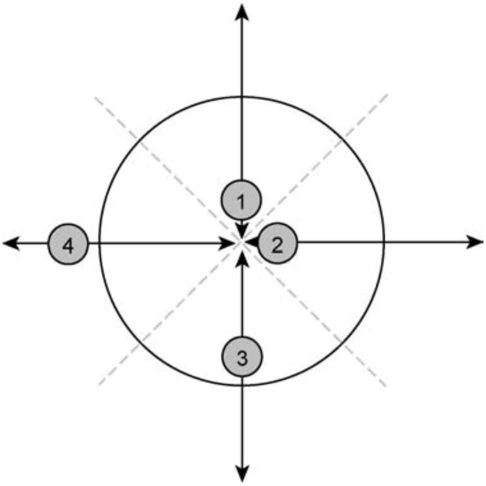
Section D
Connecting People, Data and Resources
Distributed Geovisualization



J. Dykes, A.M. MacEachren, M-J. Kraak (2005), Exploring Geovisualization, Pergamon, 732pp. 0-08-044531-4

J. Dykes, A.M. MacEachren, M-J. Kraak (2005), Exploring Geovisualization, Pergamon, 732pp. 0-08-044531-4

Exploring Geovisualization
J. Dykes, A.M. MacEachren, M.-J. Kraak (Editors)

Chapter 21

Connecting People, Data and Resources – Distributed Geovisualization

Ken Brodlie, School of Computing, University of Leeds, Leeds LS2 9JT, UK
David Fairbairn, School of Civil Engineering and GeoSciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK
Zarine Kemp, Computing Laboratory, University of Kent, Canterbury, Kent CT2 7NF, UK
Michael Schroeder, Department of Computing, City University, London EC1 0HB, UK

Keywords: geovisualization, distributed computing, computer supported cooperative work, collaborative visualization, Grid computing, environmental crisis management, ontologies, Semantic Web

Abstract

This chapter illustrates how challenging problems in geovisualization may be solved by harnessing activities on a distributed scale: bringing a range of people, a range of data and a range of resources to bear on the problem. The discussion is framed around a scenario of environmental crisis management – a flood emergency – which is typical of the challenges that only a distributed approach can solve in an effective and timely manner. We discuss some of the tools and technologies already available to meet the needs of this scenario, and some of the obstacles that remain to be overcome. We finish by reviewing the scenario against the recent Research Agenda of the ICA Commission on Visualization and Virtual Environments.

21.1 Introduction

Of growing importance in computing is the solution of previously intractable problems by aggregating a set of distributed resources. We see this in many ways: the Web has brought distributed information together; the Grid is demonstrating how distributed computational and data resources may be harnessed in an integrated way; and the use of computer supported cooperative work (CSCW) has allowed people to work together in geographically separated teams. In particular, challenging problems, interconnection of all three aspects – information, computational resource and human expertise – can yield significant benefit.

J. Dykes, A.M. MacEachren, M-J. Kraak (2005), Exploring Geovisualization, Pergamon, 732pp. 0-08-044531-4

In geovisualization itself, distribution is a key aspect in current research. In this chapter, we examine different facets of distributed working, computing and applications, and explain some of the key enabling technologies. Our approach is to pose a hypothetical scenario, and to weave a story around that scenario to show how distributed geovisualization can prove extremely useful. The scenario we choose is environmental crisis management. As the story unfolds, the notions of distributed data, resources and people are discussed in turn. We also stress the importance of conveying the message from the scientist to the decision makers. Overall, the aim is to demonstrate how current and future research in distributed geovisualization can yield real benefits to science and society – as a tool that emergency planners can employ in order to handle complex, time-critical incidents. By linking people, data and resources at very short notice, and by using visualization to provide analysis and understanding, the emergency planner can react in an informed manner. We conclude with reflection from two perspectives. Firstly, a view from Computer Science considers what visualization system designers can learn from this study; secondly, a geovisualization perspective reviews the scenario against the Research Agenda of the ICA Commission on Visualization and Virtual Environments (MacEachren and Kraak, 2001).

Selected key topics that are crucial to distributed geovisualization are explained in more detail in Key Topic boxes that should be used in conjunction with the text.

21.2 Reacting to an Environmental Crisis – Flooding

Imagine this. Several days of heavy rainfall cause rivers to rise dramatically. The town of Criss Cross lies at the confluence of Rivers Criss and Cross. Large areas of the town are threatened with flooding. How can distributed geovisualization help us prepare for and manage this crisis scenario – and help minimize future disasters?

We need to bring together resources that are widely distributed, and in a very short period. The scenario begins when the local environmental agency raises the alarm, and a disaster management group is formed to coordinate the activity and identify the key expertise that is required. One can see this group, together with the associated experts, forming a virtual organization to tackle the crisis management.

An important step is to harness scientific expertise in order to predict the likely danger areas as the flood develops. Flood modelling experts play the key role. They need to bring together distributed data, distributed computational resources (see Key Topic 1: Introduction to distributed computing) and distributed people, in order to understand the problem, predict the outcome and devise the best disaster management strategy. The emphasis initially is on exploratory visualization – using visualization as part of the analysis process, to gain an understanding of how the flood will develop. This is covered in §21.3 to 21.5, looking in turn at the three aspects: data, computation and people. Once the scientific understanding has been gained, there is a second phase in which visualization is also important: the message has to be communicated effectively to the decision makers, because difficult decisions on whether to evacuate people or not will have to be taken. This use of visualization for presentation of results to other people is covered in §21.6.

