ISSN 2186-7437

NII Shonan Meeting Report

No.167

Formalizing Biological and Medical Visualization

Renata Georgia Raidou
Barbora Kozlíková
Johanna Beyer
Timo Ropinski
Issei Fujishiro

February 24–27, 2022

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Formalizing Biological and Medical Visualization

Organizers:
Renata Georgia Raidou (TU Wien, Austria)
Barbora Kozlíková (Masaryk University, Czech Republic)
Johanna Beyer (Harvard University, USA)
Timo Ropinski (Ulm University, Germany)
Issei Fujishiro (Keio University, Japan)

February 24–27, 2022

Abstract

Medicine and biology are among the most important research fields, having a significant impact on humans and their health. For decades, these fields have been highly dependent on visualization—establishing a tight coupling which is crucial for the development of visualization techniques, designed exclusively for the disciplines of medicine and biology. These visualization techniques can be generalized by the term Biological and Medical Visualization—for short, *BioMedical Visualization*. BioMedical Visualization is not only an enabler for medical diagnosis and treatment, but also an influential component of today's life science research. Many BioMedical domains can now be studied at various scales and dimensions, with different imaging modalities and simulations, and for a variety of purposes. Accordingly, BioMedical Visualization has also innumerable contributions in industrial applications. However, despite its proven scientific maturity and societal value, BioMedical Visualization is often treated within Computer Science as a mere application subdomain of the broader field of Visualization.

To enable BioMedical Visualization to further thrive, it is important to *formalize its characteristics independently* from the general field of Visualization. Also, several lessons learnt within the context of BioMedical Visualization may be *applicable and extensible* to other application domains or to the parent field of Visualization. Formalization has become particularly urgent, with the latest advances of BioMedical Visualization—in particular, with respect to dealing with Big Data Visualization, e.g., for the visualization of multi-scale, multi-modal, cohort, or computational biology data. Rapid changes and new opportunities in the field, also regarding the incorporation of Artificial Intelligence with "human-in-the-loop" concepts within the field of Visual Analytics, compel further this formalization. By enabling the BioMedical Visualization community to have intensive discussions on the systematization of current knowledge, we can adequately prepare ourselves for future prospects and challenges, while also contributing to the broader Visualization community.

During this *4-day seminar*, which was the *150th NII Shonan meeting* to be organized, we brought together 25 visualization experts from diverse institutions, backgrounds and expertise to discuss, identify, formalize, and document the specifics of our field. This has been a great opportunity to cover a range of relevant and contemporary topics, and as a systematic effort towards establishing better fundaments for the field and towards determining novel future challenges. In the upcoming sections of this report, we summarize the content of invited talks and of the eight main topics that were discussed within the working groups during the seminar.

Executive Summary

The field of BioMedical Visualization is a living system, which receives input from its environment–in the form of new challenges. For example, it has to deal with new imaging modalities, or it has to address new applications, and it has to incorporate new technology. When receiving all this energy-rich input, as shown in Figure 1, the system becomes increasingly chaotic, oscillating between different states. At one critical point, the system will either break apart (disintegration) or it will leap to a higher, more complex order (sustainability)—now, able to handle even more challenges.

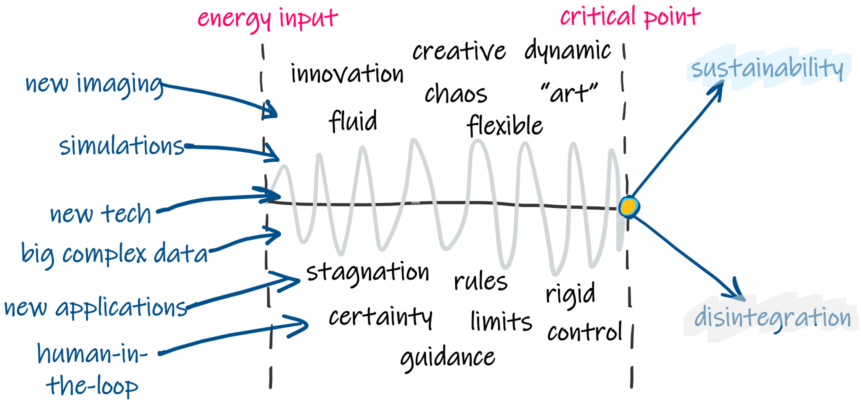


Figure 1: The field of BioMedical Visualization is a living system, which constantly addresses new challenges.

At this critical point, we have two dimensions, as shown in Figure 2. The first dimension relates to the incorporation of fresh ideas, innovation, new challenges and so forth. The second dimension relates to the establishement of rules, of a clear vision of the future and of a form of control. A balance between these two dimensions must be kept at all times. For example, if we have too much dynamics but no clear vision, the research field faces the risk of incoherence. On the other side, if we have too many rules and no innovation, the research field becomes rigid and inflexible. Both states lead to disintegration, i.e., breaking apart, of the living research field. Keeping the right balance for sustainability is what can help our field to move up and to thrive.

The main aim of the seminar was to formalize the current status of the field of BioMedical Visualization and to determine future challenges for novice researchers. While tasks and requirements in BioMedical Visualization differ from the general field of Visualization, these differences have not yet been systematized independently. A free and open discussion between all participants has been sought—opening novel, interesting paths for the future and contributing to new, creative collaborations. This effort might serve in the future as a model for other research fields within the visualization community, where interdisciplinary collaboration is also required.

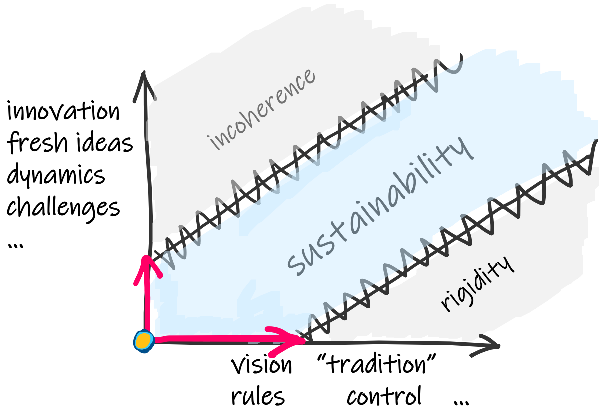


Figure 2: Keeping the right balance between novelty and vision will help the field of BioMedical Visualization to thrive in a sustainable state.

Before the beginning of the seminar, the organizers identified seven topics, which they believed were relevant for all involved participants. These were defined based on the three main pillars around which BioMedical Visualization revolves: the *Data*, the involved *Stakeholders*, and the conducted *Tasks and Processes*. On the first day of the seminar, a semi-formal discussion lead to the integration of two more topics, while the pre-determined topics were reworked (i.e., some were merged and others were split), to form a total of eight topics to be discussed. These eight topics are schematically depicted in Figure 3.

Differences between BioMedical Visualization and the parent field of Visualization
Where formalization is needed for BioMedical Visualization?
Best Practices in BioMedical Visualization
Integration & Standardization in BioMedical Visualization
Complexity vs. Knowledge in BioMedical Visualization
Reusability, Generalizability, Adoptability in BioMedical Visualization
Curriculum: How can we structure a BioMedical Visualization course?
Dissemination: Driver of Formalization in BioMedical Visualization

Figure 3: The eight topics discussed in the working groups of Seminar n.167 on "Formalizing Biological and Medical Visualization".

The upcoming sections of this report include abstracts from the invited and panel talks during the seminar, as well as a summary of the discussions with regard to the aforementioned eight topics.

Overview of Talks

I. Sustainable Applied Visualization Research

Ingrid Hotz, Linköping University, Sweden

In applied research projects, challenges and expectations are often very different from those for basic research projects with a clear technical contribution. In many cases, the anticipated outcome and the measures of impact or success are unclear. In this talk, I did not give answers to these questions but talk about experiences from three applied research projects with different outcomes. The first example is the development of a "Visual Environment for Hypothesis Formation and Reasoning in Studies with fMRI and Multivariate Clinical Data". The second example is based on a collaboration with theoretical chemists performing molecular dynamics (MD) simulations. The third example is a collaboration with mechanical engineers dealing with multivariate data from finite element simulations for the evaluation of the performance of mechanical parts. The outcome in these projects ranges from software prototypes to a talk given in an industrial forum, which resulted in the hiring of a visualization expert. While all these projects are success stories to some extent, they all also leave the question about sustainability and lasting impact open.

II. BioMedVis—Opportunities & Threats

Helwig Hauser, University of Bergen, Norway

Out of a brief assessment of the state of the art in both biological data visualization and medical data visualization, and based on a short review of current strengths as well as weaknesses, a number of opportunities are identified and discussed, including (1) new data and new forms of data, (2) multi-modal data visualization, (3) the visualization of physiology, (4) the study of entire cohorts, (5) comparative visualization, especially when more than two datasets are to be compared, (6) the study of multi-scale phenomena, (7) the study of multi-dimensional problems, (8) data science in computational medicine, biomedicine, and biology, (9) the integration of modeling and simulation, as well as (10) the more conscious focus on biomedicine, precision medicine, etc. This discussion of opportunities comes with a brief recapitalization of associated threats, such as the risk of redoing too much too often, getting lost in details, the non-availability of data, especially in the multi-scale case, the risk of working solution-oriented (instead of problem-oriented), the possibly too extended distance to the domain and an according lack of interdisciplinary collaboration, the threat of becoming partially obsolete due to machine learning, and—as "always"—a lack of funding combined with difficulties associated with the evaluation of interdisciplinary research proposals.

III. Formalizing Biomedical Visualization: Connecting Macromolecular Structures to the Big Picture

Marc Baaden, CNRS Paris,France

Macromolecular structure visualization is a historic field in BioMedViz and going back through a few of its milestones and characteristics may provide useful clues about formalizing and advancing the larger domain of BioMedViz, as well. A few particular lessons may be learned about the big divide between the computer science and domain communities, about the crucial role of software and limitations of current practices, as well as about how to get stakeholders on board of the latest evolutions in the field. A few community-wide used codes have been provided for more than 20 years now by a few research groups, many more specific tools exist with varying degrees of availability. Most conceptual prototypes however seem to become quickly inaccessible after initial publication. This situation has consequences on how innovation may—or may not—transfer from the core visualization research community to the domains of application.

IV. Essence of Data Visualization: Putting the Form in Formalizing

Martin Krzywinski, Canada's Michael Smith Genome Sciences Center at BC Cancer, Vancouver, Canada

Well-designed figures can illustrate complex concepts and patterns that may be difficult to express concisely in words. Figures that are clear, concise and attractive are effective–they form a strong connection with the reader and communicate with immediacy. These qualities can be achieved by employing principles of graphic design, which are based on our understanding of how we perceive, interpret and organize visual information. Additionally, because figures often act as a first explanations, it is critical to distinguish essentials from details and merely interesting tangents–not an easy task for the researcher who feels that all their hard-won data should be shown. While everything may indeed be important, initially some things are more important than others. Classifying aspects of the science this way always feels risky—how do I know that I know enough to justify leaving things out? Using critique by redesign, this talk distills core concepts of information design into practical guidelines for creating scientific figures.

