Building an Ontology of Visualization

D.J. Duke* and K.W. Brodlie†
School of Computing, University of Leeds, UK

D.A. Duce‡
Department of Computing, Oxford-Brookes University, UK

ABSTRACT

Recent activity within the UK National e-Science Programme has identified a need to establish an ontology for visualization. Motivation for this includes defining web and grid services for visualization (the 'semantic grid'), supporting collaborative work, curation, and underpinning visualization research and education. At a preliminary meeting, members of the UK visualization community identified a skeleton for the ontology. We have started to build on this by identifying how existing work might be related and utilized. We believe that the greatest challenge is reaching a consensus within the visualization community itself. This poster is intended as one step in this process, setting out the perceived needs for the ontology, and sketching initial directions. It is hoped that this will lead to debate, feedback and involvement across the community.

CR Categories: I.3.8 [Computer Graphics]: Applications; I.3.m [Computer Graphics]: Miscellaneous;

Keywords: Ontology, Semantic Web, Grid

1 MOTIVATION

Visualization is an interdisciplinary activity. It involves domain specialists with data that they wish to present, understand, and/or explore; it involves the resources needed to produce visualization, both computing (software systems, HPC, networking) and human (visualization specialists). And in many domains in which visualization is deployed, it routinely involves collaboration between multiple users and systems. One way of making sense of this complexity is to think in terms of the kind of interface involved:

Human-Human: Dialogue between domain experts working on a problem, or between domain and visualization experts.

Human-Computer: Description of data to be visualized, the representation required, and/or the process to be used.

Computer-Computer: Description of visualization services and data models provided, e.g., by distributed services.

Computer-Human: Display (by one or more modalities) of the representation to users of the system.

At each interface, there is an issue of how communication between the partners is expressed and grounded. While visualization remained an activity carried out by one or two experts within a local system, the issue of grounding appeared only as a need for common data formats, and a shared but informal language for describing visualization. Systems integration was accomplished through hard-coded assumptions on data format coupled with use of specific protocols, e.g. the use of MPI to support a pipeline distributed between two sites and passing data within a given format for structured grids. Present interest in formalizing shared understanding as an ontology stems from three developments:

Collaboration: support for users interacting on a problem through shared resources (data, processes), including inter-operation of visualization systems.

Components and Services: provision of capabilities as software components and web/grid services, requiring description of resource semantics to support discovery and composition.

Curation: description of data and processes to enable archiving of experimental analyses.

A fourth, orthogonal, concern is to provide a more structured basis for education in visualization. The UK is investing significant effort in developing infrastructure, systems and principles to support large-scale on-line collaborative research (e-Science). The importance of visualization in this activity was underlined at a meeting [1] that recommended as its highest priority outcome the development of a visualization ontology to address the issues listed above. This poster reports, on behalf of all who attended, the outcome of a workshop to begin the task [2]. It is intended to raise the issue of an ontology in the international arena, and aid in discovering what other programmes and groups may be developing in this direction.

2 COMPONENTS OF AN ONTOLOGY

Following guidelines for ontology development, and the cycle of interfaces mentioned above, an initial set of concepts and relationships were identified. These are but a small sample of the concepts used across visualization, but it was useful to explore what kind of organization could be discerned.

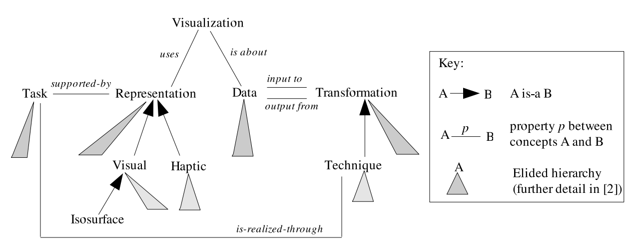

Figure 1: Top level components of the visualization ontology

The result, sketched in Figure 1, suggested that the overall ontology can be divided into four groups of concepts.

task and use: why are the data being visualized, and what is being done with the visualization?

representation: how are the data presented?

process: how can the transformation from data to representation be expressed in terms of a given set of primitives?

data: how are the data processed through the various stages of the transformation process expressed and organized?

*e-mail:djd@comp.leeds.ac.uk
†e-mail:kwb@comp.leeds.ac.uk
‡e-mail:daduce@brookes.ac.uk

Three levels of concern were also noted:

Conceptual: What is the visualization that is to be performed, independent of the visualization system?

Logical: What configuration of entities (data, processes) are required, independent of physical resources?

Physical: What physical resources are used?

These may help organize the ontology, providing distinct but inter-connected languages for declarative, procedural and system-dependent aspects. For example, a contour-plotter in the AVS toolkit might be an instance of a service (physical-level concept), which is in turn a subclass of a logical 'filter', in turn a subclass of the transformation concept.

