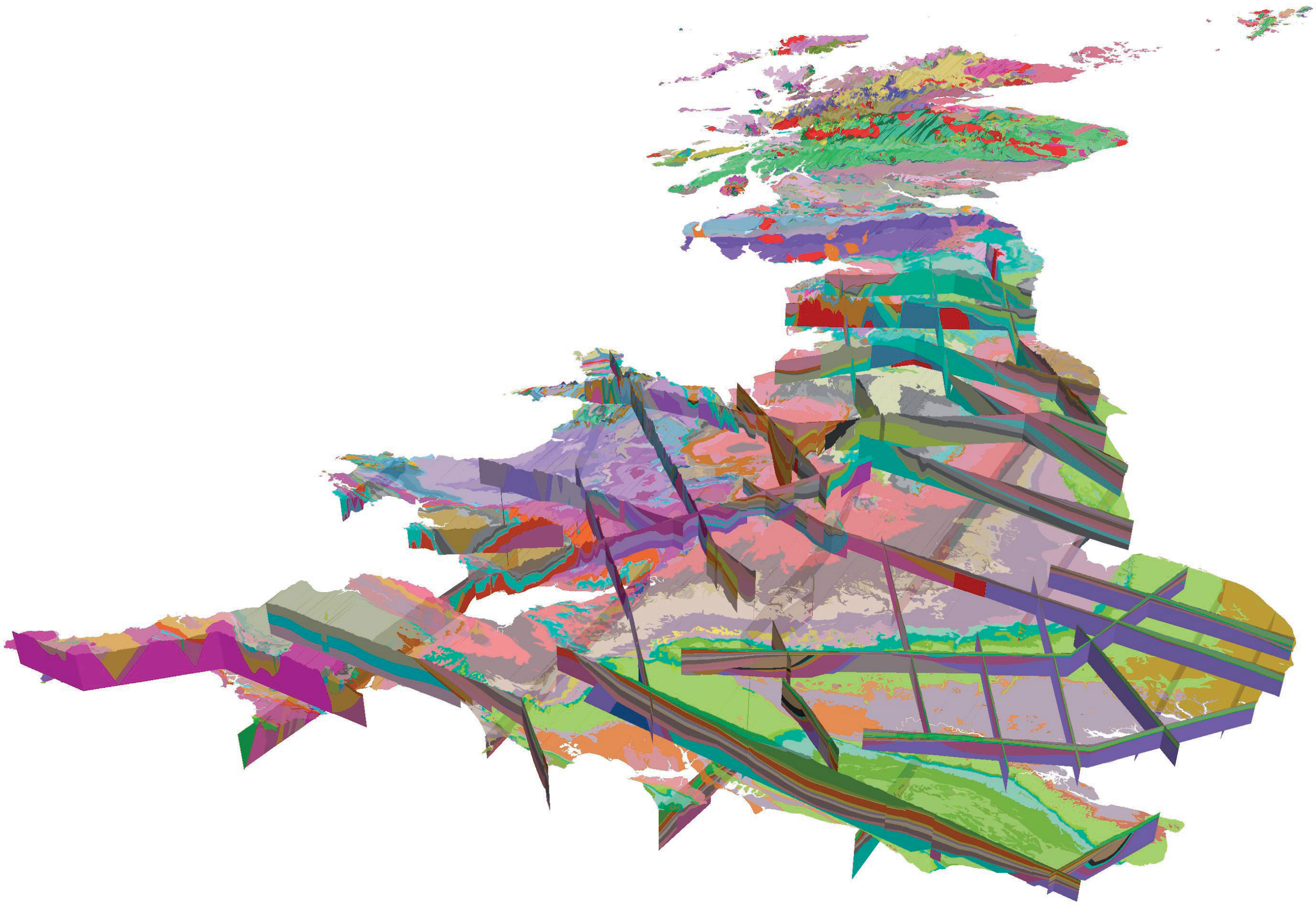
British Geological Survey

NATURAL ENVIRONMENT RESEARCH COUNCIL

Data and research for environmental applications and models (DREAM): scoping study report



BRITISH GEOLOGICAL SURVEY

INFORMATION MANAGEMENT PROGRAMME

OPEN REPORT OR/10/020

Data and research for environmental applications and models (DREAM): scoping study report

JRA Giles, A Hughes, H Kessler, C Watson and D Peach

Contributor/editor

G Baker, T Barzey-Francois, J Laxton, R Moore, M Sen, G Wealthall

Keywords
Data, Applications, Environmental, Modelling.

Bibliographical reference
Giles, J R A; Hughes, A; Kessler, H; Watson, C and Peach, D. 2010. Data and research for environmental applications and models (DREAM): scoping study report. British Geological Survey Open Report, OR/10/020. 54pp.