V. Panel Discussion: Formalizing Biological and Medical Visualization

Helwig Hauser (moderator), Stefan Bruckner, Michael Krone

In this panel, we want to stimulate the discussion on the rationale for, feasibility of, as well as benefits and drawbacks of formalizing the area of biological and medical visualization by presenting to opposing viewpoints presented by the two panelists. The viewpoints correspond to extreme positions not necessarily held by the panelists, but aim to outline the continuum of opinions on the subject.

Stefan Bruckner: I advocate the position that a too strong focus on the subfield of biological and medical visualization is counterproductive. While this area certainly addresses an important set of applications for visualization technology, it does not exhibit fundamental conceptual or methodological distinctions from other application domains of visualization. By attempting to single out medical and biological applications, we contribute to a balkanization of the field, leading to reduced exchange and hence potentially limit the impact of our work. Visualization in the life sciences is “just another application”, albeit a highly interesting and important one!

Michael Krone: I took the position that biomedical visualization (or “life science visualization”) can be seen as a field of its own and not just another application area of visualization since there are a couple of distinctive points such as the privacy issues, restrictions and ethical implications for example related to health-related data. The field also has separate conferences and meetings that focus only on biomedical visualization. Examples are EG VCBM or the BioVis symposium, which used to be collocated with IEEE VIS but is now a COSI at ISBM. Biomedical visualization can be compared to bioinformatics or medical informatics, which are part of computer science but are also independent fields of research that require a distinctive set of skills taught in specialized bachelor and master programs. A similar approach could be possible for biomedical visualization, which could be a masters program for specializing after getting a bachelor’s degree in CS or BioInf, since it is not possible to teach the skills required for biomedical visualization in one course. A proper formalization of this field could be a primer for establishing biomedical visualization as a distinct discipline.

Topic Summaries of Working Groups

We will now summarize the findings of the respective group discussions. Each of the challenges lists all participants of the respective session(s) in alphabetical order. Each of the participants took part in two groups. Each group had a moderator, whose name is denoted with bold font.

Topic 1—Difference between BioMedical Visualization and the parent field of Visualization

Johanna Beyer, Helwig Hauser, Tobias Isenberg, Renata Raidou, Timo Ropinski, **Thomas Schultz**

Motivation During this seminar, the question "*Are we just another application field within VIS?*" has been raised very often. To formalize the domain of Biological and Medical Visualization (BioMedVis), we first need to understand the characteristics of this particular domain, the aspects that unite us as a community and the points that differentiate us from the other application fields of the general Visualization (VIS) discipline. Our working group identified as the starting point of the discussion that the currently used term of "Biological and Medical Visualization" is too narrow. This term implies that the focus of our work is to provide visualization solutions for the medical or biological domain, while the actual research focus of our community is far broader than this—both with regard to the employed technical methodologies, and the addressed application domain. A common agreement is that working within the BioMedVis domain requires to obtain early and profound knowledge of the respective life science or medical domain, on top of the technical expertise in, e.g., visualization, image processing, image analysis, machine learning, and so forth. To determine a more accurate and broader term that describes our research domain, we also need to consider the disciplines, the backgrounds, and the personal research interests of the researchers that are part of our community. In the upcoming sections of this report, we initially present the identified challenges that differentiate BioMedVis from VIS, as well as the aspects that create a common ground for all researchers working interdisciplinary within the BioMedVis domain. Also, we present a number of outcomes and actionable points with regard to the identified challenges.

Challenges Clarifying the difference between BioMedVis, VIS, and adjacent subfields can help us to better address the following practical challenges:

- *Education*: Does it make sense to offer specific classes on Biomedical Visualization? How should they relate to a general class on Visualization? What are the prerequisites for learning about the details of our field? Are there specific teaching goals, particularly for BioMedVis?
- *Community*: Which expertise is required to address the major research problems within Biomedical Visualization? What is the best way to involve domain experts? From which technical fields other than Visualization do we need to involve experts? Where do we associate ourselves within the research community? Given that we have many different backgrounds, what commonalities do we have with each other?

- *Publication culture*: What are the most suitable venues for disseminating work on Biomedical Visualization? Do we need new conferences or journals, or new types of publication? What about publishing software and data?
- *Funding*: What are suitable sources for funding interdisciplinary research within Biomedical Visualization?
- *Research questions*: How do we view the roles of fundamental vs. domain-specific research questions within our field? How has the SciVis/InfoVis/VA separation influenced our community?

Outcomes The outcome of the initial part of our discussion regarding the identification of characteristics and the involved disciplines of the BioMedVis domain can be summarized with the Venn diagram in Figure 12:

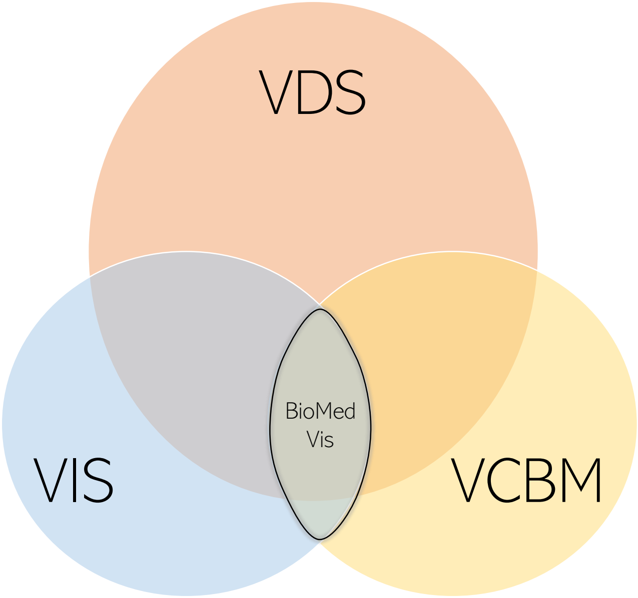


Figure 4: Venn diagram representing the position of the domain of BioMedVis in respect to the parent VIS field, the domain of Visual Computing in Biology and Medicine (VCBM) and the newcomer field of Visual Data Science (VDS). This diagram illustrates the interdisciplinarity of our research field.

What distinguishes BioMedVis from the general field of visualization is that it requires knowledge, and collaboration with partners from the life sciences or medicine. In this respect, it is similar to other application domains in visualization, and it is a regular part of visualization without any fundamental differences. However, we also view BioMedVis as part of a broader and highly interdisciplinary field, Visual Computing in Biology and Medicine (VCBM). This term does not fully overlap with visualization, but rather includes numerous sub-fields of computer science, such as image processing or computer vision, as well as statistical and machine learning techniques, which are applied to visual data from our application domain, or integrated with visualization. We identified "life sciences" as a term that is more comprehensive, and in our opinion more suitable to describe the scope of our field, than "biology". However, we believe that "life sciences" should not replace "biology and medicine", since it might be seen as excluding techniques that support the practice of medicine, as opposed to medical science / research. However, we believe that our domain should promote, e.g., in formulations of calls for papers, the term "life sciences" to facilitate the inclusion of new application domains and new practices within our research. We identified "Visual Data Science" as an emerging term that is increasingly being used to describe an approach in which data management, machine learning, and statistics, are tightly integrated with visual data analysis. We expect that its role within the domain of life sciences and medicine will increase, and believe that the VCBM community should be open to this development.

Concerning the specific challenges listed above, we came to the following conclusions:

- *Education:* We see two main potential target groups for a BioMedVis education: Students from computer science, and students from the domain sciences. In the former case, the contents should be more technically oriented, with a focus on understanding methodology and being able to develop and implement it oneself. Depending on the local circumstances, the course might be specific to life sciences and medicine, or it might be a more general class about applied visualization. A course that targets students from the application domain should focus on the correct and effective use of available techniques.

- *Community:* We see clear advantages of collocating our events, such as VCBM, with other meetings, which could be either from an application, or an adjacent technical domain. Due to the diversity of our field, we see no natural candidate for a regular collocation, but we would welcome interchanging collocations, depending on which opportunities offer themselves. In addition to that, we should promote the reverse "flow" by supporting with our participation venues of local domains of expertise, e.g., local medical or bioinformatics conferences. This could be an excellent source for new ideas and new collaborations with domain experts.

- *Format of VCBM:* We see VCBM as an opportunity to experiment with non-standard formats, such as hackathons, breakout sessions, hands-on workshops, or bring-your-own-data sessions. This would allow us to showcase to domain experts the endless possibilities that our domain offers, and it would help us gain more knowledge about the domain of application. Additionally, we believe it would increase the visibility of VCBM on the web if `vcbm.org`, where a web portal (which could prominently link to the current VCBM workshop) instead of a forward to the current year's event. Such portal could include additional relevant information for new researchers, and ways to create and promote links within the community.

- *Involvement of domain experts:* For good reasons, technical and application communities will keep their separate venues. However, we would like to see an increased exchange between these communities. As one instrument, we encourage VCBM organizers to invite presentations from domain experts, preferably from their local communities, as well as to participate more often–when/if possible–to domain venues.

- *Publication culture:* At this point, we see strong benefits from keeping VCBM as an opportunity to present also more experimental or more narrowly focused work. An interesting strategy would be to help domain experts publish in their venues a showcase of our visualization work.
- *Research questions:* In our workshop calls and assessment of submissions, we should be inclusive with respect to currently underrepresented "life science" topics, such as biomedicine (which might provide a stronger link between the currently somewhat separate directions of biological and medical applications), visualization of abstract data, machine learning.

Topic 2—Where formalization is needed for BioMedical Visualization?

Marc Baaden, Stefan Bruckner, Issei Fujishiro, **Helwig Hauser**, Barbora Kozlíková, Martin Krzywinski, Johannes Sorger

Motivation The initial aim of this topic was to detect to which degree the formalization of the BioMedical visualization (BioMedVis) is necessary, in which way and where it would be clearly beneficial, and who could benefit from that. Based on the discussion, we were trying to give examples of formalization from other fields and suggest possible future action points in this direction.

Challenges First of all, all participants of the discussion agreed that it is difficult to define the BioMedVis field from the methodological point of view, and the first part of the discussion was devoted to the subject of finding the characterizing lines of such a combined field. Both fields are facing complex multiscale phenomena and one of the main challenges in both of them is a simultaneous handling of the interplay between different types of data, such as spatial, multidimensional, nominal, and temporal. This can quickly lead to the problem of maintaining the visual elements – we can run out of colors, shapes, and other visual variables. Therefore, it can be overwhelming to solve these problems at the same time. Also, especially in medical visualization, the problems being solved were originally very spatially-oriented. Nowadays, both biological and medical visualization are facing big heterogeneity in data, they need to incorporate visual analysis into their solutions, etc.

