From "Digital Mapping Techniques '09—Workshop Proceedings"
U.S. Geological Survey Open-File Report 2010–1335
http://pubs.usgs.gov/of/2010/1335/

Why Doesn't Your Model Pass Information to Mine?

By Jeremy Giles and Holger Kessler

British Geological Survey
Kingsley Dunham Centre
Keyworth
Nottingham, NG12 5GG
Telephone: +44(0)115 936 3220
Fax: +44(0)115 936 3552
email: *jrag@bgs.ac.uk*, *hke@bgs.ac.uk*

Abstract

For several decades geologists have been making three-dimensional (3D) models. Various proprietary and open-source software tools have been developed which allow geoscientists to produce reasonable 3D representations of the geological system that they are studying. The model they produce is quite often an 'island' of independent information. In the past, this may not have been a significant problem because there were so few models available that it was unlikely to find adjacent models forming islands in the same sea area. However, that is changing, the sea is now getting crowded with island models that can't or won't communicate with each other. The problem is compounded by research in other disciplines -- hydrologists, oceanographers, and atmospheric scientists are creating environmental models which don't take account of the geological sciences or that model them in a simplistic manner. For example, in water resource management a given area can have a 3D geological model, 3D hydrogeological model, a hydrological model, and a precipitation model. These are four models, produced by four disciplines, each using different methodologies, often based in different organisations or universities; of course none of the models pass data or information between each other. Our society needs to manage the water resources, but the models that environmental scientists are producing do not provide a coherent and consistent, single picture for the policymakers. This is becoming increasingly recognised within the European Union (EU). The European Environment Agency recently completed an inventory and recognised that "over the past few decades, a myriad of models geared to depicting, simulating and projecting environmental change have been developed and applied."[1] This is one of many preparative steps for the SEIS (Shared Environmental Information System) initiative which may lead to a future EU Directive and transposition into member countries' legislation.

The British Geological Survey (BGS), a component of the Natural Environment Research Council (NERC), launched a project in early 2009 named "Data and Applications for Environmental Modelling" (DAEM), in preparation for SEIS. The aim of the project is to enable our models to pass data and information back and forth to other models. This paper describes challenges faced by the DAEM Project.

Introduction

Over the next few years, the British Geological Survey (BGS) will focus its activities on key strategic issues related to energy and environmental change. The BGS will address complex environmental challenges requiring decisions in both the short and medium term, including carbon capture and storage, radioactive waste management, natural hazards, resource security, and environmental protection. The 2009-2014 BGS Strategy (BGS, 2009) document has at its heart a number of crosscutting projects designed to address the key

[1] http://www.eea.europa.eu/publications/technical_report_2008_11

strategic issues. One of these is Data and Applications for Environmental Models (DAEM). The stated aim of DAEM in the BGS Strategy is:

> *Development, application and operational deployment of dynamic geoscience models is at the leading edge of geoscience informatics. It requires complex and sophisticated technological development, especially in the fields of data architecture and standards, spatial informatics systems and knowledge management. This project will build on the technological advances of earlier BGS projects in the fields of data architecture, information management, digital map production, digital field data capture, geographic information and 3- dimensional modelling and visualisation, to develop a data architecture and applications environment that supports the generation of spatial and process models. We will encourage wider community involvement in their testing and application and existing international collaboration, for example in developing worldwide geoscience data and mark-up languages and exchange formats, will be taken forward to incorporate methodologies and best practice for development and use of subsurface models. To maximise their effectiveness and range of applications we will adopt a policy of making our capture and modelling software and systems available to the wider community for testing, research and educational use.*

The scientific problem that DAEM will address has been well articulated by Reitsma and Albrecht (2005, 2006). They recognised that modelling the Earth system involves numerous interacting components, each of which can be further dissected into sub-components that are studied by specialists in a wide range of scientific disciplines. The problem is compounded by the number of research groups and individuals involved in creating, managing, and sharing environmental models. Add to this the existing wide diversity of modelling approaches. Then factor in the requirement to deal with both spatial and temporal data. Furthermore, much of the knowledge about the physical systems that are modelled is held, from a computing perspective, dormant in scientific papers, modelling code, and in the heads of scientists. Finally, the lack of transdisciplinary semantics, or even explicit domain-specific semantics, reduces the ability of linked models to create real understanding.

BGS intends to put in place a framework that provides scientists with data, tools, techniques, and support to address transdisciplinary environmental questions that affect human society. We intend to achieve this by building an open community that will share data, applications, techniques, and environmental models thus enabling collaboration and achieving sustainable solutions. Clearly the BGS will not achieve such an ambitious vision on its own. Instead it intends to be part of a community, playing a leading role within that community.

To achieve these ambitious goals, a considerable number of challenges will need to be faced and overcome; these are described below.

What Do We Mean By Models?

One of the difficulties of transdisciplinary working is terminology. The word "model" means different things to differing scientific communities. Therefore it is worth defining different types of model discussed in this paper:

- Conceptual models
- Framework models
- Discrete Process models
- Linked Process models.