Keyworth, Nottingham British Geological Survey 2010

BRITISH GEOLOGICAL SURVEY

The full range of our publications is available from BGS shops at Nottingham, Edinburgh, London and Cardiff (Welsh publications only) see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications, including maps, for consultation.

We publish an annual catalogue of our maps and other publications; this catalogue is available online or from any of the BGS shops.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as basic research projects. It also undertakes programmes of technical aid in geology in developing countries.

The British Geological Survey is a component body of the Natural Environment Research Council.

British Geological Survey offices

BGS Central Enquiries Desk
Tel 0115 936 3143 Fax 0115 936 3276
email enquiries@bgs.ac.uk

Kingsley Dunham Centre, Keyworth, Nottingham NG12 5GG
Tel 0115 936 3241 Fax 0115 936 3488
email sales@bgs.ac.uk

Murchison House, West Mains Road, Edinburgh EH9 3LA
Tel 0131 667 1000 Fax 0131 668 2683
email scotsales@bgs.ac.uk

Natural History Museum, Cromwell Road, London SW7 5BD
Tel 020 7589 4090 Fax 020 7584 8270
Tel 020 7942 5344/45 email bgslondon@bgs.ac.uk

Columbus House, Greenmeadow Springs, Tongwynlais, Cardiff CF15 7NE
Tel 029 2052 1962 Fax 029 2052 1963

Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB
Tel 01491 838800 Fax 01491 692345

Geological Survey of Northern Ireland, Colby House, Stranmillis Court, Belfast BT9 5BF
Tel 028 9038 8462 Fax 028 9038 8461
www.bgs.ac.uk/gsni/

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon SN2 1EU
Tel 01793 411500 Fax 01793 411501
www.nerc.ac.uk

Website www.bgs.ac.uk
Shop online at www.geologyshop.com

Executive Summary

Many scientific disciplines have been modelling during the past 5 to 10 years in order to best understand and analyse the processes and conditions within their areas of interest. This has led to a multitude of discipline specific models, modelling system software and workflows with greater or lesser success depending upon the quantity and sources of data and complexity within the scientific discipline concerned.

There is now a growing realisation that to answer the most pertinent questions of the age such as climate change, sustainable and natural resources we need to model whole Earth system science, bringing together climate, ecological, hydrological, hydrogeological, geological and socio-economic models to name but a few in order to provide the necessary framework in which decisions upon prediction and planning can be most appropriately undertaken.

This has become most apparent within the British Geological Survey (BGS) from the wide variety of differing geoscience models generated in the past few years that need to be interlinked to fully understand the subsurface. To this end the 'Data and research for environmental applications and models' (DREAM) scoping study was commissioned to assess the current situation and make some preliminary recommendations in order to make steps towards a more joined up and semantically harmonized future in environmental modelling.

Vision: Our vision is to provide people access to data, tools, techniques and support to address trans-disciplinary environmental questions impacting on human society. We hope to achieve this by being a leading member of an open community that will share data, applications and environmental models thus enabling collaboration and achieving sustainable solutions.

The investment and knowledge captured within the many existing scientific models is a significant resource and not one that could be easily replicated in any new centralised environmental modelling software. The intrusion upon existing legacy modelling workflows and knowledge held for many collaborative partners would be too much to bear. Considering these acute disadvantages of centralisation, the alternative approach of 'linked models' passing parameters at runtime is seen as more pragmatic, achievable and cost-effective solution. This solution brings together the best and most appropriate scientific models and allows the various scientific disciplines to continue development of their current models as their knowledge is enhanced.

Linkage of models has been discussed and considered by many to be the most appropriate answer and the most mature solution currently developed is the European Union (EU) supported Open Model Interchange (OPENMI). With critical underpinning activities such as data management, semantics, vocabularies and ontology's, understanding of linked model uncertainty and visualisation, OPENMI presents an opportunity to address the present disparate nature of scientific models and move forward in understanding the whole Earth.