Further we were also discussing the positioning of the BioMedVis in the context of a broader perspective, such as life sciences, or nutrition science. We pointed out that the definition of the "first principles" of BioMedVis is missing, stating how to visualize the target dataset. Here we could be inspired by the general principles proposed by Tamara Munzner [9]. In our fields, the nature of the underlying phenomena, measurements, and multifaceted-nature of the data plays a crucial role, which needs to be taken into account.

In order to better define the BioMedVis field, we conducted a thought experiment where we were trying to define the scope of a hypothetical BioMedVis journal and detect borderline cases for the editor to decide for a desk-reject. One of them could be that the paper is focusing too much on the inorganic chemistry or the lack of sufficient technical contribution from the visualization perspective. Here we discussed if incorporating only the image processing or rendering and computer graphics techniques are acceptable contributions and we came to the conclusion that such papers should still be considered as a potentially beneficial contribution to the BioMedVis field. Generally we agreed that an acceptable paper should cover biomolecular visualization for medicine, with a strong application for a disease. The scope can be broadened from human medicine to animals and plants as well. But the scope of such hypothetical journal should be also derived from the needs of the community and their tasks. Therefore, in the following discussion, we were focusing on the better definition of the target audience and who could potentially benefit from the progress in formalizing the BioMedVis field.

Here we detected the following potential beneficiaries of the formalization:

- **Students and collaborators (domain scientists)**– formalizing the field and introducing the consistent terminology can reduce the ambiguity in terms of use of the visual elements that are mapped onto specific meanings. It can decrease the entry barrier for entering the field and can help with reducing the situations when the visualization is wrongly perceived. The collaborators would benefit mainly from finding the common language between the domains.

- **Visualization designers** – similar benefits as students, with possible difference in the source of gaining the information (mostly webpages). Here it is important to state the domain conventions which should be followed by the designers.

- **Visualization researchers** – except for the same benefits as for the designers, the researchers could additionally benefit from the software standards, standards for interactions, interoperability, etc.

- **General public** – can enable to reach better literacy, can lower the burden to understand the conveyed information. In other fields, such as geovisualization, it was already proved that standardization leads to faster learning and understanding.

Based on that, we were trying to discuss how this meeting could influence the future of our community – shall we separate or rather more integrate with the general visualization community? This topic was also extensively discussed in the discussion panel of the meeting. Except for that, we were raising the issue of the lack of presence of the domain experts at the visualization venues, such as BioVis or VCBM workshops. The reason could be that the need from the domain scientists is not so strong yet. The individual collaborations are what matters the most. Currently there is still a big difference between knowing by the experts what is possible and what is needed. If there is a huge bottleneck in the field where visualization could be helpful, the experts will come and search for help.

Examples of formalization Another important topic of our discussion was to compile a list of examples which could inspire us in the formalization process. These are:

- An algebraic process for visualization design by Kindlmann and Scheidegger [7].

- Goal-oriented taxonomy of visualization, such as the Wehrend's matrix (A problem-oriented classification of visualization techniques by Wehrend and Lewis [17]). Similarly, we can be inspired here by the Tamara Munzner's (action,target) pairs [9].

- Many published attempts to formalization are either too general or too specific, but can still serve as an inspiration. Here we mentioned again the Munzner's book [9], the book on Visual Computing for Medicine: Theory, Algorithms, and Applications by Preim and Botha [11], or the paper entitled Visualization and Visual Analysis of Multifaceted Scientific Data: A Survey, by Kehrer and Hauser [6].

- Improving figures and visualizations to make them more readable is one of the main challenges. Here we can be inspired, for example, by a paper by Vuong et al. [16]. Here belongs also the discussion about introducing the bias by presenting the data in different ways. We mentioned a well-known example, see Figure 5.

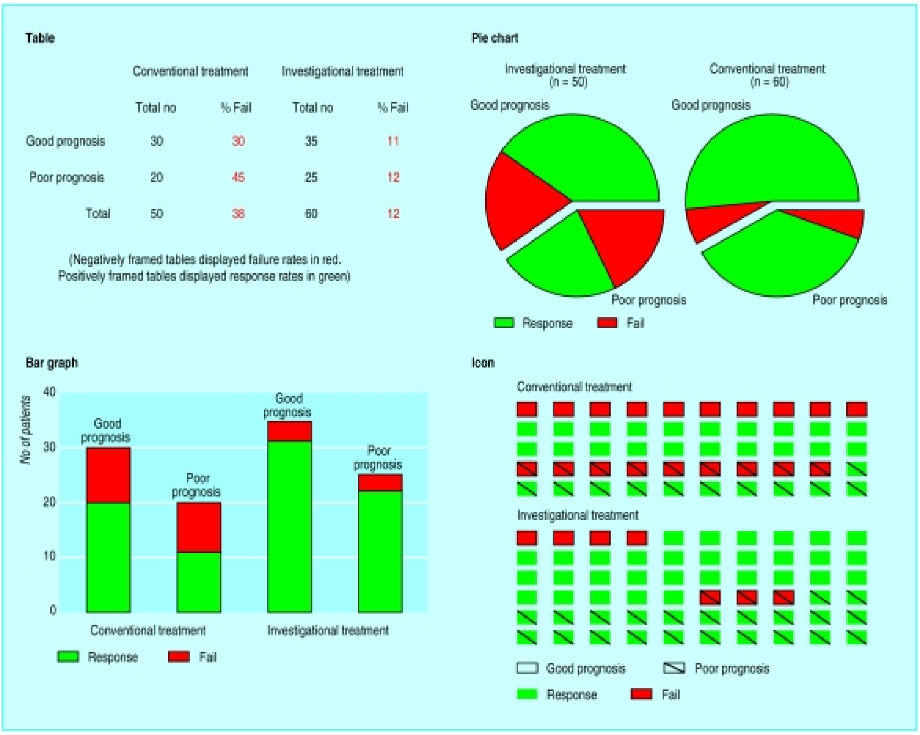
- Another crucial aspect is to handle the text in visualizations in a better way. The standards for labeling should be set, they should take into account, for example, the hierarchy in the data. Symmetry in data should lead also to symmetry in labeling (positioning the labels, their alignment, ...). The rules from typography should be applied to visualization.

It is also worth to mention that formalizing BioMedVis field is not just about defining static visualization, but also the dynamic processes which are often the crucial part of the visualized phenomena, which need to be interactively explored as well.

In general, we need to work towards better communication of science. If the visualization product is useful and is meeting the needs of the collaborators, it can be considered as successful. Therefore, if we will be able to agree on some rules, this is already very valuable. It will enable to define the interface between the two communities sharing the visualization – the visualization researchers and the domain experts, who are in need of more comprehensible and efficient ways of exploring their data.

Outcomes In this part of the discussion, we aimed to identify potential tangible outcomes which could be followed in the future. Here we suggested the following action points:

- **Invariants of good visualization** – when such invariants are defined, they could be published as a Points of View column in Nature Methods. We could come with 3 or 4 pieces and in each one of them, we would choose one invariant to be presented. Here, each of these invariants should be worked out through a selection of examples. We already detected possibly interesting invariants – to identify a clear axis of the visualization, to be conscious about the hierarchy in the visualization, to discuss the color schemes, typology, or other aspects.



Table

	Conventional treatment		Investigational treatment	
	Total no	% Fail	Total no	% Fail
Good prognosis	30	30	35	11
Poor prognosis	20	45	25	12
Total	50	38	60	12

(Negatively framed tables displayed failure rates in red. Positively framed tables displayed response rates in green)

Figure 5: The effect of the method of data display on physician investigators' decisions, taken from [4].

- **Getting visualization to the next level** – here we want to propose an Opinion piece article addressing, for example, the "graveyard" of software prototypes and how to improve the current situation in that, the lack of support for searching across images (indexing the image content by keywords), or the standardization agreement based on rules (interoperability, e.g., arcballs). Regarding the indexing system for searching across images, we can envision a similar solution to the text-based search. Each figure could be described by 5 keywords, serving for subsequent search.
- **Inspiration from the existing guidelines** – we can be inspired by the already established and used guidelines in other fields, such as the IUPAC guidelines in chemistry (`https://iupac.org/`).
- **Where to further discuss these topics** – to continue with the discussions and potentially also include more people, we could propose a workshop at the VIS conference which could help to initiate the follow-up discussions and other outcomes regarding the formalization idea.

In conclusion, we agreed that even if we are probably not able to globally formalize the BioMedVis field, the formalization can be still performed on the local level of individual collaborations, which can be of great benefit.

Topic 3—Best practices in BioMedical Visualization

Michael Krone, Tobias Isenberg, Renata Raidou, Noeska Smit

Motivation This working group was established in order to discuss if there are any best practices and recommendations that are specific to biological and/or medical visualization (BioMedVis). Such recommendations could serve as a starting point for graduate students that are new to the field. Furthermore, a structured collection of best practices from these fields could serve as a basis for the formalization based on previous outstanding examples. Such a literature survey could also help to identify white spots in the landscape of BioMedVis research where more work is required.

Challenges In order to identify what best practices are for BioMedVis, we first needed to identify whether there is anything specific about BioMedVis research versus visualization research applied to other fields in general. An issue that immediately comes to mind when working with patient-specific data. The rules for protection of such data are more strict, also under the GDPR regulations since they fall under the personal data provisions [1]. Anonymizing such data is challenging. Even if a medical imaging scan only has a number assigned to it, as soon as there exists a coupling list between these numbers and patient information, the data is not anonymous, but rather pseudonymous. In addition, data like CT scans of the head can feature identifying information even within the imaging data itself. One approach to work with sensitive data is to keep the data where it is best protected, e.g., within a secure hospital environment. The visualization software can then be developed with 'safe' data externally and deployed within the hospital environment for research purposes. For most ethical considerations, the main responsibility lies with the domain experts acquiring the data. Project partners may need to re-apply for ethical clearance if they want to re-use the data in other projects. Sometimes, domain partners are not eager to publish findings in visualization papers before they have had the chance to publish. Alternatives could be to show images of similar data or to present the results vaguely. Even when there is no direct patient data involved, any research involving people, as for example in user studies, should be carefully treated. In biology, there are also protection mechanisms in place related to preservation of rare plant species. Developing a repository of datasets to verify our techniques (benchmark) could be fruitful, for example something like the list on medvis.org[2], but at a higher level. A starting point could be data from previous VIS challenges.

Another challenge arises in BioMedVis when attempts are made to translate research software into clinical practice. Strong certification regulations are in place for any software that influences patient care decisions that need to be obtained. These patient care decisions also lead to a situation where clinicians need a high level of trust in visualization solutions before basing any decisions on them. A question arises also on who is responsible for the outcome of any such decisions when visualization software was involved in making them.

A challenge that holds for both biology and medicine is that these fields are dealing with processes that are not easily controllable. These processes are

---

[1] `https://gdpr.eu/eu-gdpr-personal-data/`

[2] `https://medvis.org/datasets/`

multi-scale both in space and time, leading to interesting visualization challenges.

