

Visualization Ontologies

Report of a Workshop held at the National e-Science Centre

7/8 April 2004

1. Background

A workshop on visualization for e-science, held at NeSC in January 2003, identified the need to establish an ontology for visualization. The meeting on the 7th and 8th of April was organized to investigate the structure for such an ontology, and to relate it to examples. As it matures, the ontology will provide a common vocabulary for describing visualization data, processes, and products, and is intended to support:

- the description and discovery of web services
- sharing of process models (pipelines) between visualization developers and users
- curation and provenance management of visualization processes and data
- collaboration and interaction across distributed sites

The meeting attracted 18 delegates, representing a range of visualization communities. It was organized by Prof. Brodlie (Leeds), Prof. Duce (Oxford-Brookes) and Dr. Duke (Leeds).

2. Format of Workshop

The workshop opened with an outline of the objectives from Prof. Ken Brodlie, followed by a presentation by Prof. David Duce. This set out the background to the workshop, covering:

- Recent work on visual web services within the e-Science gViz project and elsewhere
- General issues – for example, one ontology versus different “worlds”
- Related work, in particular the EU “Monet” programme for mathematical services
- Existing taxonomies that might serve as starting points

The latter included work by Brodlie (1992, 1993), and by Tory and Möller (2004) (from Simon Fraser University). The slides are available at the following URL:

http://wwwcms.brookes.ac.uk/~p0072753/ontology_ws/talk1/overview.htm

An invited talk was then presented by Dr. Ivan Herman, Head of W3C Offices, based at CWI Amsterdam. Dr. Herman outlined the rationale and concepts underlying the “semantic web”, in particular the role of the Resource Description Framework (RDF – Manola and Miller, 2004) and Web Ontology Language (OWL – McGuinness and van Harmelen, 2004) in providing means for making statements about web resources. The evolution of resource description from RDF through to OWL was motivated by the issue of what kinds of statements could be represented, and how one might want to reason about resources. The talk explored both how these technologies could support the objectives of the workshop, and some of the practical issues and difficulties that have been encountered in developing or capturing ontologies. The slides of this talk are located at the URL:

<http://www.w3.org/2004/Talks/0407-Edinburgh-IH/>

Following the presentation, the workshop split into groups to discuss and refine the role of ontologies in visualization. The following summarises the reports back to the plenary session and the subsequent discussion.

Using the model of (Brodlie et al) as a starting point, three potential groups of ontology users were identified: visualization service provider, visualization user, data service provider. Bergeron’s classification scheme (Bergeron, 1993) likewise provided three uses for visualization: presentation

(you know what is there), analytical (you know what you are looking for), discovery (you have no idea what you are looking for). Astronomy users, for example, usually describe problems using terms from their application domain; they are looking for answers in terms of tools, data sources and algorithms. Some users will describe their problems in terms of a file format, e.g. I have a NetCDF file, how can I visualize it?

Building on these insights, the meeting considered ways that the ontology might be used:

- Describing the configuration of a visualization system (software and hardware), for use in recording provenance of published representations, supporting discovery of web/grid services, and for providing a level of portability between different environments.
- Directing visualization research, providing a clear definition of terms, and helping identify topics that have not been well covered. It could also support visualization education, providing the basis for a curriculum.
- Supporting domain-specific tasks, for example, by serving as the source/target for transformations to/from domain data.

These roles for the visualization ontologies were subsequently mapped into a 2-by-2 square, with each quadrant defining a type of communication in terms of the entity that is constructing the input and that using the output (Figure 1):