Key Topic 1:

Introduction to distributed computing

Distributed computing makes use of a network of computers to solve a task. There are two main reasons for making use of a computer network: First, through parallel processing a task can be solved faster. Second, resources such as specialized instruments, hardware or software, e.g., are not available in one place. The common characteristics of distributed systems are that they "consist of a collection of autonomous computers, connected through a network and distributed operating system software, which enables computers to coordinate their activities and to share the resources of the system, so that users perceive the system as a single, integrated computing facility" (Coulouris et al., 2001). According to Enslow (1978) a distribution can be organized along three axes:

Data. Is data replicated in the system? Is it centrally held or do all components store their data locally? Meteorological data collected at various places is an example of data, which is locally generated. It may none-the-less still be stored at a central server used for forecasting. (Note that issues of how to disseminate large quantities of weather model data are discussed by Treinish, this volume (Chapter 20)).

Control. Is the system organized in a hierarchical fashion such as a master-slave architecture or are all components autonomous? Indeed, there is a continuum of possibilities: at one extreme, a pair of walkie-talkies is an example for autonomous operation; at the other extreme, a UNIX network is dependent on its server; between these extremes, mobile phones involve some central control, to the extent that messages are stored centrally when a particular phone is switched off.

Processors. Are processors specialized or general-purpose? Are they heterogeneous or homogeneous? A network of identical PCs, all running the same version of Microsoft Windows software, would be an example of general-purpose, homogeneous processors, while a network of mobile phones of different vendors are specialized, heterogeneous processors.

The above aspects are geared to data and resources. If we take a broader perspective, then we must involve people as well. As argued in the scenario, people based at different locations with access to different infrastructures, have to be empowered to cooperate and exchange information. In doing so, they will invariably exchange data and access other resources.

21.3 Connecting Data – Data Management for Interoperability

Geovisualization is being presented with unique challenges as spatio-temporal decision-making applications require capabilities for extracting and using relevant subsets of data from heterogeneous distributed data resources. As MacEachren et al. (1999) have noted, geovisualization can be viewed as "…a process, … (involving human visual thinking, computer data manipulation and human computer interaction), in which vast quantities of georeferenced information are sifted and manipulated in search for patterns and relationships". In the context of the example emergency management scenario, the emphasis is less on extracting insights from the data using knowledge discovery techniques than on providing the user with the tools to deal with the specific task of slicing, dicing and drilling down into the data sets to extract and visualize relevant, multiple perspectives on the data. Another aspect of data management in a distributed environment is the use of semantic data or ontologies that facilitate information sharing by enabling a user community to share a set of concepts that are widely subscribed to and accepted within an application domain. Fonseca et al. (2000) highlight the use of ontologies for information integration, which is also discussed by Kemp, this volume (Chapter 25). We consider these issues in the context of the imminent crisis due to flooding sketched in the scenario earlier.

The users in this emergency management application, the river or water resources authority, the highways authority and the civic authorities will require access to relevant subsets of their operational databases. The data will consist of spatial network data describing the river network including topological information on connectivity of tributaries, direction of flow and location and description of flood defences and barriers. The flood simulation model will require parameters from the meteorological office indicating the predicted rainfall and the terrain model of the watershed to enable identification of the areas most likely to be flooded; this analysis is best presented to the users in the form of dynamic maps with estimates depicting the risk at, for example, hourly intervals. The emergency services will need access to detailed land cover information to identify the location of sensitive areas such as those with dense populations and agricultural areas where livestock may be at risk, as well as the associated road network from the highways authority so that services and resources may be located at the most appropriate places.

From the point of view of the underlying computational support, the geoinformation system needs to be enhanced in two ways. First, the data subsetting and extraction capabilities provided by the user interface must enable users to specify, flexibly and efficiently, the required subsets of the data space that are relevant to solve the particular problem. With most analyses using georeferenced data, the two main dimensions are space and time associated with the relevant thematic or scientific attributes. In the context of the flood emergency, the spatial dimension will be extremely relevant to enable users to specify the data to be extracted from the river and road networks and the information about land use that refers to the region at risk. The temporal dimension will be of relevance in the river network data to distinguish variables in the database that identify parameters for river flows depending on the season of the year. This temporal aspect is of major importance in regions of the world where the rainfall is seasonally concentrated.

The information base that supports geovisualization can be enhanced by inclusion of categories or classification hierarchies that are considered relevant to the application domain. These hierarchies can assist users to specify the appropriate level of detail that is required to be visualized and can also be used to link concepts representing geographic phenomena with the values that occur in the underlying data sets (Fairbairn et al., 2001). Consider, for example, the ontology required to describe land cover classifications. The ontology that underlies all formal land cover classifications consists of a hierarchy of descriptors associating commonly used terms to identify each type of land cover (see Key Topic 2: Ontologies and semantic information). The land cover classifications developed in different disciplines or by different organizations share this structure, but may differ in the specifics of the hierarchy. In such cases, inconsistencies have to be resolved either on the fly or by prior agreement to enable data sharing to occur.

Key Topic 2:

Ontologies and semantic information

Semantic information can take many forms: the term often represents application domain-relevant concepts used by researchers that are not directly present in the data sets. This is a relatively recent development in Computer Science, and is linked to terms such as the Semantic Web (Berners-Lee et al., 2001; Guarino, 1999), ontologies (Guarino, 1998a,b), knowledge representation, knowledge management, knowledge discovery and datamining (Miller and Han, 2001). Ontologies in the context of information systems refer to perspectives on semantic knowledge that underpin most disciplines or application domains. Ontologies refer to a structured representation of the concepts, terms and inference rules that are commonly used by a research community that does not necessarily form part of the variables in a database, but informs the data capture and analyses that are carried out using that data. Incorporating this domain knowledge into geoinformation systems enables heterogeneous data formats, schemas and the meaning of terms to be reconciled.