Part of the discussion in this breakout group focused on previous papers that are good examples from which a formalization could be inferred. We first started with theoretical considerations of how a good BioMedVis paper should be structured. As in all visualization papers that are targeting a specific application domain, sufficient domain background knowledge needs to be included so that a visualization researcher can follow the paper. However, this section should not take up too much space in the paper and should not mention unnecessary details that might be interesting but not needed to understand the paper. Here, it is important to mention that on best practice when writing about BioMedVis should be that the text should be tailored to a specific audience. Since BioMedVis often addresses health-related problems, it is often also interesting to a lay audience. That is, not only the resulting visualization targets a general audience, but this general audience might also be interested in the visualization itself, not just the visualized data. Similarly, a paper could be targeted at domain experts wanting to work with BioMedVis for exploratory analysis and, consequently, wanting to understand it to be able to judge its reliability.

When trying to come up with examples of successful BioMedVis research, we divided them into examples that had a tangible impact since they are now used in practice and examples that we think were highly influential on subsequent research in the field. The following examples fit into the first category:

- Many techniques for visualizing the results of medical imaging, like 3D ultrasound rendering, Virtual Colonoscopy, and Curved Planar Reformations [5] are nowadays used in clinical practice for diagnostic tasks. High-quality volume rendering is also used in touch-screen tables by SECTRA, which are used in diverse settings, such as medical visualization, virtual autopsy, and outreach (museums).

- For the visualization of three-dimensional molecular structures, tools like VMD and UCSF Chimera have been developed and are now routinely used, for example, by computational chemists and structural biologists for their research. These tools include molecular representations, like Molecular Surfaces.

- Biological network visualization tools like Cytoscape are used for research in academia as well as industry.

- BioMedVis is also used in practice for educational purposes. One example for this is the biomolecular visualization tool marion, created by Ivan Viola's group at TU Wien, which is now used by the illustrator Drew Berry to create educational movies about whole cells. Similarly, medical visualization research at TU Delft has led to the creation of a Massive Open Online Course (MOOC) for anatomy education[3].

The following papers are examples for highly influential works with respect to subsequent research in the field of visualization:

- Researchers from University of Münster, Germany and Linköping, Sweden have investigated the influence of different illumination models on image

---

[3] `anatomy.tudelft.nl`

comprehension in direct volume rendering [8]. Their findings have direct implications for (especially medical) volume rendering.

- For the visualization of molecular structures, Tarini et al. wrote a paper showcasing the use of ambient occlusion and edge cueing to enhance the perception of these complex, three-dimensional structures in real time [12]. This paper is still inspiring and influencing research in this area.
- Visualization research is usually published in specialized journals that have a rather low impact factor compared to other disciplines. One early success story is the seminal work of Connolly on the visualization of smooth Molecular Surfaces, which was published in Science. The fundamentals presented in this work are used to date for molecular surface rendering. Another notable paper that was published in the high-impact journal Nature presents the visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding [15].
- BioMedVis tools like VolumeShop [2] or MegaMol are used as prototyping platforms for further research by PhD students (at least in the respective groups), leading to shorter development cycles and, consequently, an increased publication output.

Please note that the examples mentioned above are neither meant to be an exhaustive list nor necessarily the highest-ranking examples. The list just represents examples from both fields what were mentioned during the breakout discussion by the four participants. A good starting point to identify further work that could be used to infer best practices for the field are the recipients of the biannual Dirk Bartz prize for medical research, as this prize also considers implications for impact in a clinical setting.

Outcomes The working group discussed recommendations in order to increase the chances for our BioMedVis to have practical impact:

- Put an extra effort to showcase results, methods, and tools beyond publishing the paper. This increases the visibility of the work, increasing the chances of practical uptake. Concrete recommendations here are:
  - Social media outreach, e.g., Twitter, Facebook, Blogs, etc.
  - Making source code available, e.g., Github, permissive licensing, etc.
  - Making author copy of paper available, e.g., arXiv, HAL, etc.
  - Making video demos and tutorials available, e.g., Youtube, Vimeo, etc.
  - Making a project page/website with, e.g., demo (executable), example images, videos, data (if possible)
  - Presenting in application domain venue, e.g., a poster, tutorial, application/tool paper
  - Presenting at outreach event or exhibition
  - Publishing a meta/summary/overview paper in venue for larger audience, e.g., CS, popular science, etc.

  - Creating step-by-step tutorials for the application domain

- Put in extra effort to obtain deeper knowledge about the application domain and meaningful relationships. This increases the chances to develop useful solutions with features that are must-haves, rather than nice-to-haves. Concrete recommendations are:
  - Reading textbooks for the domain
  - Writing a State-of-the-art on the topic
  - Working within the domain expert's environment to observe, e.g., during a sabbatical, a day or week, attending their conferences, etc.
  - Having an application domain 'buddy', perhaps of the same seniority to increase availability throughout the project
  - Hiring people with the right domain background
  - Organizing semi-structured hands-on workshops
  - Establish a good working relationship with trust from both sides with your partners. Follow-up on ideas and give them a win-win situation (contributions in both visualization and domain research)
  - Develop software that has a higher chance of practical uptake
  - Developing stable, usable, sustainable software
  - Considering the best target for the software: integrating code into public projects, developing libraries, standalone or web tools, permissive licensing
  - Linking up with existing frameworks that are in use in the domain and developing a plug-in for these frameworks
  - For medical visualization, adhering to the ISO standards, ethical restrictions, and privacy requirements
  - Teaming up with a company, for example via an industry-funded PhD, to develop research results into a product

In addition, we have developed best practices for working with biomedical data with respect to data privacy and ethical issues:

- Follow the GDPR and recommendations from the ethical committees established at hospitals is crucial.
- Anonymize the data where possible. There is related work on data anonymization methods available.
- Use example datasets, if possible, with less ethical/privacy implications:
  - make your own datasets (scans of authors),
  - generate or simulate data (accuracy concerns),
  - explore the use of 'toy datasets' for testing, e.g., mouse data.
- Anyone handling the data should sign NDA agreements.

Topic 4—Integration & Standardization in BioMedical Visualization

Lars Linsen, David Mayerich, Shuichi Onami, Guido Reina, Baoqing Wang, Hsiang-Yun Wu

This working group discussed the two issues of (1) standardization of data and methods and (2) robust methods for integrating multi-faceted data representing individual samples.

Motivation Modern data in the fields of medicine and life sciences contain data gathered from multiple sources for individual samples. **Integrating** this information into a joint analysis is required for a comprehensive understanding. Due to variations in data acquisition, the data from different sources can be in inconsistent formats. **Standardization** is required to have comparable data. Comparable data allows for the application of standardized workflows and for comparative and ensemble/cohort analysis.

Definitions We first provide a set of definitions for terms we have found critical to this discussion:

- **Integration** – integrating data from different sources and of different types.
- **Standardization** – the process of making something conform to a standard.
- **Standard** – something used as a measure, norm, or model in comparative evaluations.
- **Multi-faceted data** – data sets with samples composed of data from multiple sources and of multiple types (i.e., patient data composed of MRI, histology slides, and doctor notes)
- **Interchange formats** – We will have to decide on something even though nobody will like it. Least a common denominator. We should include an umbrella meta-file describing how separate data are integrated for a single sample/subject.

Challenges The biomedical community is currently facing a data explosion, with multi-modal acquisition methods collecting massive amounts of data with limited tools for analysis. Visualization and analytical tools targeting biomedical data must therefore overcome barriers imposed by this data explosion. In this section, we address two tracks for coping with the expanding biomedical data sets: (1) interchange formats that can cope with the existing and emerging data types, and (2) a curated benchmark repository facilitating the standardization and objective comparisons where possible.

Data Types One of the major challenges facing formalization is the broad array of data types and competing standards. However, we believe that enforcing an additional standard on an entrenched and constantly evolving biomedical community is impractical. We therefore advocate for the establishment of a set of *interchange formats* that maximize data inclusiveness and facilitate conversion between the existing and future competing standards.

In this section, we will discuss the current variety of basic data types, along with the recommended requirements for a formalized interchange format.

- **Tabular Data** Tabular data refers to data where rows and columns of a table represent the samples (or instances) and attributes (or dimensions), respectively. The attributes can be of different types, such as numerical (with continuous or discrete range), ordinal, or categorical (or nominal). They may also contain spatial coordinate values. Tabular data are commonly stored in spreadsheet-like formats. For dynamic data, each time step is stored in an own spreadsheet with some implicit sorting. There is no common standard on how to do that. A standardization is required.

- **Images (multidimensional arrays)** One of the most common biomedical data types encountered is image data. Current standards for image storage, such as bitmaps and lossy-compressed JPEG, are fundamentally limited to two-dimensional images. This reduces the applicability to a wide range of common medical images, including MRI, CT, and PET imaging.

  Three-dimensional representations, including DICOM and TIFF files, are currently limited to three-dimensional data and have limited support for hyperspectral images, which we have seen a growing use in the form of FTIR[4] and Raman[5] methods. One common format available in hyperspectral imaging, primarily used in geospatial imaging, is the ENVI file[6]. The ENVI file format is essentially a RAW memory dump with a corresponding human-readable header file providing details about data format and annotation.

- **Text** At different steps of acquiring data, there is textual data added, such as reports in clinical data or annotations. Textual data commonly consist of an unstructured collection of characters. Structured textual information (such as markup language-based description) is less common. Standard formats across different fields of medicine and life sciences are not established.

- **Explicit Geometry** Explicit geometry occurs in various fields, e.g., when extracting the geometry from imaging data or when describing molecular structures. Example formats include glTF, OBJ, COLLADA, or PLY-files.

- **Relationship Data** Network is a common model that describes the objects and their corresponding relationships. Trees, graphs, and matrices are data structures often used for this purpose. Typical examples include JSON, xml, property graph, RDF, and others.

---

[4] https://www.sciencedirect.com/topics/earth-and-planetary-sciences/ftir-spectroscopy

[5] https://www.sciencedirect.com/topics/engineering/raman-method

[6] http://www.harrisgeospatial.com/portals/0/pdfs/envi/Reference_Guide.pdf

- **Metadata**

  Annotations Every interchange format developed should have a corresponding annotation format. This annotation format should provide (1) a semantic segmentation of the source file, (2) an identifier, and (3) links to the corresponding annotations in other files (other identifier-file pairs).

  Provenance Along with the operations that are executed on the data, we need to store all of them describing the author, date and time, and operation performed. This provenance information shall allow for reproducible results.

- **Multi-Faceted Data** Nowadays, all of the facets described above come together, e.g., when considering cohorts of people with all kinds of measured data including imaging data, health record data, and textual descriptions. This needs to be handled by a combination of multiple standard data formats. An umbrella data format is required to explain how everything comes together.

Benchmarks One major step towards the formalization may be the establishment of a benchmark repository.
Contents of a benchmark repository: This repository should contain a wide array of data sets produced by the domain experts and include a range of features:

- **Scalability of extent** – Representative data sets should exhibit a range of sizes, in both extent and resolution. Small data sets are useful for quickly evaluating the visualization and data mining techniques, while large data sets are critical for producing scalable algorithms.