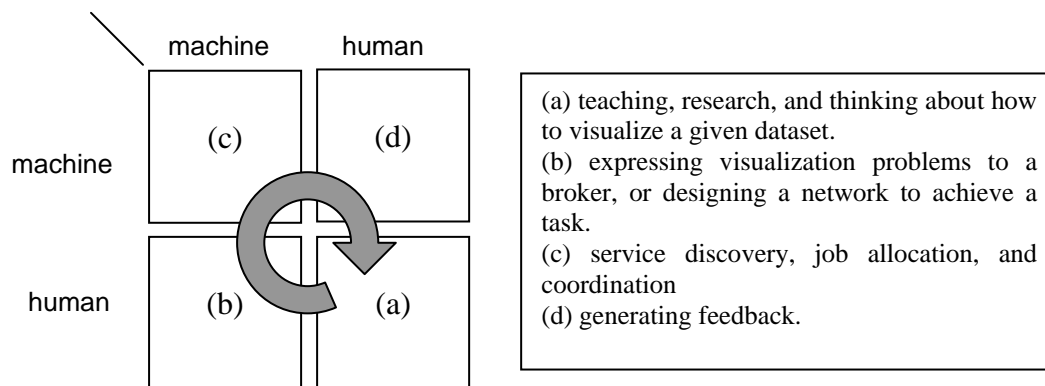


Figure 1 – the HM-square

It was noted that in the “four-box” model, visualization activity could be understood as a cycle, or as a spiral (expanding out in time): one first has to understand what one wants from visualization (H-H), phrase a service request or develop a pipeline and prepare input (H-M), allow a machine (possibly a distributed service) to operate on the input (M-M), and then interpret the output (M-H). The output hopefully gives new insight into the original problem, and/or leads to a refined or modified visualization problem.

There was some discussion of whether there is any fundamental difference between scientific visualization and information visualization. In the former there is an intrinsic geometry, but in the latter there is no geometric context. Information visualization is about discovering relationships between variables; in scientific visualization the relationships are known. The group discussed the representation used in the SequoiaView program (SequoiaView, 2004) for visualizing disc utilization that had been shown in Prof. Duce’s talk and how one might characterise the kinds of data for which that visualization technique would be appropriate. There is certainly some metric associated with data that maps onto area, and also some notion of hierarchy that guides the recursive subdivision.

Discussion also highlighted the need for a thesaurus to relate the ways that people talk about visualization in different fields. One could envisage describing a problem using application domain terminology and expressing solutions using visualization terminology.

3. Building an Ontology: where are the raw materials?

Given the consensus on the role of the ontology, the next session of group work was given over to identifying how the ontology might be developed, and what existing work might be built on.

Group 1: What do we have already?

Existing work on taxonomies and ontologies was identified and located within the H-M four box matrix.

	M	H
M	Brodlie's E-notation MathML/OpenMath Fibre bundle theory/OpenDX XML Schemas OWL-S (includes workflow) CC-PP – hardware resources UPROF Dublin Core SVG/X3D (primitives, etc.) Eades graph layout algorithms/ Herman's TVCG review paper	Brodlie's O-notation
H	Brodlie's E-notation Fibre bundle theory/OpenDX Dublin Core Lattice notation – Bergeron/ Grinstein VisAD – data model NetCDF SVG/X3D (primitives etc.)	Brodlie's E-notation

The group also listed a number of existing systems which may serve as sources of inspiration. All have made some decisions on conceptual models and organisation of functionality.

IRIS Explorer AVS OpenDX Khoros VTK VisAD PV3	PVWAVE Amira GIS systems Workflow systems/languages/OWL-S Excel/Open Office (for business graphics) Conceptual models used in teaching
---	---

The group noted that we appear to be missing a way in which to describe how one wants to *manipulate* data.

There was also discussion of whether an ontology should be judgemental in some sense? For example mapping discrete data values (e.g. share prices at the ends of a set of years) into a line graph (where points on the line other than those corresponding to the actual data values have no meaning). The view was that an ontology should say what is an invalid mapping, but ultimately the issue is a cognitive one.

Group 2: The 'E' Notation

The main activity of the group was to revisit the 'E' notation, first developed at the AGOCC Visualization meeting in 1992 (see Brodlie, 1992). This was seen as a useful classification of the type of model underlying sampled data, but it failed to capture detail of how the samples were distributed, nor how the visualization was carried out – both fundamental issues in visualization.