Ontologies can be thought of as conceptually concise bases for communicating knowledge. They represent a shared understanding of some domain, and can be explicitly represented by encoding the structure and relationships that exist between the terms within it. These encodings can be used to reason about the data as well as to facilitate interchange or exchange of information among researchers. Standards for specifying ontologies and converting them into interchange formats such as XML/RDF have been developed (Abiteboul, et al., 1999) to achieve semantic interoperability. Software tools and mechanisms for reasoning with ontologies and using them to integrate diverse data resources have also been developed. There are examples of ontologies applied to classification, system modelling, designing human–computer interaction and computer reasoning; indeed, there is recent ontology work for GIS – see for example, Fonseca et al. (2000, 2002). In each case, the shared vision is of crucial importance in an environment where people, data and resources are distributed (Paton et al., 2002).

Note that the concepts in this hierarchy are often associated with each other by *spatial containment* relationships; the terms or descriptors at a higher level of the hierarchy are at higher levels of abstraction, and correspondingly, those further down the tree structure are finer classifications of the higher-level concept. Thus, the concept of "*built-area*" could describe all types of artificial surfaces and serve, by implication, as a descriptor for the more detailed classification of areas of high population density. If the flood simulation model associates a risk factor with areas of high population then the concept could be used to identify the areas that the emergency services would wish to concentrate on. Another relevant ontology might be associated with the river network by classifying segments of the network (arcs) in terms of the volume of water that they would be able to contain and transport. While dealing with the flooding crisis, the emergency services would be able to concentrate on those sections where the danger of the river bursting its banks is highest. Similarly, a classification of the road network in terms of type and capacity of each road would enable the emergency managers to identify the routes most suitable for transporting required equipment and organizing evacuation of people in endangered areas. An obvious temporal ontology would classify the months of the year with respect to average expected rainfall to quickly identify the parameters required by the flood simulation model. Thus, generic concepts can be used to reflect users' collective understanding of the data space and to facilitate analyses to be performed on individual datasets.

21.4 Connecting Computational Resources

An important aspect of flood prediction in our scenario is simulation modelling, driven by the distributed data resources described earlier. Visualization at high enough resolution to enable accurate predictions will require computationally intensive simulation. Moreover, the computation has to be done very quickly. In fact, this is a scenario where "real-time" computing is not fast enough – the simulation needs to be performed and analysed faster than real-time if it is to be of any value at all! This will require running a large computational code on a high performance computing resource. The notion of a Computational Grid (see Key Topic 3: Grid computing) has emerged as an important concept in this situation. The idea of Grid computing is to collect a set of dispersed resources into a single utility: software such as Globus provides tools for authentication, authorization and secure file transfer between host machines on the Grid; a Globus passphrase provides single log-on to the set of resources. An introduction to the general concepts of distributed computing is given in a key topic and additional aspects of Grid computing are discussed, with particular application to datamining, by Schroeder, this volume (Chapter 24).

Visualization of Grid-based simulations is vital in order to understand the results of the simulation, and to study and explore a range of "what if" scenarios when time is limited. For example, the flood modeller will wish to consider the effects of different weather scenarios over the next few hours. Graphic tools such as IRIS Explorer (NAG, 2003) provide a visual problem-solving environment that can act as a desktop interface

Key Topic 3:

Grid computing

The Grid promises to provide transparent access to computational data and other resources across institutional boundaries (Foster and Kesselman, 1999). This integration of resources will allow researchers to tackle problems, which could not be solved previously due to limited or inaccessible computational power, storage, instruments, or human expertise.

In the late 1940s the British government came to the conclusion that two or three computers would be sufficient for the UK. Since then computing power has increased exponentially by a factor of ten every five years – and still the needs are never satisfied! An added challenge is that theoretical results suggest that single processors have their limits. How can these limits be overcome given that there have always been applications, which require more computational power than is available?

Clustering and networking are the way forward beyond these limits. In fact, all of the 500 fastest computers, judged on their ability to solve a matrix factorization problem, comprise multiple processors. An example is a machine with 9152 Pentium processors, which ranks among the top five and which achieves nearly peta-flops (1015 floating point operations per second). However, as in many other fields the Internet adds a new dimension to high-performance computing. In contrast to classical machines, efforts such as SETI@HOME and Distributed.Net go a step further by tapping into a much larger resource: they use idle cycles of some 100,000 PCs connected over the Internet. The principle is simple: Individuals, who wish to contribute, register with a server and are assigned a small part of a large problem that can be broken down into independent sub-problems. When they are finished, results are sent back to the server and new sub-problems are handed out. This is already an example of a Grid, which spans many different institutions.

However, the vision of Grid computing does not end with bundling computational power. Often the problem is that specialized instruments, large databases, human experts, or any other resource is spread across multiple institutions. The Grid aims to provide the infrastructure of protocols, services and software development kits to integrate these resources and facilitate collaboration between institutions.

A brief introduction to the Grid in general is given in Foster (2000). Further information can be found at NeSC (National e-Science Centre, 2003) and the Global Grid Forum (GGF, 2003).

to the Grid-based simulation (see Brodlie, this volume (Chapter 23)). This allows the notion of computational steering whereby the control parameters of a running simulation are changed as it proceeds, on the basis of the visualization of current results (see for example, Walkley et al., 2002).

J. Dykes, A.M. MacEachren, M-J. Kraak (2005), Exploring Geovisualization, Pergamon, 732pp. 0-08-044531-4

The interface – in terms of display and interaction technologies – available to the flood modeller can span a wide range of possibilities. At one end of the spectrum, the modeller may be located in a well-equipped laboratory, with expensive display technology such as stereo wall or CAVE (Electronic Visualization Laboratory, 2003), and connected to the simulation by a high speed network connection; at the other end of the spectrum, the modeller may be in the field and only have access to a mobile device. In an emergency situation, it is indeed very likely that the modeller will only have access to the simulation through this limited means. Some parameters change dramatically as we move across the display spectrum: for mobile devices, the resolution is very low, the network speed is low and the local processing power is likely to be more limited. Nevertheless, 3D visualization is already a possibility on a mobile device and we can expect this technology to advance significantly over the next few years. This is discussed further by Coors et al., this volume (Chapter 27).

