: 4th Workshop on Visualization for the Digital Humanities, Vancouver, Canada; 2020-10-20
26. Silvia, S., Etemadpour, R., Abbas, J., Huskey, S., Weaver, C.: When the tech kids are running too fast: Data visualisation through the lens of art history research. In: *Proceedings of the Workshop on Visualization for the Digital Humanities*. IEEE, Berlin, Germany (October 2018)
27. Simon, S., Mittelstädt, S., Keim, D.A., Sedlmair, M.: Bridging the Gap of Domain and Visualization Experts with a Liaison. In: Bertini, E., Kennedy, J., Puppo, E. (eds.) *Eurographics Conference on Visualization (EuroVis) - Short Papers*. The Eurographics Association (2015). <https://doi.org/10.2312/eurovisshort.20151137>
28. Tory, M., Moller, T.: Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics* **10**(1), 72–84 (2004). <https://doi.org/10.1109/TVCG.2004.1260759>
29. Van Den Berg, H., Betti, A., Castermans, T., Koopman, R., Speckmann, B., Verbeek, K., Van der Werf, T., Wang, S., Westenberg, M.A., et al.: A philosophical perspective on visualization for digital humanities. In: *Proc. 3rd Workshop on Visualization for the Digital Humanities (VIS4DH)* (2018)
30. van Wijk, J.J.: Views on visualization. *IEEE Transactions on Visualization and Computer Graphics* **12**(4), 421–432 (2006). <https://doi.org/10.1109/TVCG.2006.80>