The group developed a high level model of the visualization process, which is illustrated in Figure 2.

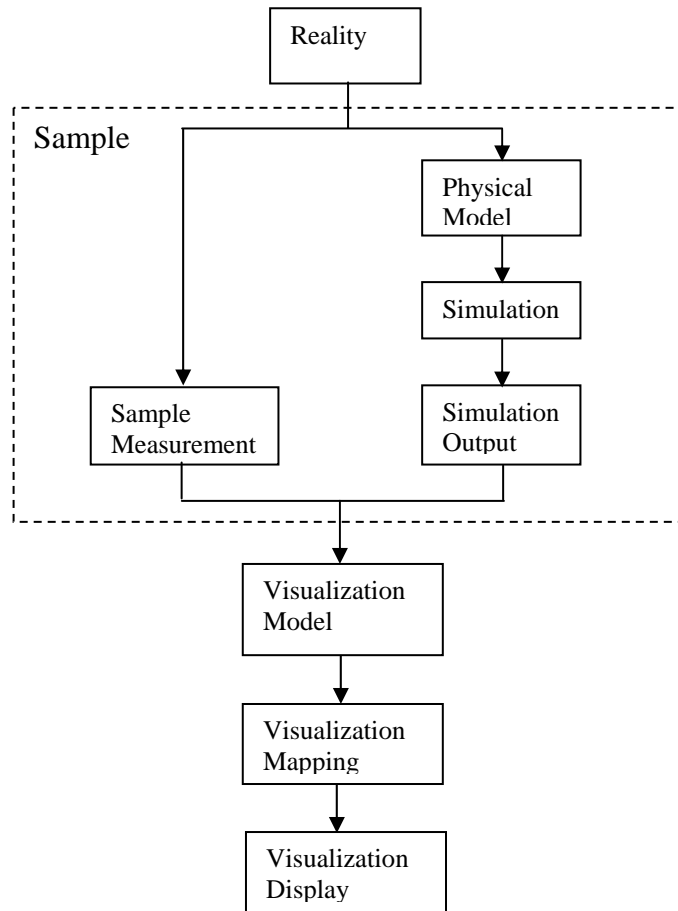


Figure 2 – Reference Model

There are 5 identifiable layers in the overall reference model.

Reality: We suppose that the purpose of visualization is to understand some physical process, which we call ‘reality’. This could be temperature in the atmosphere or it could be flow over an aircraft wing, as two possible examples. We assumed that we could represent this mathematically as a vector-valued function of many variables, defined over a domain D , say. The original E notation will largely work for this, so that for example, E^S_2 would represent a scalar function of two variables. Note that any constraints on the domain are not captured.

Sample: In this layer, we have discrete samples of data describing in some way the reality we are studying. As we can see from Figure 2, this can arise in two ways. For measured or observed data, such as temperature in atmosphere, we place equipment at points in space and capture results. For simulations (such as modelling flow over aircraft wing), there will be an initial step whereby physical laws are used to build a model which is then used as a basis for a numerical simulation. The simulation will typically be computed over a grid, and the results output at grid vertices or cell centres – giving again a set of values at discrete positions in space.

This layer was not considered at all in the ‘E’ notation. It would be nice to capture the following:

- The nature of the grid over which the samples are distributed (regular, irregular; rectangular, curvilinear; etc)
- The errors associated with sampling; note that there are errors likely in both independent and dependent variables, ie positions where measurements made as well as measurements themselves.

Visualization Model: This is the layer that was really addressed in the original 1992 E notation. It was motivated by a feeling that the classification ought to be based on the type of the reality that underlay the data, and that a key part of visualization is to empirically build an approximation to that reality. For example, in a simple graph drawing application, we are often joining points with a smooth curve in order to convey insight to the viewer. The curve is the visualization, and so it is appropriate to use this as the basis for classification, not the sampled data. Although it was not considered in the original paper, it should be noted that the sampled data is an important part of the visualization as well.