In any distributed visualization application, however, it is likely that the amount of data transfer will push the limits of the bandwidth available. It becomes extremely important therefore to make use of data compression. The chapter by Coors et al., this volume (Chapter 27) describes a mechanism for compression of 3D geometry that achieves 98% compression.

21.5 Connecting People – Collaborative Visualization

Thus far, we have seen the need to harness distributed information and distributed computing resources. But an important further aspect is access to additional human expertise. For example, the modeller requires access to information on likely weather patterns over the coming few hours. They want to collaborate with experts on local meteorology. This collaboration should take place quickly and so it is generally infeasible to expect to collocate the modeller and the meteorologist. This collaboration can however take place over the network. Both people need to be able to visualize the results of simulations, and both require the ability to interact by controlling parameters – such as likely rainfall levels over the next few hours. Indeed, situations like the one outlined may call for a range of people to be brought together in this way. Studies in distributed cognition, where a group of people (both real humans and virtual agents) gain benefit by pooling their cognitive resources, could have a useful impact here (Artman and Garbis, 1998; Hollan et al., 2000; Zhang and Norman, 1994; MacEachren, this volume (Chapter 22)).

As outlined by Brodlie, this volume (Chapter 23), a number of prototype collaborative visualization systems have been developed. In particular, the extension of IRIS Explorer to allow collaborative visualization, and indeed collaborative computational steering, is now an integral part of the product distributed by NAG (see Wood et al., 1997, for a description of the collaborative extension; see NAG, 2003, for information about IRIS Explorer). A demonstrator of its use in a disaster scenario involving the release of a toxic chemical is described by Brodlie et al. (2002).

Collaborative visualization needs to be supported by high quality video conferencing facilities. The network of AccessGrid nodes (see AccessGrid, 2003) provides an excellent environment for group-to-group video conferencing; and the VRVS desk-top video conferencing system (VRVS, 2003) offers good, low-cost person-to-person conferencing.

Just as care over size of data transfer is important in connecting people to the computing resource, so it is an important issue in connecting people to people in the collaborative visualization scenario just described. The collaboration may span continents; the collaboration may be between people with widely varying connectivity (for a detailed discussion of strategies and uses of collaborative visualization see Brodlie, this volume (Chapter 23) and MacEachren this volume (Chapter 22)).

Thus, we have seen how each component of distributed geovisualization (distributed data, distributed computation and distributed human expertise) combine to provide an effective "geovisualization" approach to the disaster scenario sketched in §21.2. Together, they allow analysis and exploration of possible outcomes. But there remains a final step – presenting the results of the analysis – and this is the topic of §21.6.

21.6 Conveying the Message

The emphasis to this point has been on scientific exploration and input to decision support – can we predict the outcome of the flood under different scenarios and an understanding of the certainty in those predictions? However, in this scenario there is an important further role for visualization, in explaining the results of the simulation to the emergency management team, so that they can plan any evacuation strategy. The focus now turns to using graphical techniques to convey a message clearly. Once again a solution might be most effectively developed through collaboration – in this case between the modeller and the management group. In a collaborative environment the latter can ask questions to which the modeller can respond using visualization to support the dialogue (see MacEachren, this volume (Chapter 22)).

At this decision support stage, we need to extend the range of models and data that are explored. For example, in any mass movement of people, as will happen in an evacuation, transport systems are central. Thus, we can envisage linking to the flood model, a corresponding transport model that will inform evacuation decisions. For example, which roads will be blocked if the flood reaches a predicted level? The communication of route information in a clear way is an enduring problem in geovisualization and developments in transport information systems and displays will make this information increasingly accessible to those who need it in a crisis; (see Fairbairn, this volume (Chapter 26)).

Finally, the information has to be made accessible to the general public. Recent experience of floods in Europe has shown that the public in an emergency situation find it extremely useful to access public service information from mobile devices. In the UK, people in flood risk areas will be able to receive warnings by text message or e-mail in a new joint Environment Agency and Met Office scheme (Met Office, 2002). Here again, visualization is important in providing the public with a clear message and route-finding information (for a discussion of geovisualization for travel assistance on mobile devices see Coors et al., this volume (Chapter 27)). In addition, mobile computing has important special functions in this scenario. Elderly people can be contacted by pager and guided to safe places by visualization supplemented with sound to provide multi-modal instructions, perhaps generated by situation-sensitive agents. People who are suspected of being in a building which cannot be accessed because of the flood can be contacted and guided to safety, again using visualization delivered to their mobile device.

The Web now plays a critical role in the dissemination of public service information. There has already been highly successful work in presenting the output from weather models and as described by Treinish, this volume (Chapter 20). He explains the importance of data compression in being able to convey the message effectively and efficiently, and presents some novel image-based rendering approaches that avoid the need to transmit expensive 3D geometry data to the browser.

21.7 Linking Data, Concepts and People

We now consider what we have learnt from the scenario that might contribute to the future development of geovisualization systems.