There is also a need to consider the relationship between data and models. A Digital Elevation Model is the result of a modelling process of the land surface. This model, in turn, can be used as input data to other models, for example a rainfall-runoff model. Care therefore has to be taken with terminology.

Conceptual Models

A conceptual model is a descriptive representation of a collection of ideas about how a system of some type functions. The process of developing a conceptual model involves gathering information of various types and developing a qualitative understanding of the physical structure or behaviour of a system. With the conceptual model in place a range of quantitative approaches can be developed to test the validity of the conceptual model and the new information can lead to its rejection or further refinement.

Framework Models

A framework model is a tool to allow scientists to integrate disparate empirical observations into a coherent whole. Such models are used to develop an understanding, in several dimensions, of information that is only partially observed. For example we frequently see three-dimensional (3D) representations of the Milky Way Galaxy. However, it is impossible to empirically observe the whole galaxy from Earth. The models are created by a mixture of observations from Earth and extrapolation from observations of other galaxies. Earth scientists use framework models to understand the geology that can only be partially observed by a range of methods. Such models capture the geologists' observations, concepts, and knowledge in a spatial framework. These may include observing outcrops, mapping topographical features, borehole logs and core, and so on. Geologists use two principal types of framework models; the geological map (on paper [*http://shop.bgs.ac.uk/bookshop/catalogue.cfm?id=2*] or GIS [*http://www.bgs.ac.uk/products/digitalmaps/digmapgb.html]*) and 3D models (*http://www.bgs.ac.uk/gsi3d/*, *http://en.wikipedia.org/wiki/GSI3D*, and *http://www.bgs.ac.uk/science/3dmodelling/zoom.html*). Figure 1 shows the differences between 2 dimensional (2D) and 3D data formats in Earth sciences.

The BGS has chosen GSI3D (Kessler and others, 2009) as the preferred geological modelling package for the production of standardised geological framework models at all scales. In simple terms, the GSI3D software utilizes a Digital Terrain Model as the model-capping surface, plus geological surface linework (maps) and down-hole borehole data to enable the geologist to construct regularly spaced intersecting cross sections by correlating boreholes and the outcrops-subcrops of units in order to produce a geological fence diagram of the area (fig. 2 A-C). Mathematical interpolation between the nodes along the sections and the limits of the units (outcrop plus subcrop) produces a solid model comprising a series of stacked triangulated objects corresponding to each of the geological units present (fig. 2 D-E). Once calculated, the block model can be analysed to solve problems as a decision support system (fig. 2 F-H).