- **Scalability of dimensions** – Data leveraging multiple dimensions, including 3D spatial, hyperspectral, and time-domain images should be available. These represent the critical aspects of current and emerging imaging modalities.

- **Scalability of density and complexity** – Data sets often contain important data embedded in lower-dimensional subspaces. This is commonly seen in hyperspectral data, where the spectra are routinely reduced using dimension reduction (e.g., PCA, ICA). Alternatively, sparse data requires different representations. A benchmark repository should draw from sparse, dense, and reducible data.

- **Anomalies** – Abnormalities, such as outliers, systemic artifacts, and noise, are consistently present in biomedical data. Visualization and analytical methods must be robust, and therefore a benchmark repository should house a realistic range of data exhibiting these anomalies.

Creation of a benchmark repository Formalization requires that benchmarks are well curated and well described. We propose a peer-review process for hosting within a curated repository. Examples of peer-reviewed repositories are available for protocols (Journal of Visualized Experiments) and scientific data associated with publications (Nature Scientific Data). The peer review process should require the submission of the data and associated annotations, along with a publication focused on the needs of domain experts. This publication should describe the methods behind data collection, challenges in data collection (e.g., image anomalies), and the desired outcomes for any informatics or visual analysis.

Benchmarks for Calibration One expected barrier in building a benchmark repository is the wide range of systems involved in data acquisition. We also recommend establishing a set of normalization standards with the specific goal of mapping data acquired using specific instruments (ex. Nikon, Siemens, etc.) to a normalized standard. This software should ideally recognize and adjust for protocol variations, and may integrate standardized phantoms imaged across a variety of systems.

Certification Standardized performance on benchmarks can lead to certification. To define the certification process we need input from stakeholders to define the requirements.

Standardized workflows for visualization We have no standard system that everybody accepts. The main reason for that could be that there are no proper repositories where modules of the system can be added (by people not being in-house developers). Having standardized workflows (and maybe even certificates) would clearly increase the trustworthiness of the visualization tools and its acceptance by stakeholders.

A standardized workflow could be, on an abstract level, described by:

1. Preprocessing
2. Information extraction
3. Mapping to a visualized extraction
4. Interaction mechanisms
5. Go to 1,2,3, or 4

Figure 6 provides a sketch of this workflow. The implementation of a standardized workflow for different application scenarios is necessary.

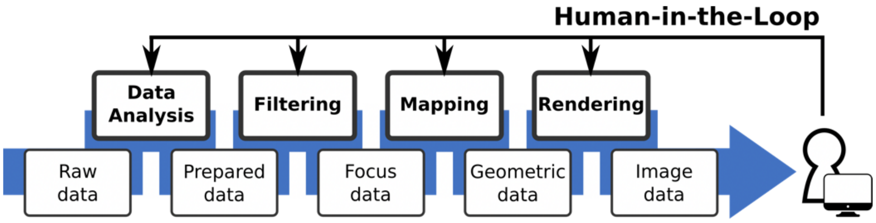


Figure 6: Visualization pipeline.

Outcomes We have discussed the need for data integration as well as standardization of data and methods. We have seen that data is multi-faceted and that there is a need to incorporate meta-information, such as annotation and provenance data. Hence, an umbrella meta-file is necessary for interchange formats.

Benchmark data sets may be a way towards standardization, but would require a supervision system to control the collection. Standardized workflows help to systematize the creation of visualization tools. However, a common system for joint software development of the community does not exist, although it would be desirable.

Topic 5—Complexity vs. Knowledge in BioMedical Visualization

MICHAEL KRONE, LARS LINSEN, DAVID MAYERICH, **Pere-Pau Vázquez**, BAOQING WANG

Motivation Biomedical data are steadily increasing in complexity, especially due to the fields of computational biology and bioinformatics. Despite the recent advances of combined automated and visual data analysis, we (and also domain experts) still have limited knowledge about the data. For this case, automated data analyses are not suitable, while semi-automated analysis is still considered tedious.

Challenges There are many challenges that we need to be addressed in order to increase the knowledge about data when the datasets become so large. For the rest of the discussion, we define some terms that are related to this problem:

- **Heterogeneity:** Multi-dimensional data composed of different categories of information (numerical, categorical, textual, etc.).
- **Complexity:** Describes how challenging/costly a visualization is to interpret.
- **Knowledge:** Awareness or familiarity gained by a visualization experience.
- **Richness:** Amount of information contained or embedded in the data.

With this in mind, we addressed several challenges:

- How do we increase knowledge?
- Does the increase in data richness lead us to the increase in knowledge?

Initially, one may think that the increase in the resolution or richness of a dataset naturally leads to an increase in knowledge about the data. However, we need some way to transmit this information to the human. This is challenged by the fact that the data may be too complex. Therefore, traditional exploratory visualization techniques may be difficult to apply due to the high number of parameters implied. Thus, applying such techniques to highly rich data, may generate a very complex visualization, that may be difficult to understand by the users. At this point, we may wonder how to measure the complexity of a visualization. And it may also be interesting to be able to set a line between a simple, easy-to-understand visualization versus a too complex one. Unfortunately, the complexity of a visualization is difficult to measure. Even using the same visualization technique, the result may be easy or difficult to interpret depending on other factors, such as the data. We discussed on it, and concluded that some of the factors that are highly correlated to the complexity of the visualization result are the following elements:

Number of independent parameters: The higher the number of independent parameters to represent, the more complex the visualization will be, since we need to devote more visual channels to it, and human perception is limited.

Dynamic range of each dimension: The number of values a certain parameter can take may further complicate the representation of the variables, because larger ranges of the perceptual domain will be required.

Number of items: The number of items in each dimension largely varies among datasets. For the general, case, we can assume it will be high, and therefore require a lot of space if individual identification is needed.

Spatial separability of items: Accumulation of items around certain values may make the visual representations cluttered, while if they can be separated, visualizations may become more understandable.

Sparsity: The distribution of the elements also plays a role, since more space may be required for the representation.

Hetereogenity: The dimensions of the data may be of different forms (e.g. text, image...) some of those are easier to represent and more space efficient.

So, for increasing the knowledge, we believe that the improvement of visualization techniques as well as the creation of new automated algorithms that are able to extract information for the data, are required. However, this is not easy, a number of problems arise: *i)* The improvement in capture devices as well as simulation algorithms leads to increase in data sizes, that may generate storage or data access problems. *ii)* The increase in data richness also requires higher computational power to process the data. This is also true for better algorithms. Transferring new knowledge to algorithms implies higher computational costs. *iii)* Finally, the size of screens may be also a limiting factor, not only for desktops, but even more for portable devices, such as tablets and smartphones.

Outcomes As result of our discussions, we propose some strategies that may lead to improve our ability to obtain knowledge from large data sources.

An increase in data richness does not (always) need to lead to increased knowledge. However, this can be tacked by designing new and better algorithms for data extraction, or new and better visualization techniques. We believe that, as the knowledge increases, there will be a lower need to develop tools for exploratory visualization, since the number of confounding variables will decrease. And, at the same time, the need for tools for confirmatory analysis, as well as results explanation, will increase. On the other hand, even with new, more powerful algorithms that work on complex data, may get automated results. However, especially in the case of biomedical data, the result can be unexpected (e.g., a tumor located in a difficult region to treat). Therefore, the need of exploratory tools will still be required.

Topic 6—Reusability, Generalizability, Adoptability in BioMedical Visualization

Marc Baaden, **Stefan Bruckner**, Issei Fujishiro, Thomas Höllt, Ingrid Hotz, Shuichi Onami, Guido Reina, Noeska Smit

Motivation The rapid advances in BioMedical visualization have resulted in a large collection of prototypes, visual designs, techniques, and development kits. These tools are domain- and problem-specific, and cannot become more generalized and extensible to other domains. How can we design generalized applications, which are practical and reusable? How to promote and motivate sustainability in BioMedical visualization? Can we design a basis for next generation BioMedical visualization tools? What can we learn from past successes, where prototypes underwent certification processes and were adopted into clinical practice? And, finally, can we determine a way of sharing our designs and results to the parent field of Visualization? How can we contribute back to the broader Visualization community?

Challenges One of the main challenges that manifested itself during our discussions is a perceived lack of incentives for going through the extra effort of creating and maintaining reusable and maintainable visualization software. Papers are still the main outcomes of many research projects, and with successful publication there are few benefits for providing access to the research results, such as the developed software. This is grounded in our community not requiring researchers to make code available alongside a publication, in contrast to practices in other communities. Not having a common framework to dock onto also makes interchange of results, comparison, and cataloguing of the existing solutions unnecessarily hard. Healthy ecosystems, such as R in the statistics community, make it possible to require a compatible implementation alongside a publication, but the fragmentation of our community will prevent such an approach in the near future. Furthermore, in our discussions we found that many of us believe that there is a certain general unwillingness in the visualization community to adopt frameworks developed by other research groups – as opposed to creating one's own solution – leading to the fact that there is a high amount of duplication. An increased focus on larger collaborative community project would lead to a much more efficient use of resources.

A related issue is the lack of funding mechanisms for positions dedicated towards software engineering. As PhD students, for instance, can only use a limited amount of their time for software development tasks (and may even lack the required competence), our discussions clearly indicated that stable financing for such positions is essential. While some countries, such as the UK, have made advances in this direction with an institution at the national level [1], in many others we have clearly identified limited (often only short-term positions tied to specific projects) resources dedicated towards such efforts [3, 14, 13].

One consequence of this lack of resources is also that many visualization software packages suffer from issues related to quality assurance in software engineering, affecting aspects such as stability, interoperability, and ease of integration. Efforts need to be made to create solutions that feature clear and well-defined interfaces, and that can be easily used in a wide variety of different environments. Most visualization researchers, as well as domain researchers [10], never receive formal training in software engineering, making it difficult to design software to be sustainable even mid-term. Best practices regarding permalinks or DOIs for code repositories are rarely followed, so even if software is made available alongside a publication, it is not consistently referenced and thus hard to find.