The overarching effect of the data management enhancements described in §21.3 is to dynamically link the data to the visualization. There are several consequences of this. Space-time windows to be visualized can be specified according to the requirements of a particular problem and several thematic dimensions can be visualized with reference to these windows at an appropriate level of aggregation. Thus, the visualization system can link spatio-temporal dispositions in order to highlight relationships between several attributes within the problem space and answer questions such as: "where is the likelihood of flooding highest and how can we use the road network to position resources in the areas of greatest need?" In the computational steering scenarios envisaged in this chapter, the active links go further than simply between the fixed "background" data such as terrain information and the visualization. Instead, we can imagine the visualization system bringing together three distinct components. These are the flood modelling simulation and two data sources: the fixed background knowledge such as terrain data and road network information; and controllable aspects (such as height and position of sandbags) that can be used to steer the simulation. Thus, we can explore more action-oriented questions such as: "where is the best place to build a flood barrier?" One can imagine providing a user with a graphical tool to insert barriers interactively, and seeing an immediate simulation of the effect.

A very interesting challenge in this area is to be able to produce application-oriented visualizations from general purpose visualization components. It is infeasible to create *de novo* visualization software for each application; yet a visualization is only effective if it is embedded in the application domain; (see Kemp, this volume (Chapter 25)). A way forward may be to tag data with some additional semantic information to guide the user towards an appropriate visual representation. In this context, as a simple example, we could imagine tagging the flood level at a location (which would appear to a visualization system as a scalar number) with metadata describing what the number represented (in this case water level) and supporting information (such as a "danger" water level). The visualization system could use this information to apply appropriate symbolism to the flood surface. The exchange and dissemination of semantic metadata is facilitated by standards such as the extensible mark-up language XML. But when it comes to integration of data from distributed sources, there is a new problem: How can it be ensured that mark-up used by the meteorologist and the flood prediction expert have the same meaning? To guarantee a shared understanding of the data's meaning, they must use a shared ontology, or have mechanisms to connect between ontologies. To this end, the Semantic Web, an effort by the World Wide Web Consortium, has the potential to support inter-operability (see Key Topic 4: Semantic Web). It builds on XML and provides a flexible, expressive language to describe and refer to a distributed concept hierarchy. Thus, the schema of local data sources refers to globally defined concepts.

21.8 Link to Research Agenda

So far, we have presented a sample scenario that forms a framework within which distribution can be examined and key technologies can be described. Further, we have

Key Topic 4:

Semantic Web

In its early days, the World Wide Web, with its standards like the hypertext mark-up language (HTML), placed a focus on the visual presentation of text and images. There was no effort to label the text or images in a way that indicated their content, or meaning. As the Web has grown, so it has become more important to be able to process Web content automatically, and to integrate different sources of information in some sensible way. Therefore, there has emerged the concept of the Semantic Web, which extends current standards and technologies to automate information integration. The main problem the semantic Web needs to address is how to allow a machine to "understand" the content of a Web page. XML, the extensible mark-up language, has been the initial approach to address this problem. Instead of tagging text with rendering information, XML allows content providers to mark-up their text with meaningful tags. So, instead of writing < *b* > *1234* </*b*> specifying that the phone number 1234 should appear in bold face, XML can express < *phone number* > *1234* < / *phone number* > stating that 1234 is a phone number. But does this mean machines can understand the content of a Web page? Unfortunately, the answer is negative, as < *phone number* > is nothing but a string of characters to the machine, no different from < *aaaa* > or < *phone no* > or < *fon no* >. What is required is a mechanism to define global ontologies, which can be referenced where appropriate.

An ontology is a concept hierarchy, which defines a domain (see Appendix A21.3 for further detail). The relationship between concepts may be rich and therefore XML alone is not sufficient. A modelling language is required that allows one to express, for example, whether a concept subsumes another one, whether it is an aggregation of other concepts, etc. Currently, representation languages and systems such as OntoLingua, Loom, FrameLogic are defined. But to build on the success of XML, it is advantageous to use it as a basis to express the ontology. This is the approach followed by the Ontology Mark-Up Language and the Resource Description Framework Schema (RDFS). A new proposal extending RDF and RDFS is the ontology interchange language (OIL). RDF and RDFS are already in use in the library community and may become an accepted standard. A successor of OIL is DAML + OIL, jointly developed by a group of European and US-American scientists. An example of a large ontology (70,000-nodes) is SENSUS, an extension and reorganization of WordNet (Whitney and Patil, 1995).

There are also efforts in the GIS community to develop ontologies, which can act as reference points and facilitate easier system integration. The synchronization and exchange of structured georeferenced data between governmental agencies, public work departments and utility organizations is a continual problem for most national mapping agencies. The OpenGIS Consortium (OGC) has developed an XML encoding for the modelling and transfer of geographic features and display on Web servers called the Geographic Markup Language (Open GIS Consortium Inc., 2003b). Badard and Richard (2001) discuss the use of XML for exchanging updates to geographic databases between GIS maintained by diverse national agencies. Visser et al. (2001) go further and discuss how the syntactic and structural capabilities of XML-based integration can be extended to include semantic integration by making such information explicit using frame-based modelling features and description logic reasoning as encompassed in the DAML + OIL proposal. Zaslavsky et al. (2000) describe a system that achieves logical integration by creating correspondences between databases of spatial information using XML and mediator middleware.

Berners-Lee et al. (2001) provides a useful overview and additional material can be found at semanticWeb.org (Decker and Sintek, 2003), OntoWeb (Fensel and Ding, 2003) and through the World Wide Web consortium (Miller et al., 2003).

considered the overall components of distributed systems – the people, data and resources – and used these to highlight specific issues such as architecture, interoperability and collaboration. How do these relate to geovisualization?

The introductory chapter of this book has explained the importance of the recent ICA-led research agenda, summarized in “Research challenges in Geovisualization” (MacEachren and Kraak, 2001). The agenda identified both precise research questions and broader problem areas in contemporary geovisualization that need to be addressed. We now identify selected links between that agenda and our particular scenario and the broader field of distributed systems.