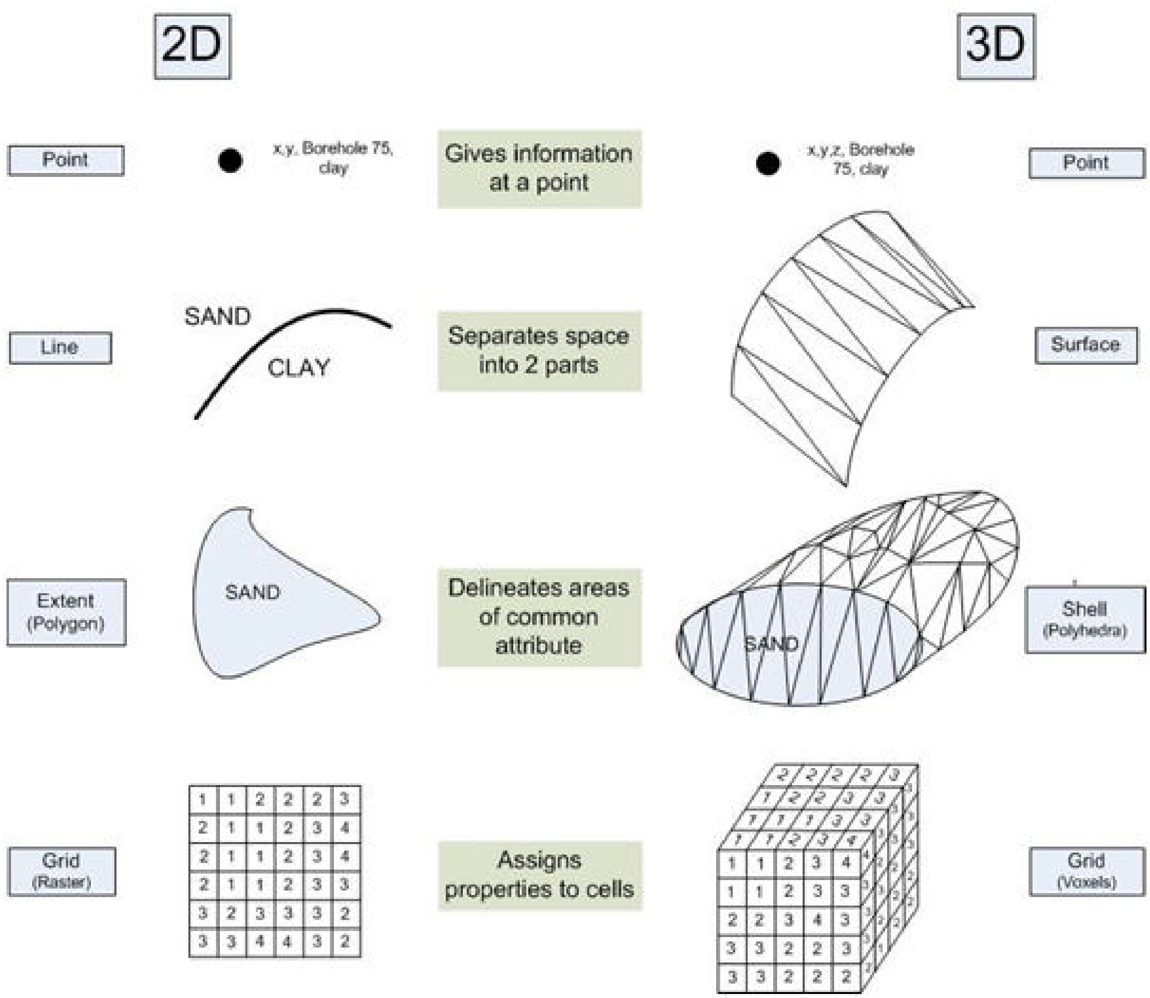

Figure 1. Data structures in 2D and 3D.

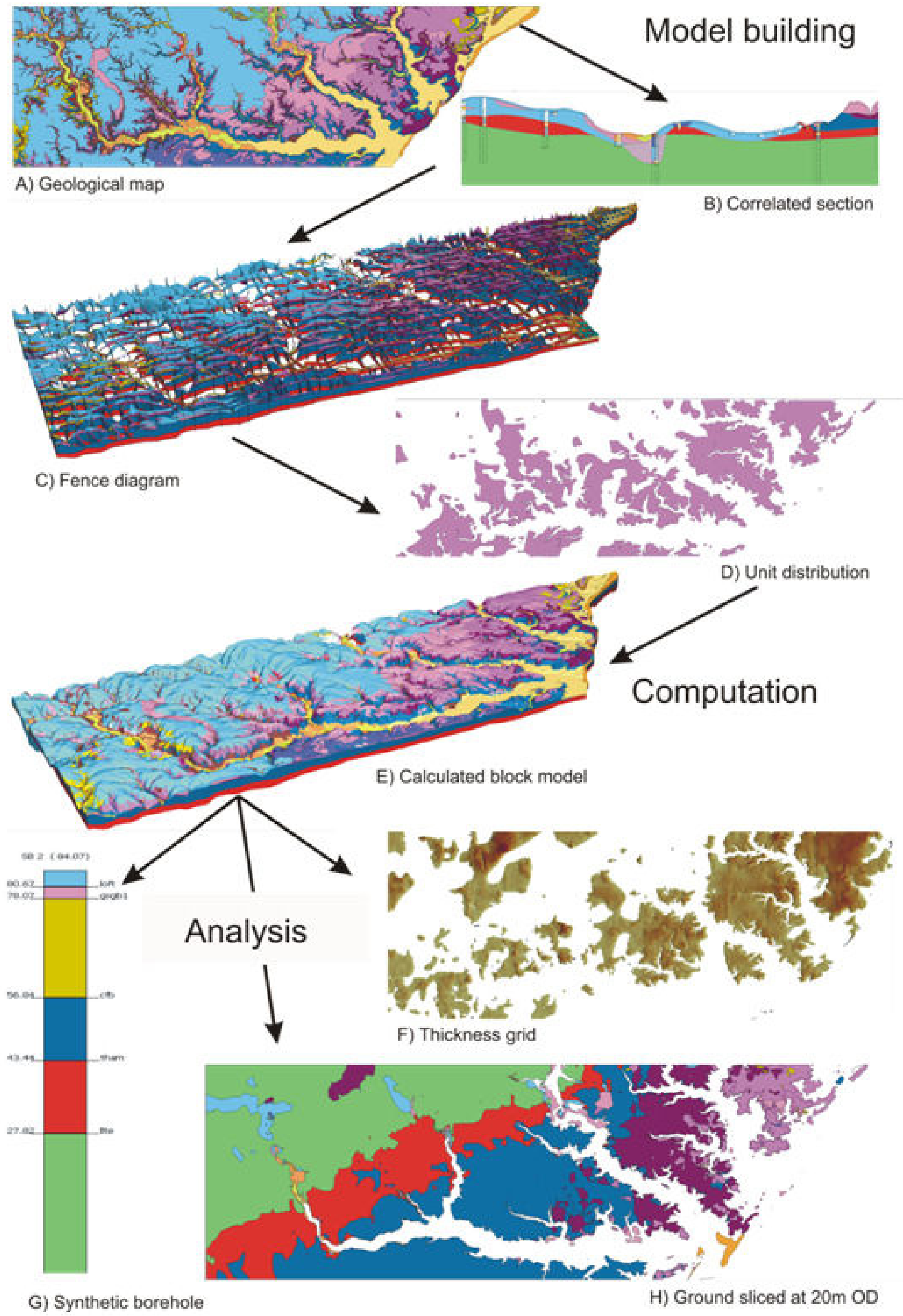

Figure 2. The GSI3D modelling workflow (from Kessler and others, 2009).