Visualization Mapping: In the 1992 work, the classification of the model was used as the basis for building equivalence classes of techniques. Thus contour plots, image views and surface plots all represent visualization models of type E^S_2 . This has been developed further in work by Groen (2004), published via the following web page: <http://www.nat.vu.nl/~pwgroen/scivis/tax.html>.

In a further development, Brodlie (1993) tried to split the E notation into two parts: the underlying field (much as the 1992 effort) and the view. The view classification was new, and was an attempt to distinguish methods from the equivalence classes established in 1992. Thus contour lines, which only show information on 1D curves within the 2D space, were distinguished from image plots, which show information everywhere.

Visualization Display: The eventual display is a set of pixels on the screen – which could be described in some form of notation as well.

One could imagine therefore an extension to the E notation, with a notation for each step in the model of Figure 2.

Group 3: Information Visualization

The role of this group was to look at the use of ontology in information visualization applications. Information visualization is less mature than scientific visualization, and there are (apparently) generic concepts and techniques. The group began by noting examples of what would be considered ‘information visualization’, as opposed to ‘scientific visualization’:

- A graph showing citation links between papers
- A file system viewer, e.g. SequoiaView
- A model of epidemic spread
- Engine maintenance and failure tracking.

An immediate role for an ontology was to provide a thesaurus, addressing issues that range from problem domain specific, through to generic visualization concerns, e.g.:

- Different names for one engine part
- Medical terminology
- Definition of what constitutes a citation

It was further felt that domain ontologies had a distinct role, for example they could be used as tools to phrase or restrict questions that were being asked as part of the visualization process. In the context of mining citation data, for example,

- Limit scope of citation model to particular subject area; this implies that “subject area” is part of the domain ontology
- Decide whether multiple references to a given paper should be viewed as distinct or as one, i.e. what constitutes a reference according to the ontology?
- Provide a vocabulary for characterizing subject areas of paper, i.e. how is the space of possible subject areas structured or denoted?

In terms of the box model, the following issues were identified:

1. *Human-Machine (input description):*

- (a) how is the visualization to be used, i.e. for exploration, confirmation, or presentation? Describing the use of a particular visualization would require a domain ontology with which to express the communicative goals that appear as part of a generic model.

- (b) what data do is available? This might need to accommodate how the data is encoded. The issue of encoding raised the question of whether the MM box might include translators, or if not, whether an ontology-driven system might be able to advise on the encodings that were available.
- (c) what visualization is required? This could be expressed in an implicit, declarative way through a description of the tasks that the representation should support (e.g. find, compare, explore), or in a more explicit, procedural form by setting out transformations to be applied to the data, e.g. layout a graph using a force-directed placement and highlight the minimal spanning tree.

It was noted that the choice of representation might be contingent on the context of use, for example “visualize the usage of my disk space by file type” could be realized by a simple pie chart, or by a tree-map. The latter provides more context and detail supporting exploration, while the former might be more appropriate if one was making a case for a larger harddisk to the head of school!

The discussion also pointed to the issue of whether capabilities of visualization services are captured explicitly in ontology, or are inferred by tools or brokers.

An ontology may require different kinds of slots, e.g. capturing issues of function, representation, use, or these different concerns might be expressed by separate ontologies that are linked, e.g. Bertin’s work (Bertin, 1983) on semiotics might underpin an ontology of graphical representation.

2. *Processing (Machine-Machine):*

- (a) it was unclear where visualization output went, i.e. how is the border between machine services and user presentation identified?
- (b) there may need to be an ontology of different kinds of visualization service, i.e. those that require manual configuration of resources, versus those that generate a presentation automatically (procedural versus declarative).