The agenda considers four broad themes – the (visual) representation of geospatial data; the integration of knowledge construction and geocomputation procedures with geovisualization; the design of interfaces to large, complex datasets; and the impact of human factors on the usability of geovisualization environments. Also identified are a number of crosscutting challenges that span different themes and are seen as particularly challenging; these include new display and interaction technologies, and collaborative work. A number of these themes and challenges specifically impact upon distributed geovisualization, notably on the use of very large, distributed datasets, their processing and the impact of recently emerging approaches in Grid computing. Grid computing approaches are being implemented to allow access to distributed computing resources, datasets and processing power, but they also raise challenges in each of the agenda themes; for example, the visual overlay of disparately-sourced geospatial data, visual datamining across computer networks, matching the concept of the Digital Earth as an interface to the Grid as a network, and using the Grid for collaborative group working in geovisualization. Further, the agenda considers aspects such as the dynamic interfaces needed for large dataset visualization and representation on mobile and distributed devices and for group work.

Within the flooding scenario, the effective representation of the various complex, multi-dimensional and multi-variate geospatial datasets that have been assembled is essential. Efficient representation methods are required for both central and distributed data handling.

Cartographic modelling is predicated on the skilful *abstraction* of reality, its spatial and attribute properties. How can we apply new representation techniques, such as virtual environments, sound and haptic enhancements and multiple views to the model used to manage the flooding scenario? The limited size and resolution of typical mobile devices make the use of good abstractions extremely important, so that clutter is avoided and key messages are communicated. Multi-modal techniques, using sound and haptics (such as vibrations), have great potential in mobile devices, as a means of extending the bandwidth of communication to the user, without any increase in size. In the example described here, one could well imagine successful use of vibration to alert a person to important flood warnings, and the combination of oral and visual instructions to expedite evacuation.

The agenda also addressed the importance of other *new representation tools*, including interactivity, animation, hyperlinking and dynamism. Many of the representations created to manage the flooding scenario will incorporate such functionality. Simulation of flood states in the near future, for example, will rely on animated display of dynamic data, may embed hyperlinks to remote data sources, such as meteorological and satellite remote sensing servers, and will be used by workers who demand the ability to interactively modify parameters. The physical appearance of the representations will be dependent on both the *technology* used and the *data available*. We have already indicated that the wide range of display devices in our scenario – from simple mobile phone display to advanced high resolution, multimedia, collaborative output – is likely to yield highly variable map output. Such output is the device-dependent manifestation of the preverbal message (PVM, see Coors et al., this volume (Chapter 27)). Similarly, possible

J. Dykes, A.M. MacEachren, M-J. Kraak (2005), Exploring Geovisualization, Pergamon, 732pp. 0-08-044531-4

representations will be dependent on the characteristics of the data, such as its resolution (both spatial and temporal), its accuracy and reliability, its geographical range, its age, etc. Clearly, the data available to a flood simulation system will drive its possible representations, as well as its integration, analysis and predictive output. The nature of some data may make effective representation difficult, for example, how do we represent the dynamism of a flood surge?

The research agenda concluded that, in addition, representation is *task dependent* and clearly each particular stage of the flooding scenario is likely to yield different requirements for data display: the map of emergency evacuation routes on a laptop in a police vehicle will differ significantly from the integrated display of multivariate and multi-dimensional data in the flood modelling centre, and neither will resemble the overview map broadcast on the regional television news.

The research agenda next examined areas of knowledge discovery and geocomputation, and their impact on geovisualization. It suggested that, because "geospatial data volumes are so large and the interactions among variables so complex that human vision cannot be successful in isolation" (MacEachren and Kraak, 2001, p. 6), an integration of computational and visual approaches is necessary. The primary aim is to achieve "added-value" and insight from a combination of contemporary visualization tools and advanced *exploratory data analysis*. Distributed methods can contribute, in particular, to collaborative datamining which may yield novel approaches to problem solving. In addition, such collaboration enables geographically remote data collectors, analysts and decision makers to effectively contribute to time-critical scenarios (such as the flooding scenario described earlier) and to manage the computational challenges presented by such circumstances. The research agenda makes a distinction between automated or machine-driven datamining for the purposes of *knowledge discovery*, and that initiated by human beings. The use of agents exemplifies the former, whilst the human can bring new methods to datamining, interpret results more effectively and extract localized and unanticipated patterns (Gahegan et al., 2001). In a flooding situation it can be imagined that initial warnings, reports on the current level of floodwater, and simulations of possible future progress are each the preserve of a sophisticated, but automated, monitoring system. The display of data from such a system may well be completely non-cartographic. As soon as the human manager enters the system, however, the tasks are likely to change (e.g., from recording flood levels to managing a vulnerable community) and the extraction of useful knowledge could benefit from a truly spatial display. Whether such a display is a schematic map on a paging device or a real-time aerial photo image is immaterial: each is capable of being examined, questioned, explored and analysed. Such map use is likely to be beneficial to all players in the scenario, from the evacuation orderly on the ground to the hydrological scientist in charge of water-level prediction. Wider consideration of visualization as enabling technology for the complete knowledge construction process is addressed in Gahegan et al. (2001). The control, overview and integration of all the knowledge discovery tasks for flood trend extraction, flood prediction analysis, flood scenario building and knowledge evaluation and flood simulation presentation, can be undertaken by a visualization architecture.