Outcomes We have identified several measures that could help to address the outlined challenges:

- There should be a curated list of visualization software categorized according to application field. This list should be periodically updated to identify "dead" projects, potentially motivating their developers to resurrect them. Similar repositories are already available in the bioinformatics community, e.g., https://biii.eu/, https://omictools.com, https://bivi.co.
- Visualization software authors should be provided with a set of guidelines on how to make their software available. These guidelines should address the needs of other visualization developers, as well as domain users interested in utilizing the software. We created a draft for these guidelines structured according to multiple tiers which we include below.
- These guidelines and their different tiers could also be used as a basis for awarding "reproducibility" and "adoptability" badges[7], which would be one way to create additional incentives for authors. These badges could also be awarded as part of a contest (similar to the IEEE VIS contest, for instance).
- Overall, the community should strive to award and value contributions in software development, by making sure that these efforts are sufficiently taken into account in, e.g., tenure decisions. In particular, initiatives that span multiple institutions should be rewarded accordingly.
- We also identified a lack of standards for different visualization software frameworks to interoperate, making it difficult to combine or compare their respective functionalities. Here, we propose to form a working group that suggests simple measures that allow for better communication between applications (e.g., standardized import/export of camera parameters, option for standardized 3D controls, formalization of selection and ordering with respect to a data set). The focus of this working group should be the suggestion of simple and feasible mechanisms for this that can be easily adopted, as opposed to a complete, but potentially highly complex standard.
- We could establish a venue for pure application/software output without the need for a scientific paper. Similar to tools journals in other fields, e.g., NAR special issue on web tools[8].
- To increase the impact of VIS tools in the domain, BioMedVis venues could benefit from being co-located with domain events, such as VCBM with MICCAI or BioVis with ISCB/ISMB. Senior visualization people should get more involved with these communities.

---

[7] compare open science badges, https://cos.io/our-services/open-science-badges/

[8] https://academic.oup.com/nar/article/46/W1/W1/5047249

Appendix: List of tiered requirements for a potential badge system

General prerequisites

- Follow FAIR[9] principles
- Consider best software targets
  - Implement solutions in established software frameworks
  - Libraries for easy integration in other tools
  - Web version (no installation)
  - Standalone proprietary desktop tools

Requirement tiers for reproducibility (developer target)

Minimum

- Provide Readme/Brief description of functionality
- Provide exact source code used with paper (e.g., packaged source code, commit id/release when using versioned hosting, etc.)
- Provide necessary instruction for building the code on one common operating system
- Provide a list of prerequisites (soft- and hard-ware)
- Provide citation information
- Provide a Point of Contact: e-mail address
- Provide licensing information
- Provide specification for supported data formats

Better

- Use versioned source code; provide repository link, e.g.
  - Github
  - Gitlab
- Provide rules for contribution
- Provide documentation, e.g., Doxygen
- Provide necessary instruction for building the code for multiple platforms
- Provide input/output/sample dataset
  - Can be data from public repositories
- Provide a script/GUI/Instructions to reproduce images in paper

Best

- Provide advanced User Support, e.g.

---

[9] https://www.go-fair.org/fair-principles/

  - Mailing list
  - Issue tracker
  - Communication platform (Slack)
- Provide maximally permissive licensing
- Provide DOI, e.g., through `zenodo.org`
- Provide datasets used in the paper
- Follow software design principles
  - Unit tests
  - Modular libraries
  - CI/CD
- Provide screencast style tutorials

Requirement Tiers for adoptability (end-user target)
Minimum

- Provide executable/docker/web application
- Provide citation information
- Provide user point-of-contact, e.g., e-mail address
- Provide specification for supported data formats

Better

- Provide a script/GUI/instructions to reproduce images in paper
- Provide a basic manual/tutorial on how to use
- Provide support for common data formats in the domain

Best

- Provide DOI, e.g., through zenodo.org
- Provide advanced User Support. e.g.
  - Mailing list
  - Issue tracker
  - Communication platform (Slack)
- Provide full manual/help system
- Provide screencast style tutorials

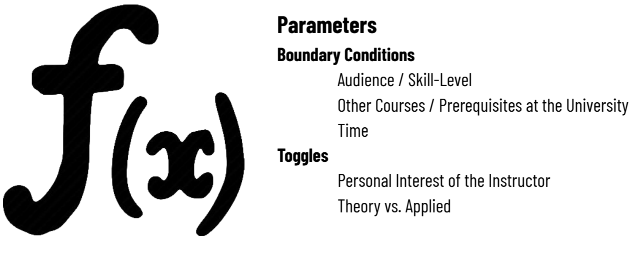
Topic 7—Curriculum: How can we structure a BioMedical Visualization course?

Johanna Beyer, Daniel Haehn, Thomas Höllt, **Ingrid Hotz**, Timo Ropinksi, Thomas Schultz, Pere-Pau Vázquez

Motivation A possible approach to creating a formalization of biomedical visualization is to use teaching as a driver for this formalization. Also, we observe that the challenges while creating a curriculum are similar to the general difficulties of formalization of our field. We investigate how we can structure a biomedical visualization course that covers the diversity of students, content, and research interests. Our goal is to teach the full spectrum of students from computer science, technical biomedical fields, as well as the life sciences.

Challenges Coming up with a curriculum requires taking into account the boundary conditions of the course to find the right content for the right audience. We identify the following challenges for teaching biomedical visualization:

1. Our students come from diverse fields with different backgrounds and different goals. A curriculum for computer science students should focus on technical visualization methods and programming. In contrast, the curriculum for life science students should focus on the application of visualization in the domain.
2. A BioMedVis course must include the basics of the area the students are least comfortable. This can, for example, be the basis of hypothesis formation for CS students or basic scripting for the life science students.
3. It is necessary to find the right level between teaching basic knowledge in visualization and the specific demands of the applications. There are only a few application-specific techniques and methods that are only applicable in BioMedVis and not in a general VIS context.
4. Different universities have different courses that can be used as prerequisites. We need to adjust a curriculum to other courses available in the environment and the level of the course.



Parameters

Boundary Conditions
- Audience / Skill-Level
- Other Courses / Prerequisites at the University
- Time

Toggles
- Personal Interest of the Instructor
- Theory vs. Applied

Figure 7: A formalized curriculum has to respect the boundary conditions of a course and take into account toggles based on the instructor's preferences.

Existing Materials We found the following interesting links to the existing materials. These links are a good starting point and a valuable source for building a curriculum.

- Bernhard Preim's Lecture `https://medvisbook.com/courses-and-tutorials/prof-preims-international-medvis-lectures/`
- Bernhard Preim's Book `https://medvisbook.com/about/`
- Course at Delft `https://www.tudelft.nl/en/health/education/medical-visualization/`
- Visualization in Medicine Book `https://www.sciencedirect.com/book/9780123705969/visualization-in-medicine`
- UChicago MS in Biomedical Visualization `https://catalog.webhost.uic.edu/gcat/cat1012archive/AHBVIS.shtml`
- Richard Robb's Book `https://www.amazon.com/Biomedical-Imaging-Visualization-Analysis-Richard/dp/0471283533`
- Omic's course `https://nbisweden.github.io/workshop_omics _integration/`

Outcomes The generation of a curriculum is not a rigid process. To cope with the challenges identified above, we propose to build up the curriculum from basic building blocks. To create a learning path, a teacher selects the appropriate blocks for a specific course. This involves defining a custom learning path throughout a complex space of building blocks, compared in Figure 8. We have identified three main topics that make up the building blocks: basic techniques, domain-specific techniques, and use cases. The use cases represent an essential part of teaching the theory with ties to complex domain applications.

These building blocks should include basic and advanced concepts that we can also tune according to the boundary conditions and toggles. An Overview also helps at the beginning of the course to keep students motivated and hungry for knowledge.

Example Curriculum: We defined the following building blocks to cover many aspects of biomedical visualizations.

- Overview/Motivation/Intro to the field
  - Tasks
  - Diagnosis
  - Surgery/Intervention
  - Clinical Research
  - Training / Education
- Basic Blocks
  - Data Types
    - Data Structure
    - Interpolation / Data Structure

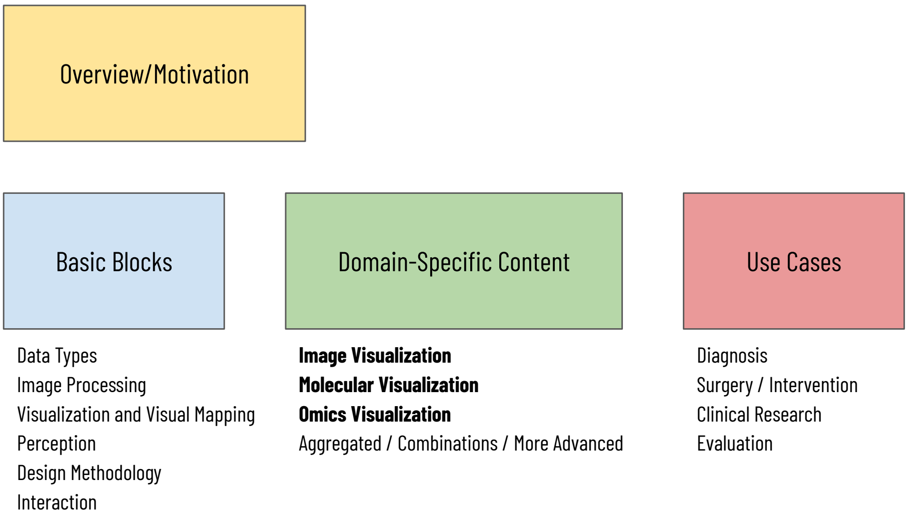


Figure 8: We distinguish between different types of building blocks. The Overview/Motivation allows increasing the students' interest in biomedical visualization at the start of the course. Basic blocks can then adjust different audiences to similar starting levels. Domain-specific content includes the main areas of biomedical visualization. Finally, use-cases tie in the theory with domain applications.

    - File Formats
    - Basic Data Acquisition
  - Image Processing
    - Alignment/Registration
    - Segmentation
    - Filtering
  - Visualization and Visual Mapping
    - Direct Volume Rendering
    - Isosurface
    - Flow Vis
      - Streamlines
      - Particles
    - Info Vis
    - Dimensionality Reduction
  - Perception
  - Design Methodology
    - Evaluation
  - Interaction
    - Advanced Queries
    - Brushing/Linking

        - Pan/Zoom
        - Filter

- Tasks / Drivers / Use-Cases
    - Diagnosis
    - Surgery / Intervention
    - Clinical Research
    - Evaluation
- Image Visualization
    - Structural
        - 2D/3D/4D
        - Projection Methods
            - MPR/CPR
            - Virtual Endoscopy
        - Diffusion Imaging
            - Tractography/Glyphs
    - Functional
        - fMRI
        - Bloodflow
        - Nuclear Imaging
- Molecular Visualization
    - Rendering Representations
        - Ball and Stick
        - Surfaces
        - Meso Scale
    - Molecular Surfaces
    - Abstraction
    - Tunnels
    - Dynamics
- Omics Visualization
    - Sequence Alignment
    - Sequence Search
- Aggregated / Combinations / More Advanced
    - Multi-Modal Visualization
    - Ensemble Visualization
    - Comparative Visualization
    - Quantitative Visualization
    - Illustrative Visualization

Retrospective The field of biomedical visualization is very complex. It is seemingly impossible to define the ultimate curriculum. Therefore, we propose a repository of building blocks that can be combined based on the specific boundary conditions and toggles of an individual course.