Discrete Process Models

A discrete process model simulates a particular process within the environment. For example, one of the most familiar of the Earth systems is the hydrological cycle (see figure 3). The cycle is made of a number of discrete processes that include:

- Rainfall
- Evaporation/Transpiration
- Unsaturated zone flow
- Groundwater flow.

Each of these processes can be modelled separately to gain an understanding of each element with the whole system, such as groundwater flow.

The BGS has developed groundwater models that more closely represent the structure of hydrogeological systems, producing flexible models that can both conform to aquifer geometry and simulate processes at different scales. In collaboration with the University of Birmingham and the Environment Agency, the BGS has developed the ZOOMQ3D (*http://www.bgs.ac.uk/science/3dmodelling/zoom.html*) as a discrete process model that is able to effectively model flow in a saturated groundwater system.

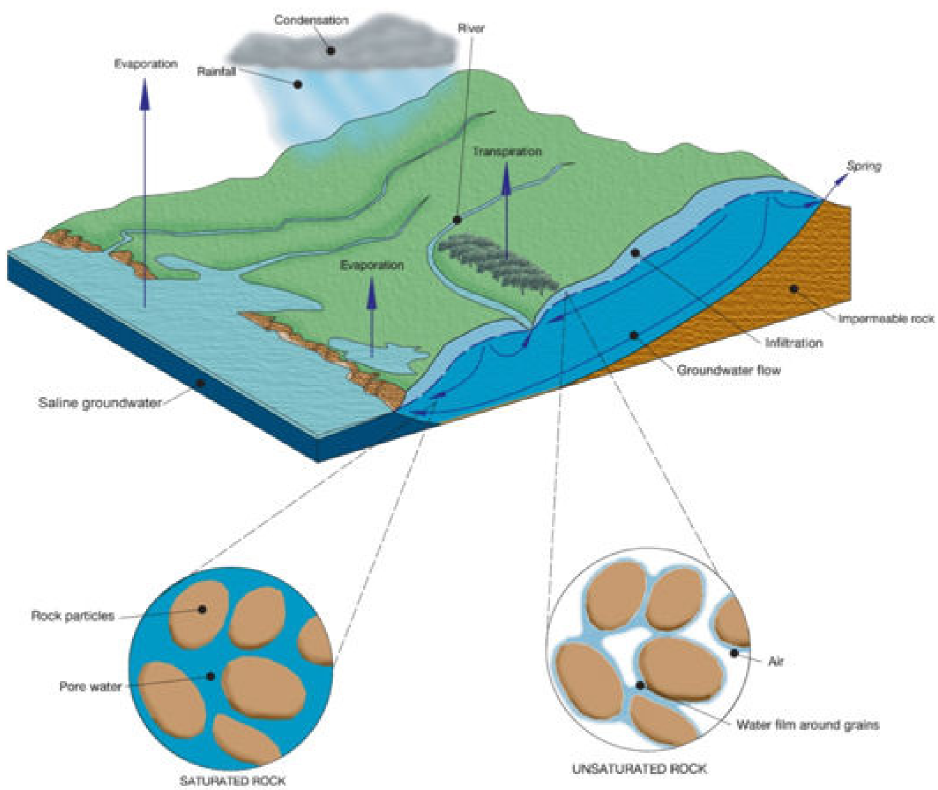

Figure 3. The Hydrological Cycle.

Linked Process Models

When a number of discrete processes have been successfully modelled, an expert can create new knowledge by taking the outputs of theses models and making an assessment of all or part of the whole system. In the case of water/groundwater, experts may make an assessment of groundwater recharge. To do this they may look at a climate model, a rainfall model, a catchment hydrological model, and a geological framework model.

Until recently it has been difficult to create a system to replicate what the expert does in the above process. The only realistic approach was to replace the existing models with a single new model that attempted to replicate the functions of the existing discrete process models. This has been a slow and expensive process that creates a further model that requires maintenance.

The alternative approach is to link two or more existing discrete models together at run-time so that they can pass parameters between each other. This effectively allows one model to query another model for a key parameter that it requires. This approach has a number of advantages:

- It is more cost effective,
- It is more agile, thereby allowing rapid development, and
- It allows the best of any existing models to be used and reused.

The Challenges the DAEM Project Faces

To achieve the vision, a range of challenges needs to be overcome. A DAEM Scoping Study Project has been established to report by end of March 2010 on the approach that will be adopted for each of these challenges in the implementation project. These challenges are:

- Software – Select the most appropriate software methodologies to achieve DAEM ambitions.
- Ontology and Semantics – The process of linking models also requires linking the concepts and classifications of those disciplines and the language used to describe them. To achieve DAEM goals requires ontological and semantic alignment.
- Scale – Environmental processes operate at scales ranging from microns to the scale of the solar system.
- Uncertainty – Understanding the uncertainties within a single model can be difficult. Understanding the uncertainties across a system of linked models represents a considerable challenge that must be addressed.
- Heterogeneity – Natural systems are heterogeneous, consisting of multiple components each of which may have considerable internal variation. Modelling Earth systems requires recognition of the inherent complexity.
- Data – Ready access to well managed data, in appropriate formats, associated with rich metadata, is essential for success.
- Intrusion – Any solution must leverage the investment in existing models rather than attempt to replace them.
- Standards – DAEM will have succeeded when its outcomes are recognised as formal international standards.
- Visualisation – Environmental models are most easily understood by their users when the output is an easy-to-interpret visualisation.
- Culture Change – DAEM must promote collaboration between researchers both within, and across, disciplines.
- Workflows – DAEM should reduce the chaotic nature of modelling of multidiscipline environmental issues and enable ordered, repeatable processes to be put in place.