3. *Presentation (Machine-Human):*

- (a) visualizations are inherently interactive, and information visualization requires a range of sometimes domain-specific interaction tools to explore and navigate abstract spaces. Some of these, e.g. colour mapping with legend, are common with scientific visualization, others, e.g. techniques for filtering, zooming, and level of detail control, may not be. In either case, interaction must be accommodated in the ontology.
- (b) ‘Visualization’ encompasses non-visual modalities (sonification, haptics); should these be requested explicitly? The choice of a non-visual modality may place constraints on how services are configured, e.g. for haptic rendering, a simple 2D layout of a graph may be preferred over 3D.
- (c) The importance of taking account of domain constraints was noted, for example a default colour map for astronomical applications might have blue mapped to a high (hot) value, while for oceanography, blue should be mapped to low values (below sea level).
- (d) A link to domain data was also seen as necessary in recording provenance.

4. First sketches of an ontology.

The opening discussion on day two emphasised that a visualization ontology was not in itself a solution to visualization problems, and that even a minimally complete ontology was beyond the scope of the meeting. The aim was to identify the “bricks” that could later be used as part of the “wall”. Three items of work were then identified for break-out groups:

1. Top level of hierarchy
2. Nodes and links; isosurface as an example
3. E notation, express in OWL

The following sections report discussion within the groups and discussion in the plenary feedback session afterwards.

Group 1: Top level of hierarchy

Following advice on ontological engineering (Noy & McGuinness, 2001), the group began by trying to identify the concepts that were part of visualization. It was accepted that this list was both tentative and incomplete; ontology creation is an iterative activity.

Data	Parameter	Filtering/Sampling	Algorithm
Time	Coordinate System	Abstraction	System
Dimensionality	Dataset	Metaphor	Data Source
Data Type	Rendering	Device	Domain Ontology
Format	Representation	Task	Scenario
Network/pipeline	Transformation	Interaction	Static/Animation
Data Model	Process		

The gViz 3-layer model (conceptual [C], logical [L], physical [P] – see Brodlie et al, 2004), introduced by Prof. Duce in his opening talk, was felt useful as a way of capturing different levels of concern within the ontology. The group sketched out how some of the concepts listed above might be linked, giving the “ontology” in Figure 3:

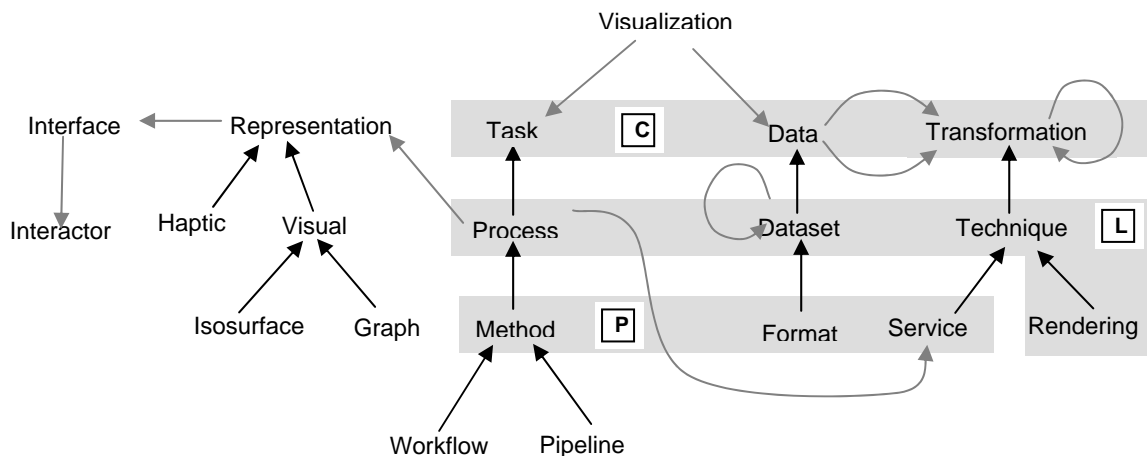


Figure 3. Top level sketch of visualization ontology.