Topic 8—Dissemination: Driver of Formalization in BioMedical Visualization

DANIEL HAEHN, BARBORA KOZLÍKOVÁ, **Martin Krzywinski**,
JOHANNES SORGER, HSIANG-YUN WU

Motivation Dissemination is a central part of academia, transferring knowledge to peers, students, and the public. Without dissemination, research is created in a bubble and innovation is stiffled. To consistently create effective dissemination, the researcher needs a way to formulate their message. Providing form to the message implicitly drives a formalization of the underlying work through common vocabulary, metaphors, and conventions. Dissemination can thus be seen as a driver of formalization.

During this meeting, we discovered that the community reflected the assumption that the output of research feeds into the process of dissemination and the process of formalization (9a). While this is undoubtedly true–the creation of new abstractions and schemas for advancing thinking and analysis does arise out of research–the dissemination process has an equally powerful (or even greater) influence on the creation of formalisms.

Challenges Because it is extraordinarily difficult to meaningfully create a deliverable that generally improves on the process of how research impacts formalism, we aimed to instead provoke discussion by proposing a light-weight formalization for dissemination.

Motivated by the initial question of how we can improve science dissemination, we aimed to establish the forms which dissemination can take, which finally led us to create a formalization of dissemination itself.

The driver of our discussions were questions that address general aspects of the communication process (1) the presenter, (2) the audience, (3) the purpose of dissemination (intended outcome), (4) the medium (e.g., video, audio) and platform (e.g., social media, meeting), and finally (5) the overall packaging, which includes aspects such as tone, style, length, lifespan, level of engagement, interactivity, and accessibility.

We consider that at this level of abstraction it is not necessary contextualize the scheme to BioMedVis because this is something that is encapsulated in the payload of the message, which is always up to the researcher to determine.

Outcomes

Schema Our formalism resulted in a dissemination framework. The framework can be split into seven aspects of dissemination, each describing an essential part of dissemination (9), e.g., who is the sender, who is the receiver of the message, through which medium is the message conveyed? Each aspect is described by a list of representatives (e.g., "To": the public, peers, students, etc). We see the list of aspects as exhaustive, while the list of representatives can be further extended.

Special consideration need to be given to "the message", and "the package". The message is the domain-specific payload of the dissemination. Itself it can be broken down into sections (e.g., traditionally: introduction, aims, methods, results, conclusion) but this is up to the researcher to arrange and should be done so to match the profile of the dissemination. The package describes the way in which the message is delivered, independent of the medium. The package contains several sub-aspects, such as the style and tone of a message, the length, the engagement factor (immersive/interactive), etc.

The initial intent of this framework was to make the "design space" of dissemination explicit. However, the entire framework can also be seen as a checklist, clarifying (to the sender of a message), who they want to reach and how they want to reach them – thus acting as a guideline for science dissemination. Making the extent of dissemination explicit, opens several options for applying this framework.

A single dissemination profile (e.g., for a tutorial video) would be specified by items in the checklist. A group of multiple profiles would be considered a "campaign", which is a collection of disseminations over various time and various media and platforms. For example, one set of disseminations might aim to inspire while another clarify.

Online platform We propose a web-based platform that can act as a checklist builder and repository of effective examples, tips and caveats for specific combinations of settings (e.g., public inspiration video of less than 2 minutes in length) (10a). This functionality would allow researchers to plug into communication strategies and gain awareness of innovative options for sharing their knowledge and research.

Extensibility This framework would classify existing disseminations, assigning an explicit dissemination profile to them. The profile space itself could be formalized more rigorously into structured metadata (similar to exif data in images and subsections such as IPTC) that would be attached to each dissemination (and drive semantic searching).

As a domain of study becomes more mature, our dissemination checklist would be expanded to include a list of all (or merely those that are relevant) formalizations in the field. The researcher would use this list to identify those that are adhered to in the dissemination and, equally importantly, those that could be adhered to but are not (thereby embed in the checklist a challenge of how best to adhere to formalisms). This would help creating communication that is consistent and emphasize the value of closing the loop from dissemination to formalization.

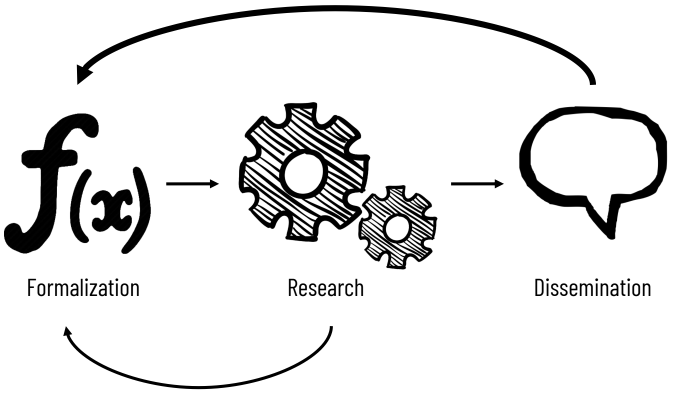
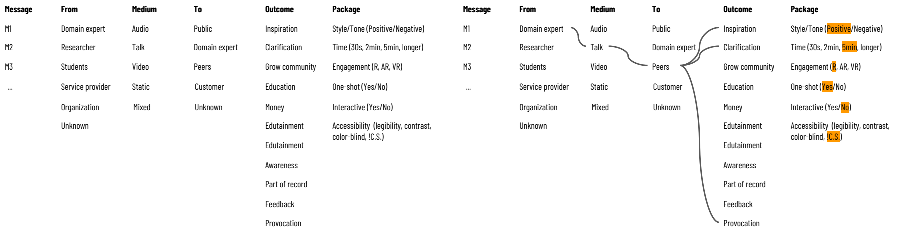


Figure 9: Relationship between formalization, research and dissemination. Formalization is influenced by both research and dissemination and we argue that the latter has a stronger effect.



Message	From	Medium	To	Outcome	Package
M1	Domain expert	Audio	Public	Inspiration	Style/Tone (Positive/Negative)
M2	Researcher	Talk	Domain expert	Clarification	Time (30s, 2min, 5min, longer)
M3	Students	Video	Peers	Grow community	Engagement (R, AR, VR)
...	Service provider	Static	Customer	Education	One-shot (Yes/No)
	Organization	Mixed	Unknown	Money	Interactive (Yes/No)
	Unknown			Edutainment	Accessibility (legibility, contrast, color-blind, !C.S.)
				Edutainment	
				Awareness	
				Part of record	
				Feedback	
				Provocation	

(a)

Message	From	Medium	To	Outcome	Package
M1	Domain expert	Audio	Public	Inspiration	Style/Tone (Positive/Negative)
M2	Researcher	Talk	Domain expert	Clarification	Time (30s, 2min, 5min, longer)
M3	Students	Video	Peers	Grow community	Engagement (R, AR, VR)
...	Service provider	Static	Customer	Education	One-shot (Yes/No)
	Organization	Mixed	Unknown	Money	Interactive (Yes/No)
	Unknown			Edutainment	Accessibility (legibility, contrast, color-blind, !C.S.)
				Edutainment	
				Awareness	
				Part of record	
				Feedback	
				Provocation	

(b)

Figure 10: Collected factors for the campaign builder, including (a) a discussed checklist, and (b) an example.

Conclusions

During this *4-day seminar*, we brought together 25 visualization experts from diverse institutions, backgrounds and expertise to discuss, identify, formalize, and document the specifics of our field. The main aim of the seminar was to formalize the current status of the field of BioMedical Visualization and to determine future challenges. A free and open discussion between all participants has been sought—opening novel, interesting paths for the future and contributing to new, creative collaborations. This has been a great opportunity to cover a range of relevant and contemporary topics, and as a systematic effort towards establishing better fundaments for the field and towards determining novel future challenges. Eight main topics have been discussed in the form of working groups and were presented in this report. These are the following:

1. Differences between BioMedical Visualization and the parent field of Visualization.
2. Where formalization is needed for BioMedical Visualization?
3. Best Practices in BioMedical Visualization
4. Integration & Standardization in BioMedical Visualization
5. Complexity vs. Knowledge in BioMedical Visualization
6. Reusability, Generalizability, Adoptability in BioMedical Visualization
7. Curriculum: How can we structure a BioMedical Visualization course?
8. Dissemination: Driver of Formalization in BioMedical Visualization

List of Participants

- Marc Baaden, CNRS Paris,France
- Johanna Beyer, Harvard University, USA
- Stefan Bruckner, University of Bergen, Norway
- Issei Fujishiro, Keio University, Japan
- Daniel Haehn, University of Massachusetts Boston, USA
- Helwig Hauser, University of Bergen, Norway
- Thomas Höllt, Leiden University Medical Center / TU Delft, Netherlands
- Ingrid Hotz, Linköping University, Sweden
- Tobias Isenberg, INRIA, France
- Barbora Kozlikova, Masaryk University, Czech Republic
- Michael Krone, University of Tübingen, Germany
- Martin Krzywinski, Canada's Michael Smith Genome Sciences Center at BC Cancer, Vancouver, Canada
- Lars Linsen, University of Münster, Germany
- David Mayerich, University of Houston, USA
- Shuichi Onami, RIKEN Center for Biosystems Dynamics Research, Japan
- Renata Raidou, TU Wien, Austria
- Guido Reina, University of Stuttgart, Germany
- Timo Ropinski, Ulm University, Germany
- Thomas Schultz, University of Bonn, Germany
- Noeska Smit, University of Bergen, Norway
- Johannes Sorger, Complexity Science Hub Vienna, Austria
- Pere-Pau Vázquez, Universitat Politecnica de Catalunya, Spain
- Baoqing Wang, Keio University, Japan
- Hsiang-Yun Wu, TU Wien, Austria

Figure 11: The participants of NII Shonan Meeting No.167 on “Formalizing Biological and Medical Visualization”.