An examination of knowledge discovery indicates methods whereby insight can be achieved. There are still unresolved questions of how visualizations actually prompt the human mind to *creative thinking*. This issue, along with other specific questions of human–computer interface design and efficiency, were also considered by the research agenda. A number of interface technologies were addressed, ranging from distributed Web based gateways into geospatial data, through multi-sensory environments (including immersive virtual worlds), to collaborative resources, and also mobile and telephone devices. It is clear that each of these technologies can have impact on distributed and collaborative geovisualization. For example, the initial *metaphor* that governs the appearance and use of a graphical depiction of spatial data has significant impact on its subsequent information extraction capability; (see Noy, this volume (Chapter 12)). A focussed visual guide to an evacuation procedure is more likely to be effective than a free-format maze or game-like interface to geospatial data. Interfaces for collaborative geovisualization can include a significant number of innovative and valuable techniques that can build on a simple video-conferencing metaphor. Use of *3D and virtual environment* interfaces, perhaps enhanced by avatars, could present the flooding scenario to groups of research workers. The avatars, representing human beings within the visualizations, could allow the real-time location and monitoring of human resources, or alternatively could be programmed to simulate panic-stricken evacuees. In this way, the effectiveness of emergency planning and management can be ascertained, although the level of realism for such effective scenario building still needs to be determined.

A final research theme identified in the agenda that is linked to all of the preceding aspects, relates to cognition and usability. We need to ascertain whether *virtual environment interfaces* to geospatial data and *animated representations* of such data are effective. *Dynamic representations*, in which user control over the display can alter its appearance, are similarly of unproven value. Of interest to distributed system users are the differences between *individual and group use* of geovisualization. Five individual variables can be highlighted here – expertise, culture, sex, age and sensory capabilities. These varying user characteristics mean that geovisualization interfaces and tools need to be adaptable – capable of being altered by the user to meet their own preferences and abilities. Alternatively, it is possible to imagine adaptive tools which adjust "on-the-fly" to their individual user. Such users may also benefit from specific training and guidance. For group and participatory geovisualization, further issues are raised. The ways in which such *group work* is undertaken can vary: same place–same time (imagine a group charged with "just-in-time" planning gathered in one location and exposed to visualizations of the current flood situation); same place–different time (more strategic planning may involve different groups – forecasters, security officials, elected bodies – accessing the same visualizations, but in successive sessions); different place–same time (collaborative remote work among central control room and field staff); and different place–different time (on-site emergency workers retrieving a forward plan for managing the flood situation from a central simulation which was run some hours previously). The impact of variation in access to *technology* (field workers using mobile phones, central staff using "memory walls" or high resolution CAVEs) make the design of usable geovisualizations particularly difficult.

There are a host of both practical and conceptual issues raised in the research agenda that impact on the creation, delivery and use of distributed geovisualizations. At the practical data level, integration of disparate and dispersed geospatial datasets, varying in projection, scale, level of generalization and other characteristics, is problematic: research led by the Open GIS Consortium is addressing such data integration issues (Open GIS Consortium Inc., 2003a). Another active research area addresses the practical challenges of large volumes of geospatial data matched to limited bandwidth, small screen mobile devices for access to critical data in the field. The developing capabilities of such devices for delivering high resolution, multi-dimensional, colour output are crucial to the successful dissemination of geovisualizations (Gartner and Uhrliz, 2002). Of similar impact is the technology used for collaborative working in the office, where matching the view of one group in one location to that of another group elsewhere may be critical. From a conceptual viewpoint, issues raised included the growing impact of experiential technologies, such as CAVEs and game-like environments, and their delivery and effectiveness. An examination of the dynamic of group work enabled by geovisualization is also of importance. What is necessary, ultimately, is a "human-centered approach to geovisualization" (MacEachren and Kraak, 2001, p. 9) which focuses on the user- and the task-dependence of all geovisualization.

21.9 Conclusions

This chapter has highlighted how Distributed Geovisualization – connecting people, data and resources – can deliver real benefit to science and society. We have taken a real world problem of flood management, and showed that a distributed approach can lead to more effective crisis management. Of course, this is just one scenario, but it is typical of many others such as the management of forest fires, oil slicks, radiation leaks and toxic chemical release. In all of these cases, we need to harness a combined force of data, resources and people, at very short notice and on a global scale. In order to understand the science, and in order to convey that understanding to others, geovisualization is fundamental enabling technology.

Acknowledgements

The authors would like to thank the referees and the editors for their perceptive comments which have improved the final version of this chapter. We also thank Heiko Bleschschmeid for his contribution to our early discussions.

References

Abiteboul, S., Buneman, P. and Suciu, D., (1999) *Data on the Web: from Relations to Semistructured Data and XML*, Morgan Kaufmann Publishers.

Access Grid, (2003) Access Grid (www.accessgrid.org). Online: http://www.accessgrid.org (23/10/03).

Artman, H., and Garbis, C., (1998) "Situation awareness as distributed cognition", *Ninth European Conference on Cognitive Ergonomics*, ECCE-9, University of Limerick, Ireland. (August 24–26, 1998).

Badard, T., and Richard, D., (2001) "Using XML for the exchange of updating information between geographical information systems", *Computers in Environment and Urban Systems*, 25(1), 17-31.

Berners-Lee, T., Hendler, J., and Lassila, O., (2001) The Semantic Web, Scientific American. Online: www.sciam.com/article.cfm?articleID = 00048144-10D2-1C70-84A9809EC588EF21

Brodlie, K. W., Mason, S., Thompson, M.R., Walkley, M., and Wood, J.D., (2002) *Reacting to a Crisis Proceedings of first UK e-Science All Hands Conference*, Sheffield, 2002. Online: http://www.visualization.leeds.ac.uk/CovisaG (23/10/03).

Coulouris, G., Dollimore, J., and Kindberg, T., (2001) *Distributed Systems: Concepts and Design*, 3rd ed., Addison-Wesley.

Decker, S., and Sintek, M., (2003) SemanticWeb.org Online: http://www.semanticweb.org (23/10/03).

Electronic Visualization Laboratory, U.o.I.a.C. (2003) evl: electronic visualization laboratory. Online: http://evlweb.eecs.uic.edu/ (23/10/03).