A few comments are relevant. First, it is interesting that the ontology makes no distinction between information and scientific visualization. Second, there are some obvious structuring issues; “isosurface”, for example, arises as both a representation and as a kind of technique, and linking this passive and active view of concepts will require some care. Also, as a representation, an isosurface could be represented via different modalities, e.g. visually, and haptically. It is not clear what the relationship between these are, is it the same sense of isosurface? In terms of capturing a specific property of a dataset, the answer might be “yes”, but from the viewpoint of a user, the two representations may well have quite different affordances. Finally, one outcome of the exercise was a feeling that it was possible to see a visualization ontology as a modular structure; in the work carried out in the group, four distinct concerns were observed, as illustrated in Figure 4:

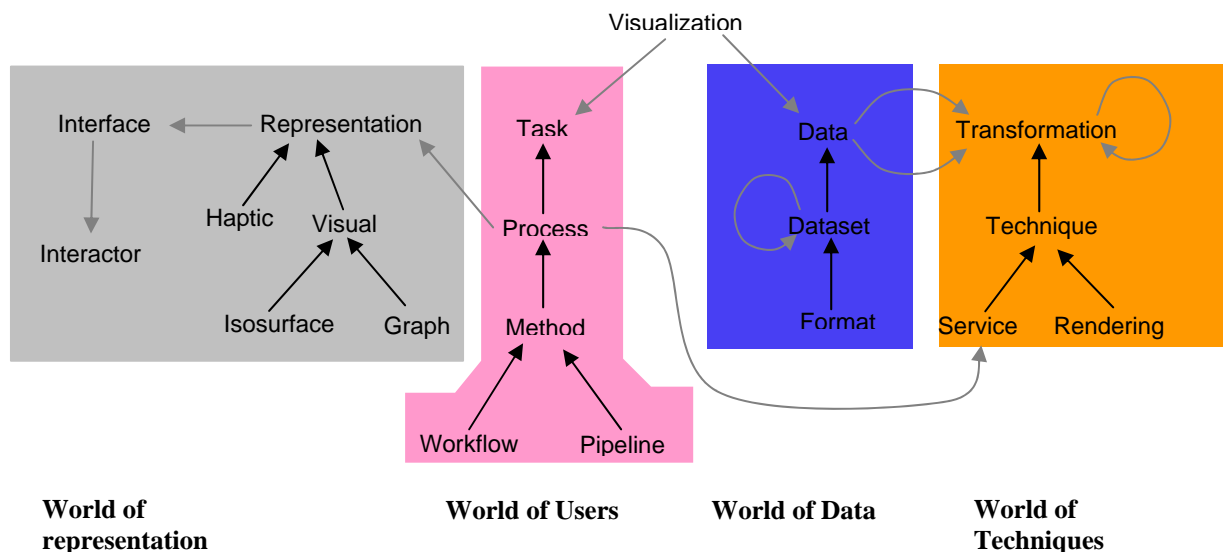


Figure 4 – Structure within the sketch; possible component ontologies.

Group 2: The E Notation

The group attempted to build an ontology in OWL, from the E notation first developed at the AGOCCG Visualization meeting in 1992 (see Brodlie, 1992, and Section 3 of this report). A start was made on this activity, using an ontology-building tool, Protégé.

In addition the group discussed whether visualization could be expressed as a set of fundamental operations. During the discussion, the group revisited some work done several years ago (possibly 1995), but never published, by David Duce, Gary Stead and Ken Brodlie. This work aimed to express visualizations in terms of a small set of operators. These are summarised here:

Operations on Independent Variable Space:

- **EMBED** – this raises the dimension of the independent variable space, as happens when a 2D scalar function is expressed as a surface plot in 3D
- **PROJECT** – this lowers the dimensionality of the independent variable space, as happens when a slice is taken through a set of data

Operation on Dependent Variable Space:

- **EXTRACT** – this restricts the domain over which a function is defined, to points at which it has a specified value, as happens in contour drawing

Operation which switches Independent and Dependent Variables:

- **INVERT** – this takes a multivariate function of one variable and creates a scalar function in a multidimensional space, such as happens when phase space plots are drawn

Visualization Operation:

- **MAKE_VISUAL** – this takes a multivariate function of 2 or 3 variables and maps selected variables to visualization aspects, such as colour, opacity, sound, etc

This work was discussed by the group, and it was found that many visualization techniques could be expressed in terms of these operations. For example, a line graph can be seen as EMBED followed by MAKE_VISUAL. While the work was clearly far from complete, it was felt to have potential and was worthy of a fresh look.