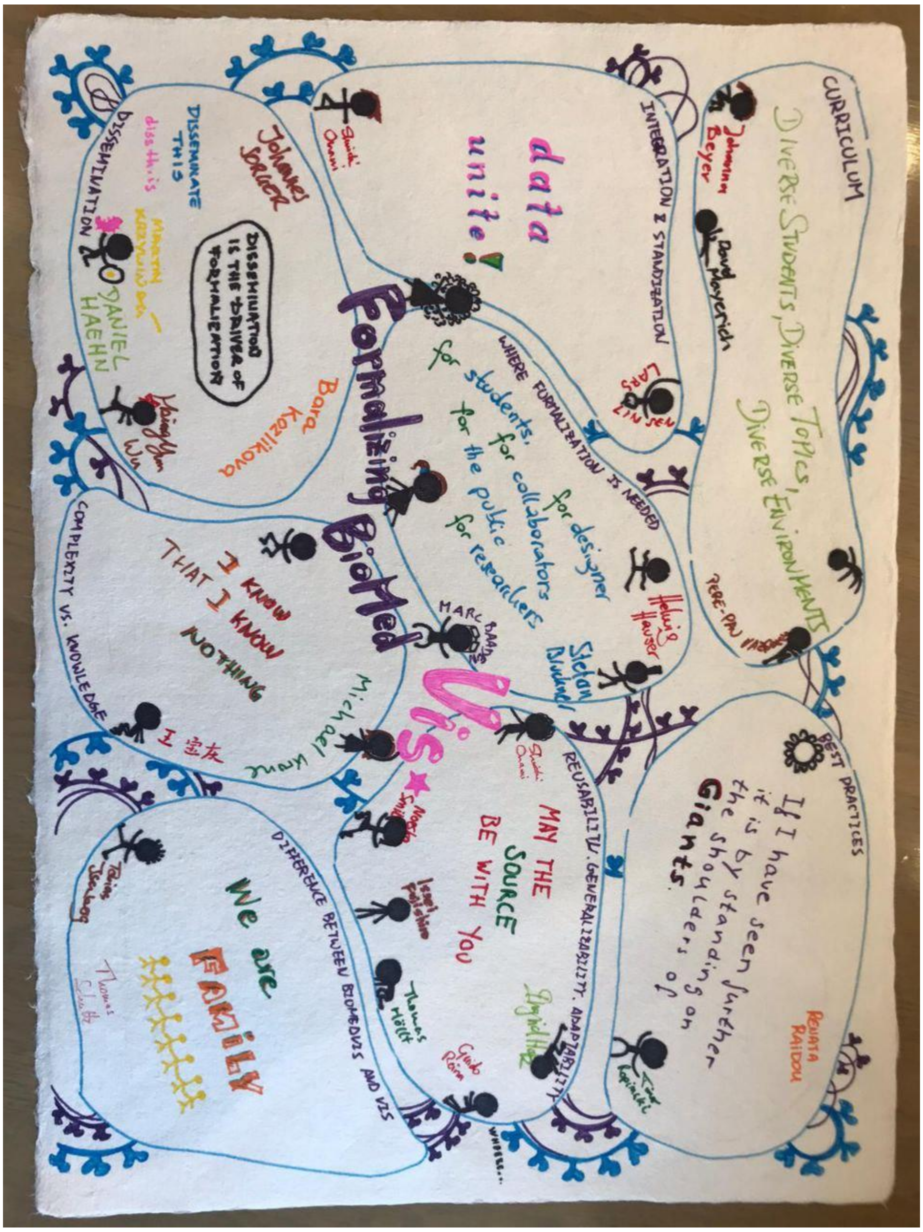
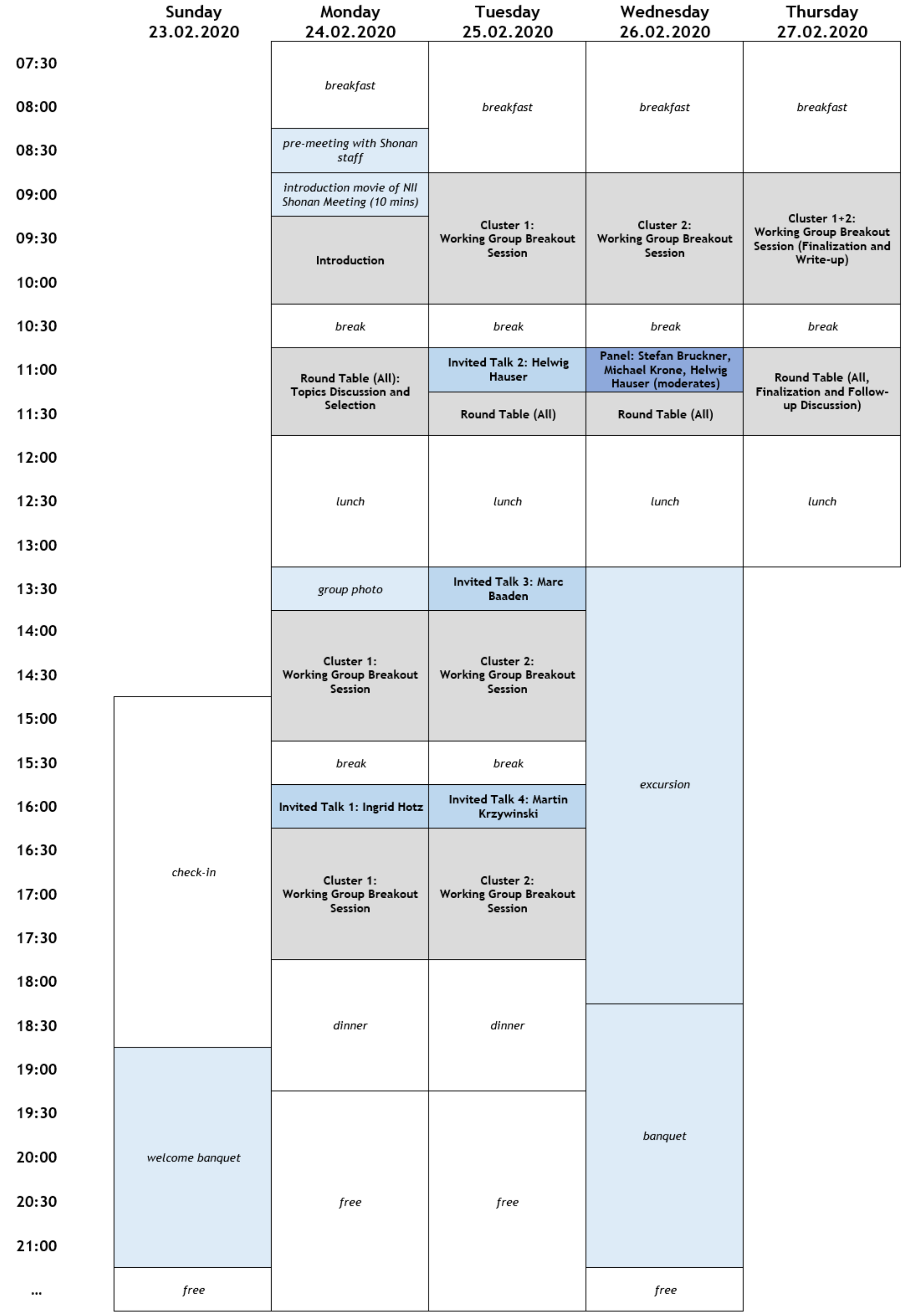


Figure 12: The poster of NII Shonan Meeting No.167 on “Formalizing Biological and Medical Visualization”.

Meeting Schedule



	Sunday 23.02.2020	Monday 24.02.2020	Tuesday 25.02.2020	Wednesday 26.02.2020	Thursday 27.02.2020
07:30		*breakfast*	*breakfast*	*breakfast*	*breakfast*
08:00					
08:30		*pre-meeting with Shonan staff*			
09:00		*introduction movie of NII Shonan Meeting (10 mins)*	**Cluster 1: Working Group Breakout Session**	**Cluster 2: Working Group Breakout Session**	**Cluster 1+2: Working Group Breakout Session (Finalization and Write-up)**
09:30		**Introduction**			
10:00					
10:30		*break*	*break*	*break*	*break*
11:00		**Round Table (All): Topics Discussion and Selection**	**Invited Talk 2: Helwig Hauser**	**Panel: Stefan Bruckner, Michael Krone, Helwig Hauser (moderates)**	**Round Table (All, Finalization and Follow-up Discussion)**
11:30			**Round Table (All)**	**Round Table (All)**	
12:00		*lunch*	*lunch*	*lunch*	*lunch*
12:30					
13:00					
13:30		*group photo*	**Invited Talk 3: Marc Baaden**	*excursion*	
14:00		**Cluster 1: Working Group Breakout Session**	**Cluster 2: Working Group Breakout Session**		
14:30					
15:00	*check-in*				
15:30		*break*	*break*		
16:00		**Invited Talk 1: Ingrid Hotz**	**Invited Talk 4: Martin Krzywinski**		
16:30		**Cluster 1: Working Group Breakout Session**	**Cluster 2: Working Group Breakout Session**		
17:00					
17:30					
18:00		*dinner*	*dinner*		
18:30				*banquet*	
19:00	*welcome banquet*				
19:30		*free*	*free*		
20:00					
20:30					
21:00					
...	*free*			*free*	

References

[1] A. Brett, M. Croucher, R. Haines, S. Hettrick, J. Hetherington, M. Stillwell, and C. Wyatt. Research Software Engineers: State of the Nation Report 2017, Apr. 2017.

[2] S. Bruckner and M. E. Groeller. Volumeshop: An interactive system for direct volume illustration. Technical Report TR-186-2-05-04, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Favoritenstrasse 9-11/E193-02, A-1040 Vienna, Austria, Apr. 2005. human contact: technical-report@cg.tuwien.ac.at.

[3] de-RSE e.V. - Society for Research Software. de-RSE.org - Research Software Engineers (RSEs) - The people behind research software.

[4] L. S. Elting, C. G. Martin, S. B. Cantor, and E. B. Rubenstein. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. *BMJ*, 318(7197):1527–1531, Jun 1999.

[5] A. Kanitsar, D. Fleischmann, R. Wegenkittl, P. Felkel, and M. E. Groeller. Cpr - curved planar reformation. Technical Report TR-186-2-02-06, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Favoritenstrasse 9-11/E193-02, A-1040 Vienna, Austria, Mar. 2002.

[6] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):495–513, March 2013.

[7] G. Kindlmann and C. Scheidegger. An algebraic process for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2181–2190, Dec 2014.

[8] F. Lindemann and T. Ropinski. About the influence of illumination models on image comprehension in direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1922–1931, 2011.

[9] T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, 2015.

[10] O. Philippe, M. Hammitzsch, S. Janosch, A. van der Walt, B. van Werkhoven, S. Hettrick, D. S. Katz, K. Leinweber, S. Gesing, S. Druskat, S. Henwood, N. R. May, N. P. Lohani, and M. Sinha. softwaresaved/international-survey: Public release for 2018 results, Mar. 2019.

[11] B. Preim and C. P. Botha. *Visual Computing for Medicine: Theory, Algorithms, and Applications*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2 edition, 2013.

[12] M. Tarini, P. Cignoni, and C. Montani. Ambient occlusion and edge cueing for enhancing real time molecular visualization. *IEEE transactions on visualization and computer graphics*, 12(5):1237–1244, 2006.

[13] The Netherlands Research Software Engineer community. The Netherlands Research Software Engineer community.

[14] The US Research Software Engineer Association. US-RSE.

[15] V. van Unen, T. Höllt, N. Pezzotti, N. Li, M. J. Reinders, E. Eisemann, F. Koning, A. Vilanova, and B. P. Lelieveldt. Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types. *Nature communications*, 8(1):1–10, 2017.

[16] J. Vuong, S. Kaur, J. Heinrich, B. K. Ho, C. J. Hammang, B. F. Baldi, and S. I. O'Donoghue. Versus—a tool for evaluating visualizations and image quality using a 2afc methodology. *Visual Informatics*, 2(4):225 – 234, 2018.

[17] S. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *Proceedings of the First IEEE Conference on Visualization: Visualization '90*, pages 139–143, Oct 1990.