Enslow, P. H., (1978) "What is a distributed system?", *Computer*, 13-21.

Fairbairn, D., Andrienko, G., Andrienko, N., Buziek, G., and Dykes, J. A., (2001) "Representation and its relationship with cartographic visualization: a research agenda", *Cartography and Geographic Information Sciences*, 28(1), 13-28.

Fensel, D., and Ding, Y., (2003) OntoWeb. Online: http://www.ontoweb.org (23/10/03).

Fonseca, F., Egenhofer, M., Davis, C., and Borges, K., (2000) "Ontologies and knowledge sharing in urban GIS", *Computer, Environment and Urban Systems*, 24(3), 232-251.

Fonseca, F., Egenhofer, M., Agouris, P., and Câmara, G., (2002) "Using ontologies for integrated geographic information systems", *Transactions in GIS*, 6(3), 231-257.

Foster, I., (2000) Internet computing and the emerging grid, *Nature Web Matters*, 7 December 2000. Online: http://www.nature.com/nature/webmatters/grid/grid.html (25/10/03).

Foster, I., and Kesselman, C., (1999) *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann Publishers.

Gahegan, M., Harrower, M., Rhyne, T.-M., and Wachowicz, M., (2001) "The integration of geographic visualization with databases, data mining, knowledge construction and geocomputation", *Cartography and Geographic Information Science*, 28(1), 29-44.

Gartner, G., and Uhrliz, S., (2002) *Maps, multimedia and the mobile Internet, Geowissenschaftliche Mitteilungen (Technische Universitat Wien)*, 60, 143-150.

GGF, (2003). Global Grid Forum. Online: http://www.gridforum.org (23/10/03).

Guarino, N., (1999) OntoSeek: Content-Based Access to the Web, *IEEE Intelligent Systems*, pp. 70–80. (May–June 1999).

Guarino, N., (1998) "Formal ontology and information systems", In: Guarino, N., (ed.), *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy.*

Amsterdam: IOS Press, pp. 3-15,
Online: http://www.oceanlaw.net/orgs/maps/ices_map.htm (June 6–8 1998).

Guarino, N., (1998) "Formal ontology and information systems", In: Guarino, N., (ed.), *Formal Ontology in Information Systems (FOIS'98)*. Amsterdam: ICES, pp. 3-15, Online: http://www.oceanlaw.net/orgs/maps/ices_map.htm

Hollan, J., Hutchins, E., and Kirsh, D., (2000) "Distributed cognition: toward a new foundation for human–computer interaction research", *ACM Transactions on Computer–Human Interaction*, 7(2), 174-196.

MacEachren, A. M., and Kraak, M. J., (2001) "Research challenges in geovisualization, cartography and geographic information science", *Special Issue on Geovisualization*, 28(1), 3-12.

MacEachren, A. M., Wachowicz, M., Edsall, R., Haug, D., and Masters, R., (1999) "Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods", *International Journal of Geographical Information Science*, 13(4), 311-334.

Met Office (2002) Meteorological Office Press Release. Online: http://www.met-office.gov.uk/corporate/pressoffice/2002/pr20020321a.html (23/10/03).

Miller, H. J., and Han, J., (2001) *Geographic Data Mining and Knowledge Discovery*. London: Taylor and Francis.

Miller, E., Swick, R., Brickley, D., McBride, B., Hendler, J., and Schreibe, G., (2003) W3C Semantic Web. Online: http://www.w3.org/2001/sw/ (23/10/03).

NAG, (2003) Online: http://www.nag.co.uk (23/10/03).

Open GIS Consortium Inc. (2003a) Open GIS Consortium, Inc. (OGC).
Online: http://www.opengis.org (23/10/03).

Open GIS Consortium Inc. (2003b) Geography Markup Language (GML 3.0).
Online: http://www.opengis.org/docs/02-023r4.pdf (23/10/03).

Paton, N., Atkinson, A., Dialani, V., Pearson, D., Storey, A., and Watson, P., (2002) Database Access and Integration Services on the Grid.
Online: http://www.cs.man.ac.uk/grid-db/papers/dbtf.pdf (23/10/03).

VRVS (2003) Virtual Rooms Videoconferencing System.
Online: http://www.vrvs.org (23/10/03).

Visser, U., Stuckenschmidt, H., Wache, H., and Vogele, T., (2001) "Using environmental information efficiently: sharing data and knowledge from heterogeneous sources", In: Rautenstrauch, C., and Patig, S., (eds.), *Environmental Information Systems in Industry and Public Administration*. Hershey, PA: Idea Group Publishing, pp. 41-73.

Walkley, M., Wood, J. D., and Brodlie, K. W., (2002) "A distributed co-operative problem solving environment", In: Sloot, P. M. A., (ed.), *Computational Science – ICCS 2002*, Springer, Berlin, Heidelburg, New York, pp. 853-861.

Whitney, R., and Patil, R., (1995) Sensus Ontosaurus.
Online: http://mozart.isi.edu:8003/sensus2/ (23/10/03).

Wood, J.D., Wright, H., and Brodlie, K. W., (1997) Collaborative visualization, *IEEE Visualization 97 Conference*. ACM Press, New York, pp. 253-259.

Zaslavsky, I., Marciano, R., Gupta, A., and Baru, C., (2000) XML-based spatial data mediation infrastructure for global interoperability, *Proceedings of the Fourth Global Spatial Data Infrastructure Conference*, Cape Town, South Africa.

Zhang, J., and Norman, D. A., (1994) "Representations in distributed cognitive tasks", *Cognitive Science*, 18, 87-122.

J. Dykes, A.M. MacEachren, M-J. Kraak (2005), Exploring Geovisualization, Pergamon, 732pp. 0-08-044531-4