Group 3: Nodes and links

The motivation for this topic is that visualizations are commonly described in terms of a collection of transformations, processes that transform from the data domain to the visual domain. In general a visualization is described by a network of transformations that can be represented by a graph of nodes and links. The following top-level concepts were identified:

- Node
- Port
- Link (relationship between Port) [allows feedback loops]
- Parameter (may need to define synonyms)

It was difficult to find terms that do not carry undesirable connotations. The motivation for the choice of concepts was:

1. Nodes are linked to other nodes through ports. Port can be thought of as an “opportunity” for a node to communicate; to pass something to another node or to receive something from another node.
2. A link is a relationship between ports. The definition does not preclude feedback loops.
3. The group had a long discussion about parameters. The word is overloaded which does not help (hence the need to define synonyms). Are parameters different to ports? A parameter might be a colour map or the value of an isosurface. Is this different to the data that a module is transforming? Parameters may have default values. We recognised that parameter values may be supplied by other nodes. We decided that parameters are different, but we followed the approach taken in visualization systems such as AVS and IRIS Explorer in that if a parameter is to be supplied by another node, the recipient must open up a port to receive the parameter value. Otherwise the connection between a node which generates a value which is received by another node as a parameter would not be a link between the two nodes.

The concept of port was elaborated further.

- Port belongs to Node
- Port has a type
- Types of Port in a Link must match (we don't define at this level what we mean by match)
- Port has subclasses for input and output (InputPort and OutputPort are disjoint)
- InputPort has subclass Control (a Control port could receive a parameter value)
- Link is a relationship from an OutputPort and to an InputPort (thus links are directed)

The concept of a pipeline could then be defined:

- Pipeline: a set of nodes and links that form a connected digraph

It was decided that there was also a need for a concept that includes a set of nodes and links that are not necessarily connected into a single pipeline. The term NodeSet was coined for this purpose.

- NodeSet: a set of nodes and links that do not necessarily form a single connected digraph

The reason for introducing NodeSet was to allow for collections of pipelines and isolated nodes. This gives a way to talk about evolution of a visualization activity, for example, the process of constructing a visualization is a transformation between a sequence of NodeSets as nodes and links are added or deleted. It also gives a way to talk about some aspects of collaborative visualization, for example the totality of a collaborative visualization session in which participants are using private as well as shared visualizations.

The group then looked briefly at how the concept of isosurface would fit into this structure. In this part more questions were asked than answers given.

1. isosurface is a kind of Node
2. Is there some intermediate structure?
 - a. Is there a class of "iso" nodes to which isosurface belongs?

- b. Is there structure such as filter, map, render, in which isosurface would be a Map node (in filter map render), transformation node in some other scheme?
 - c. Is there multiple inheritance in the sense, say, that iso surface is an iso node and a Map node, but iso nodes are not subsumed by Map nodes? The classifications might be from different worlds, e.g. a mathematical world and a system world, so isosurface is a kind of iso in one world and a kind of map in another.
3. What is isosurface mathematically? Sampling makes the definition tricky. We did not pursue the question.
 4. How would one describe isosurface for purposes of searching? One possibility might be:
 - a. Takes as input unstructured data - could characterise in E-notation
 - b. Computes an isosurface (defined in a mathematical ontology),
 - c. Outputs geometry (e.g. as VRML - subclass)
 5. What types do we have: data, geometry? Data has substructure (E-notation describes some of this). Does geometry have substructure, if so, what? Are there nodes that only generate, say, hyperbolic geometry?
 6. How would one describe what a specific instance of isosurface takes as input and generates as output?

5. Closing Plenary

It was agreed that the first goal should be to finish what was started at this workshop.