These challenges are discussed in greater detail below.

Software

At the heart of the DAEM vision is the ambition to link existing environmental models together to gain a more complete understanding of the environment and the processes that occur within it. A number of systems exist that demonstrate that this is possible. For example, Caldwell and others (2009) reported a custom designed system. The work relates to the economically important Pacific salmon fisheries. The fish breed in the major rivers such as the Sacramento River of California. Competition for fresh water resources in California and climate change are affecting the survival of the juvenile fish. Their presentation entitled "An Integrated Framework for Improved Stream Temperature Predictions to Mitigate Fish Mortality" described a state-of-the-art modelling system with statistical analysis and prediction methods. The system allows a comprehensive set of decision support tools to be developed that will best guide water resource management decisions.

An alternative approach is offered by the OpenMI Association, which has produced an open standard for exchanging information between OpenMI-compliant models at run-time. The demonstration project, financed by the European Commission – Life Programme (*http://ec.europa.eu/environment/life/* and *http://www.openmi-life.org/*) is centred on the transnational Scheldt River Basin. Water management in the basin is distributed among many different authorities and operators in three countries: Belgium, France, and

Netherlands. Over recent years most of them have adopted modelling technologies to understand the hydrological/ hydrogeological system that is under their responsibility. The introduction of the European Water Framework Directive requires water management to be integrated. Existing models have been developed independently, so integration is far from straightforward. The OpenMI Standard has provided an option that enables the existing models to work together. Four use cases were defined within the Scheldt Basin, in which various aspects of model linking will be tested. By the end of the project, it is hoped that water managers will have better insights into how interactions between water systems may affect strategic decisions (Devroede and others, 2008).

Ontology and Semantics

Ontology is the branch of metaphysics that deals with the nature of being (Oxford English Dictionary) whilst semantics is the branch of linguistics concerned with meaning (OED). These two subjects are closely related. Ontologies are used to define a real world object or concept, such as a mineral. For example, how do we distinguish a feldspar from other minerals, how do we distinguish a plagioclase feldspar from all other feldspars, and how do we distinguish a labradorite from all plagioclase feldspars? Semantics enable us to exchange information and knowledge about an object or concept that exist in an ontology. In environmental science, considerable effort is put into both the study of ontology and semantics. Within a particular scientific discipline there will have been a significant history of identifying objects, defining concepts, and developing the semantics to communicate information and knowledge about them. Within a particular scientific domain the level of common agreement on both ontologies and semantics should be high enough for humans to understand each other without too much confusion. It must be remembered that human communication relies on a wealth of domain knowledge in conjunction with inference skills. Clarification is sought by iterative questioning when doubt about meaning remains. Communications between computers currently are largely transactional. Information is requested and exchanged and there are simple, automated tests to ensure that transactions were completed as anticipated. However, there is little domain knowledge held by either computer in a transaction, neither of which has any significant inference ability, to verify that the transaction was both successful and that knowledge exchanged was correct (Reitsma and others, 2009).

The DAEM vision is to link together existing environmental models to gain a more complete understanding of the environment and the processes that occur within it. Linking models together requires more than a software solution. It requires a clear understanding of both the relationships between the concepts used within a given model and the mapping of those concepts into any models that are linked to it. This requires that the BGS has a mature understanding of the ontologies and semantics that it uses and has the ability to communicate these to others both in a human readable and machine readable format. It also requires that the BGS encourage its peer organisations to adopt the same approach.

The Web Ontology Language (OWL) is a language for processing Web information. It can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. This representation of terms and their interrelationships is called an ontology. OWL is designed for use by applications that need to process the content of information instead of just presenting information to humans. It has advanced facilities for expressing meaning and semantics and representing machine-interpretable content on the Web.

Scale

The environment is affected by processes that operate from the micron scale to the solar system scale and potentially beyond. Studies of aquifers polluted by dense non-aqueous phase liquid (DNAPL) have shown that a model of the behaviour of the pollutant within the pore spaces between the grains of the sedimentary material contributes to remediation of the polluted sites (Gooddy and others, 2002; Wealthall 2002). At the other end of the scale is space weather, which requires monitoring and modelling of the state of the space environment. It requires understanding of the behaviour of energetic particles as well as the changes in electric and magnetic fields. The main interest is in conditions in near-Earth space, though space weather is important throughout the solar system. The significance of space weather lies in its potential impact on manmade technologies on Earth and in space, for example, on satellites and spacecraft, electricity power grids, pipelines, radio and telephone communications, and on geophysical exploration.