Contents

Executive Summary iii

Contents v

1 Introduction 1

1.1 Purpose 2
1.2 Intended audience 2
1.3 Principle assumptions 2

2 What do we mean by models? 3

2.1 Conceptual models 3
2.2 Framework models 3
2.3 Discrete process models 4
2.4 Linked process models 6

3 Science case for DREAM 7

3.1 Drivers 7
3.2 Benefits and risks 8

4 The challenges the DREAM project faces 9

4.1 Software 9
4.2 Ontology and semantics 10
4.3 Scale 11
4.4 Uncertainty 12
4.5 Heterogeneity 13
4.6 Data 13
4.7 Intrusion 15
4.8 Standards 15
4.9 Visualisation 16
4.10 Culture change 16
4.11 Solution workflows 17

5 The approaches we will adopt to deal with the challenges 18

5.1 Software approach 18
5.2 Ontology and semantics approaches 19
5.3 Scale, uncertainty and heterogeneity approaches 21
5.4 Data approach 21
5.5 Intrusion approach 22
5.6 Standards approach 23
5.7 Visualisation approach 23
5.8 Culture 24
5.9 Solution workflows 25

6 What will the project deliver? 26

6.1 Year 1 deliverables 26
6.2 Year 2 & 3 deliverables 28
6.3 Year 4+ deliverables 29

7 Critical success factors—how do we know if we have succeeded? 30

8 References 31

Appendix 1: Project mandate 33

Appendix 2: Concept note 34

Appendix 3: Glossary of terms 37

Appendix 4: Abbreviations 40

Appendix 5: Description of OpenMI 41

Appendix 6: Review of existing standards 46

1 Introduction

Over the next five years the British Geological Survey will focus our activities on key strategic issues related to energy and environmental change. We will address complex environmental challenges requiring policy decisions and actions in both the short- and medium-term, including carbon capture and storage, radioactive waste management, natural hazards, resource security and environmental protection.

'BGS will develop a more holistic focus on modelling and the prediction of environmental; change and its impacts'

BGS will play a major role in the delivery of the Natural Environment Research Council (NERC) strategy—'Next Generation Science for Planet Earth' and the 'Living with Environmental Change' (LWEC) programme. Through the activities of surveying, monitoring and research, in collaboration with the national and international community, the BGS will develop a more holistic focus on modelling and the prediction of environmental change and its impacts. Currently many scientists adopt a reductionist approach in modelling the environment. A well constrained model is developed to gain an understanding of a single component, or a small number of components, of an Earth system within a single scientific discipline. As a result numerous environmental models exist that operate independently of each other. This approach does not lead to a comprehensive model of the Earth systems or the wider environments. As part of the DREAM project a model inventory for NERC was created (Barkwith 2010), which identified over 120 independent environmental models operating within NERC. Clearly this represents a considerable resource and opportunity

Our vision is to provide people access to data, tools, techniques and support to address trans-disciplinary environmental questions impacting on human society. We intend to achieve this by being a leading member of an open community that will share data, applications and environmental models thus enabling collaboration and achieving sustainable solutions.

to merge some of these models to gain a more holistic understanding of environmental systems.

A key element of this strategy is to put in place a framework that provides people with data, tools, techniques and support to address trans-disciplinary environmental questions impacting on human society. The BGS aims 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.

1.1 Purpose

This is the final report of the Data and Applications for Environmental Modelling[1] Scoping Study lead by Jeremy Giles during 2009-2010. The objective was to take the Project Mandate and develop the ideas that it contained to create a Project Brief. The DREAM project mandate was published in the BGS Strategy 2009–2014 and forms Appendix 1: Project Mandate.

1.2 Intended audience

The intended audience of the scoping study is primarily NERC staff and senior management. It is intended to provide the BGS Senior Staff with sufficient information to set the direction and identify the resources for the project during the period 2010–2014. It will manage the expectations of the BGS focused interdisciplinary and collaborative projects:

- Battlegrounds of environmental change–cities, catchments and coasts;
- Strategically valuable rock formations for secure and sustainable resources; and
- Sea-bed geology for sustainable marine management.

It will inform senior staff across NERC and the community of scientists it supports of the planned project and opportunities for collaboration. It will provide staff with an information resource that will enable them to rapidly develop follow-on projects, collaborative projects and grant applications.

1.3 Principle assumptions

It is assumed that:

- An existing framework of technologies exists to achieve the project goals and that our requirements will be met by their adoption and implementation.
- A solution will be found to the challenge of estimating the scientific and technological uncertainties associated with linked models and that the resulting models will then be recognised as having veracity.
- A community of practice exists or can be formed, in which the BGS may play a leading role.
- The community of practise will be able to articulate its case in such a way as to attract sufficient funding to sustain the required activities.
- The required skill base will be found within, or available to, the community of practice.

[1] Throughout this report, DREAM is used to represent the project that produced this scoping study whereas the term Environmental Modelling Platform is the name of the system that will evolve to enable modelling.

2 What do we mean by models?


©Getty Images

One of the difficulties of trans-disciplinary 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 report:

- Conceptual model
- 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 (DEM) is the result of a land surface modelling process. 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.