1. For data ontology, continue the development of the ontology and look for some examples of the benefits that an ontology would give, for example through inferencing. There is a paper from the ontology group in Manchester that has some nice examples of inferencing; try to develop something like this for visualization, something that is specific enough to be presented to the wider community.
2. Look at the nodes and links work in the context of OWL-S (OWL-S, 2040). [We did not have access to the OWL-S work in the meeting.] This would be the basis for a wider activity in that area.
3. Top level hierarchy (Figures 3 and 4). Write up and explore relationships to other ontologies.

There was agreement that another workshop is needed to progress the work. It was suggested that we aim to organise a BOF session at the IEEE Visualization conference in October 2004; thus another workshop is needed before then, in September or early October.

Our thanks to the National e-Science Centre and staff for the support and help with organizing and running the meeting.

References

- D. Bergeron, 1993. Visualization Reference Models (Panel Position Statement), Proceedings of IEEE Visualization, G.M. Nielson and D. Bergeron (Eds), IEEE Computer Society Press
- J. Bertin, 1983. Semiology of Graphics: Diagrams, Networks, Maps, University of Wisconsin Press, (From the French Edition, 1967)
- K.W. Brodlie, 1992. Visualization Techniques, in Scientific Visualization - Techniques and Applications, edited by K.W. Brodlie, L.A. Carpenter, R.A. Earnshaw, J.R. Gallop, R.J. Hubbard, A.M. Mumford, C.D. Osland and P. Quarendon, Chapter 3, pp 37-86, Springer-Verlag.
- K.W. Brodlie, 1993. A Classification Scheme for Scientific Visualization, in Animation and Scientific Visualization, edited by R. A. Earnshaw and D. Watson, pp 125-140, Academic Press.
- K.W. Brodlie, D.A. Duce, J.R. Gallop, J.P.R.B. Walton and J.D. Wood, 2004. Distributed and Collaborative Visualization, Computer Graphics Forum, 23(2), pp. 185-213
- P.W. Groen, 2004. Scientific Visualization – Taxonomy, <http://www.nat.vu.nl/~pwgroen/scivis/tax.html>
- Protégé Project Home Page, 2004. <http://protege.stanford.edu/>
- F. Manola and E. Miller, RDF Primer, W3C Recommendation 10 February 2004, <http://www.w3.org/TR/rdf-primer/>

- D.L. McGuinness and F. van Harmelen, 2004. OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004, <http://www.w3.org/TR/owl-features/>
- N.F. Noy & D.L. McGuinness, 2001. Ontology Development 101: A Guide to Creating Your First Ontology, Technical Report KSL-01-05, Stanford Knowledge Systems Laboratory; see <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html>
- OWL-S (Semantic Web Services Home Page), 2004. <http://www.daml.org/services/owl-s/>
- SequoiaView Program Home Page, 2004. <http://www.win.tue.nl/sequoiaview/>
- M. Tory & T. Möller, 2004. A Model-Based Visualization Taxonomy, Submitted to IEEE Symposium on Information Visualization; see <http://www.cs.sfu.ca/~mktory/personal/publications/CMPT2002-06.pdf>

Appendix A: Attendees

Prof. Nick Avis	Cardiff University
Prof. Ken Brodlie	University of Leeds
Mr. Stuart Charters	University of Durham
Prof. David Duce	Oxford Brookes University
Dr. David Duke	University of Leeds
Dr. Ron Fowler	CCLRC-RAL
Miss Fiona Hemsley-Flint	Oxford Brookes University
Dr. Hongchen Fu	University of Southampton
Dr. Ivan Herman	W3C Offices, The Netherlands
Dr. Siu-wai Leung	University of Edinburgh
Mr. Tony Linde	University of Leicester
Prof. Malcolm Munro	University of Durham
Mr. Max Ong	University of Sheffield
Dr. Helen Purchase	University of Glasgow
Dr. Mark Riding	University of Manchester
Mr. Musbah Sagar	Oxford Brookes University
Dr. Lakshmi Sastry	CCLRC-RAL
Dr. Helen Wright	University of Hull