Solutions that are developed during the DAEM Implementation Project must be able to handle the range of scales that are found in nature. The strap-line "*from pore to catchment and beyond*" well describes the requirement of the hydrological cycle, whose management is so critical to the well-being of an overcrowded island like Britain. There are two challenges relating to scale:

- Developing process models in heterogeneous environments where critical parameters may be at microscales and also at kilometre scales? For example, fluid flow in a rock body may be controlled by variations in pore throat diameter, measured at the micron scale, and changes in formation lithology may be measured at the kilometre scale.

- The uneven distribution of the available data, a common problem in geology. This leads to the requirement to 'upscale' and 'downscale'

    - Upscaling is the problem of generalising from highly detailed local data to a more regional understanding.

- Downscaling is the reverse problem to upscaling, in which limited regional-scale information is leveraged to produce a more detailed local-scale understanding.

The challenge is to ensure that solutions produced by the DAEM Implementation Project take full account of the range of scales required in environmental modelling and are not restricted to only a limited scale range.

Uncertainty

All scientific models have associated uncertainties, whether such uncertainties are recognised by the modellers or not. The problem of uncertainties has long been recognised by statisticians and scientists (Chatfield, 1995).

Oreskes (2003) described the complexity paradox. As understanding increases, the natural reaction of any scientist is to add complexity to their models. In other words, as data are collected and understanding correspondingly improves, more and different processes can be added to any model. However, as more processes are added, the model requires more parameters, each with an associated uncertainty. Therefore, the overall uncertainty in the model increases. Oreskes described the paradox thus:

> *"..the attempt to make models capture the complexities of natural systems leads to a paradox: the more we strive for realism by incorporating as many as possible of the different processes and parameters that we believe to be operating in the system, the more difficult it is for us to know if our tests of the model are meaningful."*

So a more complex model better captures the nuances of the natural system, but it is more difficult to determine whether the model successfully reproduces the natural system. This has important implications for complex systems of linked models, such as those proposed for the DAEM Project. Whilst the overall system is better represented, there is an important issue as to how the modelling system can be tested against the observed response.

The uncertainties inherent in the linking of models are poorly understood and little research in the area has been undertaken. The limited numbers of models that have been linked together, to be used as predictive tools, seem to have avoided addressing the issue of the combined uncertainty.

It is the objective of the DAEM Project to link together framework and process models to produce a more complete understanding of the natural environment. Without a clear understanding of the uncertainties inherent in the combined models, the predictions they produce will have little credibility.

Research is being undertaken in this field. For example the GoCad Research Group, based at Nancy Universite in France, is becoming increasingly interested in uncertainty. Professor Caumon, Nancy Universite, recognises the success of 3D modelling and its growing importance as a major tool in natural resource management. However, it is important that modellers consider two other dimensions in their models, time and uncertainty. Geostatistical simulations have shown that one 'best' model is always limited in describing the reality and may lead to wrong predictions.

Heterogeneity

Natural systems are heterogeneous. This is often masked in small-scale models, which may be generalised. But for large-scale models there needs to be recognition of the inherent heterogeneity contained within them. The problem was articulated by Sivapalan and others (2003) in the International Association of Hydrological Sciences (IAHS) Science Plan.

Earth systems are made up of many individual processes that are related but which can vary independently. The variation may reflect natural cycles that may occur over a short time scale (for example, the season) or longer term (for example, orbital forcing and resulting climate change). Time-series data from observations of component processes within Earth systems may not fully capture the natural complexity because the duration of the observation may be inadequate. On top of this is the issue of human-induced change causing perturbations in time-series records, which increases the heterogeneity of these records.

The result of heterogeneity is to make the assessment of uncertainty more challenging.

Data

Well-managed data in the correct format with associated complete metadata are essential to the development of a comprehensive understanding of the natural environment. By well-managed we mean data that meet the eight dimensions of data management articulated by Feineman (1992). The eight dimensions are

- Accuracy
- Completeness
- Fidelity
- Lineage
- Quality
- Accessibility
- Security
- Timeliness.

These eight dimensions naturally fall into two groups. The first five dimensions reflect quality and the remainder refer to management.

Data Quality

High-quality datasets have exceptional completeness, accuracy, fidelity, and a clear lineage. The quality dimension is therefore a function of the dimensions of completeness, accuracy, fidelity, and lineage.

When users discover inaccuracies in a dataset they lose confidence in the data and in the data-management system in which it is stored. Effort should be made to ensure that the datasets are error free or that the error limits of the data are known, documented, and published.

Dataset catalogues can be frustrating when the datasets listed are missing or incomplete. For example a GIS dataset can be of limited value if it is missing its projection file. Completeness means all potentially available data are readily available on demand.

In the geosciences, many datasets are abstractions from the analogue originals. For example the majority of borehole logs are still transmitted as paper records, and a selection of the information is abstracted from the original for a specific purpose. The process of abstraction is potentially error prone. A dataset is described as having high fidelity when the digital representation of the information accurately reflects the original source.