3 TECHNICAL BASELINE

A full survey with citations is beyond the scope of a poste. Rather, the aim here is to point to the variety of models and taxonomies that are available under the grouping of concepts given above; [2] provides a number of key references. Models of process and data tend to be specific to visualization, and most have been developed after the McCormick report appeared. Models of representation and use can be found work on semiotics, HCI, and computer graphics, and have a longer history.

3.1 Process Models

The notion of a 'visualization pipeline' as a conceptual model owes much to the work of Haber and McNabb; its use as a logical model, popularized by the AVS system, established a vocabulary of concepts for building and managing visualization applications. Linking 'components' to realize an overall 'service' is not an idea specific to visualization; the semantic web community is developing a promising *web service ontology* (OWL-S) for tasks including discovery, composition, interoperation and execution.

However, in addition to pipelines, a broad vocabulary should also accommodate models as diverse as spreadsheets, 'spray' rendering and event-based composition. It not yet clear how deeply such models can be unified within a 'generic' ontology such as OWL-S; can the concepts of the ontology capture important differences in process models, or will the process part of the ontology require disjoint branches for different techniques (at least at the physical level)? Further work is required to understand how OWL-S might be used in the process part of the visualization ontology.

3.2 Representation

Tufte's books on visual representation illustrate how the use of pictures to understand 'data' predates modern interest by some millenia! Tufte builds on Bertin's 1967 semiology of graphics that sets out a taxonomy of marks and representation systems; both in turn draw on the semiotic approach established by Peirce around a century ago. An interesting bridge between this work and the more formal approach of Brodlie et.al. (see below) was set out by Keller and Keller in a book on visual cues, in which example representations are indexed by properties of the data field (numbers of dependent and independent variables), and by the visualization goal.

Data (field) and representation are coupled; some models of representation are specific to classes of data. Thus for example work in flow visualization and graph visualization have distinct categories of representations. Many visualization problems call for the composition of multiple representations; an ontology will need concepts to distinguish for example graphical superimposition from separate representations linked via some common frame of reference.

3.3 Data Models

Regularities (geometric and/or topological) have long been used to structure and organize data, and these all underpin fundamental distinctions, e.g. between structured and unstructured grids. Several taxonomies of data and its organization have been been developed, in particular work by Butler, Pendley, and Bergeron and Kao; other approaches are also notable, for example the lattice model of Hibbard et.al. is significant in addressing error and uncertainty.

A classification scheme for data proposed by Brodlie emphasises the importance of the underlying field, i.e. the phenomenon that is captured within the data. Building on this, Tory and Möller have developed a visualization taxonomy based on data models rather than data. These two groups of taxonomies (e.g. Bergeron and Kao, and Brodlie), are complimentary; the former concentrate on how data is structured, while the latter the link between data and representation. A synthesis of these two contributions should help to develop the 'data' branch of the top-level ontology.

3.4 Users, Tasks and Goals

In 1993 Bergeron suggested that at an abstract level visualization goals could be classified as descriptive, analytic or exploratory. This provides a useful starting point for linking visualization with research in HCI and graphics on presentation synthesis. Here work e.g. by Feiner and Zhou, building on the seminal contribution of Mackinlay's APT and Casner's BOZ systems, has lead to more detailed and operational taxonomies of visual task. Keller and Keller's visual cues book also uses taxonomies, in this case of action (task) and representation, building on work by Wehrend. Also relevant are models such as the operator-function approach of Chi and Riedl that locates user interaction within a conceptual 'pipeline' model .

To borrow the language of MacEachren, these taxonomies define 'private' views of data. Thinking of visualization as a collaborative activity should impact at least two aspects of the ontology: how the meaning of representations is grounded within a particular community, and how task, process and representation are shared during collaboration. There is relevant work in the CSCW community on cooperation, coordination and communication.

4 CONCLUSION

An ontology provides a vocabulary by which users and systems can communicate. It is only useful if it reflects the consensus of a community. We have sketched how a visualization ontology might be organized, and pointed to significant bodies of work on which it might draw. Through the semantic web and the grid, some form of ontology will inevitably be needed to describe the interface to visualization services. It is up to the community to decide at what point, and at what level, to collaborate in such an activity.

The approach outlined here recognises different uses for an ontology. High-level views (the cycle of activity, and levels of concern) have been found useful in organizing some of the baseline. Is this the right way to start? Are the abstract models helpful? What other research can help populate the ontology? How should the ontology evolve? We hope that activity in the UK will build on the preliminary meeting and address these questions. More importantly, we seek comments, criticisms on the work, and suggestions for how to reach and involve the international community.

REFERENCES

[1] National e-Science Centre. Visualization for e-science. http://www.nesc.ac.uk/esi/events/130/workshop_report.pdf.

[2] National e-Science Centre. Visualization ontologies. http://www.nesc.ac.uk/talks/393/vis_ontology_report.pdf.