Many datasets are processed a number of times before they are in a usable form. The history of the processing is known as the lineage of the dataset. A dataset has a good lineage when the original source of data is known, as well as details of all subsequent processes and transformations. Seismic reflection data are a good example. The original data collected in the field are processed through a number of steps to produce a dataset that can be studied by a seismic interpreter. At each stage of processing there are a number of values that can be assigned from a range of processing variables. To fully understand the dataset the interpreter may need to know the processing steps undertaken and the values assigned to the key variables. In other words, the interpreter needs to understand the entire lineage of the dataset.

Data Management

Well-managed datasets are those that are easily accessible, contain timely data, and are stored in a secure environment.

Scientists spend considerable amounts of time searching for and formatting datasets so that they are usable. Well-managed datasets are said to be accessible when the dataset is easy to locate and retrieve from a data store, they are available in the format in which it is normally used, and the intellectual property rights are clearly understood and articulated. Where the data volumes are large there must be adequate, rapidly accessible storage and high-speed access to the data store.

Such accessibility is predicated on good security. The datasets, and their related documentation, are protected from unauthorized access, inappropriate use, and partial or total loss.

Users become frustrated with datasets that do not contain the most up-to-date information. Such a dataset has poor timeliness. This is usually due to processing or inputting delays. Work-rounds are often implemented by users resulting in loss of control and multiple copies in use by the community. A timely dataset represents the current state of knowledge, or the state of knowledge at the time the data collection/synthesis is recorded and described.

Intrusion

Intrusion is an important concept in relationship to the DAEM Project. A single organisation will not succeed if it proposes an approach that assumes all other organisations will abandon their existing approaches, and the associated investments, and adopt the new approach. It would be too intrusive if the DAEM Project were to propose such an approach. The project team must respect the existing diversity of approaches.

The wonderful thing about environmental models is that there are so many of them to choose from.[2] Numerous environmental models have been produced to aid the study of various aspects of the natural environment. A study by the European Environment Agency (2008) produced a report called "Modelling Environmental Change in Europe: Towards a Model Inventory." The report looked at more than 80 models that had been recently used in environmental assessments by the European Environment Agency. This is not an exhaustive list but gives an indication of the numbers of models that exist. These models represent a major investment in time and resources to produce and maintain. Individuals and teams have considerable intellectual capital invested in the models they have created and are reluctant to abandon their work and adopt an alternative model. The DAEM Project must not start from the assumption that it will develop new environmental modelling software that will replace the existing software. Such an intrusive approach into the existing environmental modelling community must be avoided.

The challenge is to ensure that solutions produced by the DAEM Implementation Project take into account the existing range of environmental models that exist and leverage the significant investment, rather than committing considerable resources into trying to replace well-established models.

Standards

A wide range of standards is applicable to the domain of environmental modelling. DAEM should not add to these unless absolutely necessary. The DAEM vision must be to adopt and support the development of existing standards rather than create standards that rival existing ones. Where new standards are required, these should be rapidly progressed

---

[2] "*The wonderful thing about standards is that there are so many of them to choose from.*" Misquoting a statement attributed to Rear Admiral Grace Hopper

through to national and international standards. The adoption of this approach will reduce the potential conflict within the community and will reduce the risk of having to reengineer systems at some later date when one standard becomes dominant.

Visualisation

Environmental models are most easily understood by end-users when the output is an easy-to-interpret visualisation. To be successful in improving the understanding of environmental science and to provide knowledge to decisionmakers and policymakers, it is essential that DAEM outputs have clear visual interfaces that are simple to use.

An example of such a system is WaterSim (*http://watersim.asu.edu*). This is an Internet-based simulation of water supply and demand for the Phoenix metropolitan area that integrates information about climate, land use, population growth, and water policy. Adjustable settings allow the user to gage future water-supply conditions in response to climate change, drought, population growth, and technological innovation, as well as policy decisions about the nature of the region's built environment, landscaping practices, and recycled water. The systems and the science behind them still need well-written documentation at a range of levels from executive summaries to detailed user guides written for the non-specialist. WaterSim, for example, has extensive online documentation including:

- WaterSim Tutorial
- WaterSim Examples
- Teacher's Guide to WaterSim
- Students Handout for WaterSim.

It is clear that we need to learn lessons from existing environmental courseware about communicating science in an easily understandable way. Another example is the 'Carbon labs' in The Habitable Planet (*http://www.learner.org/courses/envsci/index.html*).

Culture Change

Individuals, small groups of researchers, or open communities develop and use environmental models. The majority of models are used by the individuals and research groups that develop them. Internationally recognized models such as MODFLOW (*http://www.modflow.com/*, a USGS-developed tool used by hydrogeologists to simulate the flow of groundwater through aquifers) are the exception. Few of the environmental models that are produced are designed to work with other environmental models. The majority are stand-alone systems that provide only a partial and incomplete picture of the environment. A study by Barkwith (2010) identified over 120 models in use within NERC.

The plethora of environmental models makes it difficult for non-specialists and for decisionmakers and policymakers to choose the appropriate models and to have confidence in the model results.

For DAEM to be successful there will need to be considerable collaboration, and promoting this change is one of the principal challenges for the project. It will require influencing research funders to promote collaboration in grant application and to recognize the important of transdisciplinary research. Communities that use large instruments, such as astronomers and high-energy physicists, have developed means of collaboration that recognize individual contribution whilst promoting collaboration.

Solution Workflows

Tackling multidiscipline environmental questions requires individuals or groups from each discipline to contribute information from their area of expertise. When all of the information is combined in the correct sequence, the resulting workflow contributes to the solution.

In practice the exchange of information is at times chaotic, often manual, time consuming, and poorly documented. It is difficult to reliably automate or audit such information flows without having agreed standards in place.

To produce a range of answers based upon a variety of scenarios often requires a significant amount of manual reprocessing. Each time a new scenario is modelled, there is a danger that the steps taken are inconsistent with previous model runs, leading to solutions or answers that cannot be reliably compared.

DAEM should encourage project leaders to consider up front not only which subject experts, data sources, and systems are required to provide an answer but also how information should be exchanged, and in which formats, and to formally document this in a workflow. Ideally the way a workflow is documented actually controls how system interfaces are defined.

Conclusion

BGS intends to put in place a framework that provides scientists with data, tools, techniques, and support to address transdisciplinary environmental questions impacting on human society. To achieve this goal, the DAEM Project was established as part of the BGS Strategy (2009). A scoping study was set up in 2009 to report early in 2010 on the challenges that have to be addressed and on the approach to be adopted. These challenges have been described above. We are confident that a suitable approach to addressing these challenges can be found. However, many of these challenges will only be solved by geological survey organisations and other environmental agencies working together to solve them. It is the aim of the BGS to create an open community that will share data, applications, techniques, and environmental models, thus enabling collaboration and achieving sustainable solutions. Clearly the BGS will not achieve such an ambitious vision on its own. Instead it intends to be part of a community and play a leading role within that community.

References

Barkwith, A., 2010, DAEM Model Inventory: British Geological Survey Open Report OR/10/010, 92 p.

British Geological Survey, 2009, British Geological Survey Strategy 2009-2014 -- Applied Geosciences for our Changing Earth: British Geological Survey document.

Caldwell, R.J., Danner, E., Pike, A., Rajagopalan, B., Melton, F.S., Lindley, S., and Nemani, R.R., 2009, An integrated framework for improved stream temperature predictions to mitigate fish mortality: Eos, Transactions, American Geophysical Union, v. 90, no. 52, Fall Meeting Supplement, abstract IN34A-03.

Chatfield, C., 1995, Model uncertainty data mining and statistical inference: Journal of the Royal Statistical Society, series A, part 3, no. 158, p. 419-466.

Devroede, N., Ronse, Y., Van Assel, J., and Vereeckenc, H., 2008, Demonstration of integrated modelling using the OpenMI in the Scheldt River Basin, in Sànchez-Marrè, M., Béjar, J., Comas, J., Rizzoli, A., and Guariso, G. (eds.) Proceeding of the 4th Biennial Meeting of the International Environmental Modelling and Software Society, Barcelona.

European Environment Agency, 2008, Modelling environmental change in Europe: towards a model inventory: Copenhagen European Environment Agency Technical Report 11/2008, ISBN 978-92-9167-983-6, doi: 10.2800/13800.

Feineman, D.R., 1992, Data management yesterday, today and tomorrow: Presentation to the PETEX ’92 Conference, London, November 25, 1992.

Gooddy, D.C., Bloomfield, J.P., Harold, G., and Leharne, S.A., 2002, Towards a better understanding of tetrachloroethene entry pressure in the matrix of Permo-Triassic sandstones: Journal of Contamination Hydrology, v. 59, no. 3-4, p. 247-265.

Kessler, H., Mathers, S.J., and Sobisch, H.G., 2009, The capture and dissemination of integrated 3D geospatial knowledge at the British Geological Survey using GSI3D software and methodology: Computers & Geosciences, v. 35, p. 1311–1321.

Oreskes, N., 2003, The role of quantitative models in science, in Canham, C.D., Cole, J.L., and Lauenroth, W.K. (eds.), Models in Ecosystem Science: Princeton, NJ, Princeton University Press, p. 13-31.

Reitsma, F., and Albrecht, J., 2005, Modelling with the Semantic Web in the geosciences: IEEE Intelligent Systems v. 20, no. 2, p. 86-88, doi:10.1109/MIS.2005.32.

Reitsma, F., and Albrecht, J., 2006, Modelling with the Semantic Web in the geosciences: School of Geosciences University of Edinburgh, online papers archived by the Institute of Geography.

Reitsma, F., Laxton, J., Ballard, S., Kuhn, W., and Abdelmoty, A., 2009, Semantics, ontologies and eScience for the geosciences: Computers & Geosciences: v. 35 no. 4, p. 706-709, doi: 10.1016/j.cageo.2008.03.014.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O’Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., and Zehe E., 2003, IAHS Decade on Predictions in Ungauged Basins (PUB) 2003-2012: Shaping an exciting future for the hydrological sciences: Hydrological Sciences, v. 48, no. 6, p. 857-880.

Wealthall, G.P., 2002. Predicting DNAPL source zones in fractured rock: Ph.D. thesis, Dept of Civil and Structural Engineering, University of Sheffield, UK.