2.1 Conceptual models

A conceptual model is essentially a descriptive representation of an idea or collection of ideas. A model may, for instance, represent a single thing (e.g. the 3D distribution of the Sherwood Sandstone Group in the Manchester area), whole classes of things (e.g. the electron), and even very vast domains of subject matter such as the physical universe. The variety and scope of conceptual models is due to the variety of purposes had by the people using them. 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 the system. With the conceptual model in place, a range of quantitative approaches can be developed to test the validity of the conceptual model, with the new information leading to its rejection or further refinement.

2.2 Framework models

A framework model is a tool that allows 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. Geologists use framework models to understand the geology that can only be partially observed by a range of methods. They capture the geologists' observations, concepts and knowledge in a spatial framework. Observations may include rock exposures, mapping topographical features, borehole logs and core, etc. Geologists use two principle types of framework models; the geological map[2 3] (on paper or Geographic Information System) and 3D models[4]. Figure 1 shows the differences between 2 dimensional (2D) and 3D data formats in Earth sciences.

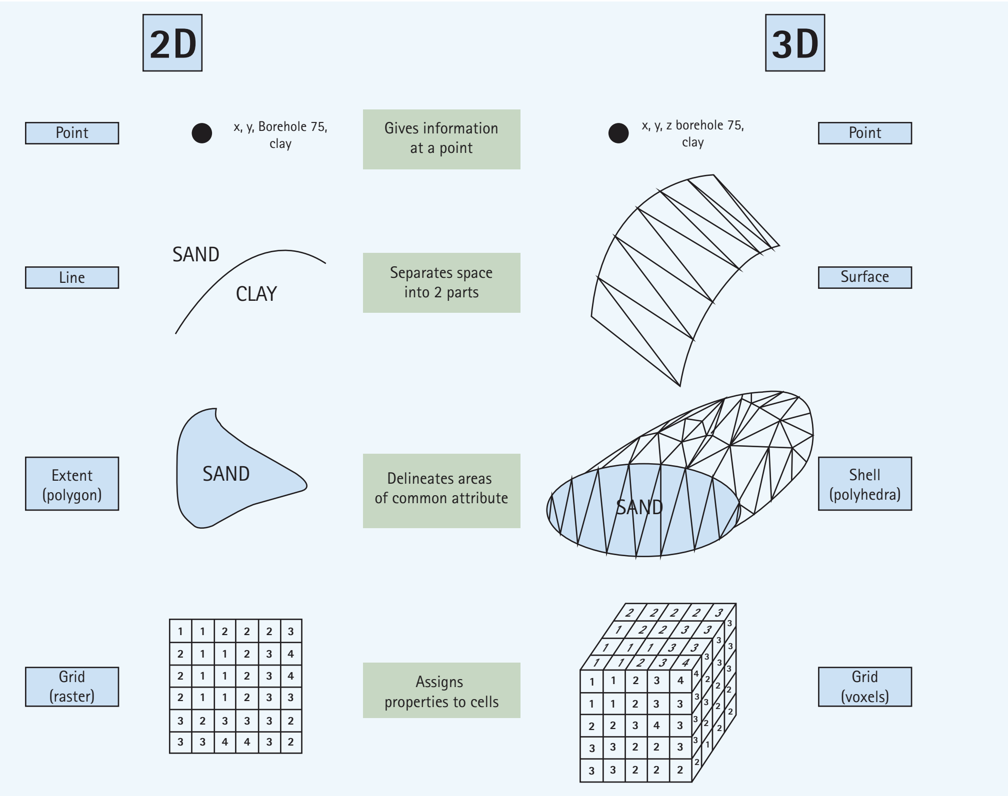


Figure 1: Data structures in 2D and 3D.

The BGS have chosen Geological Surveying and Investigation 3D (GSI3D)[5 6] as the preferred geological modelling package for the production of standardised geological framework models at all scales (Kessler et al 2009). In simple terms, the GSI3D software utilizes a Digital Terrain Model (DTM) as the model capping surface, plus geological surface line-work (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 to produce a geological fence diagram of the area (Figure 2 A-C). Mathematical interpolation between the nodes along the sections and the limits of the units (outcrop plus subcrop) produces a solid model comprised of a series of stacked triangulated objects corresponding to each of the geological units present (Figure 2 D-E). Once calculated the block model can be analysed to solve problems as a decision support system (Figure 2 F-H).

2.3 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 which include:

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

[2] http://shop.bgs.ac.uk/bookshop/catalogue.cfm?id=2
[3] http://www.bgs.ac.uk/products/digitalmaps/digmapgb.html
[4] http://www.bgs.ac.uk/science/3dmodelling/mapstomodels.html
[5] http://en.wikipedia.org/wiki/GSI3D
[6] http://www.gsi3d.org.uk

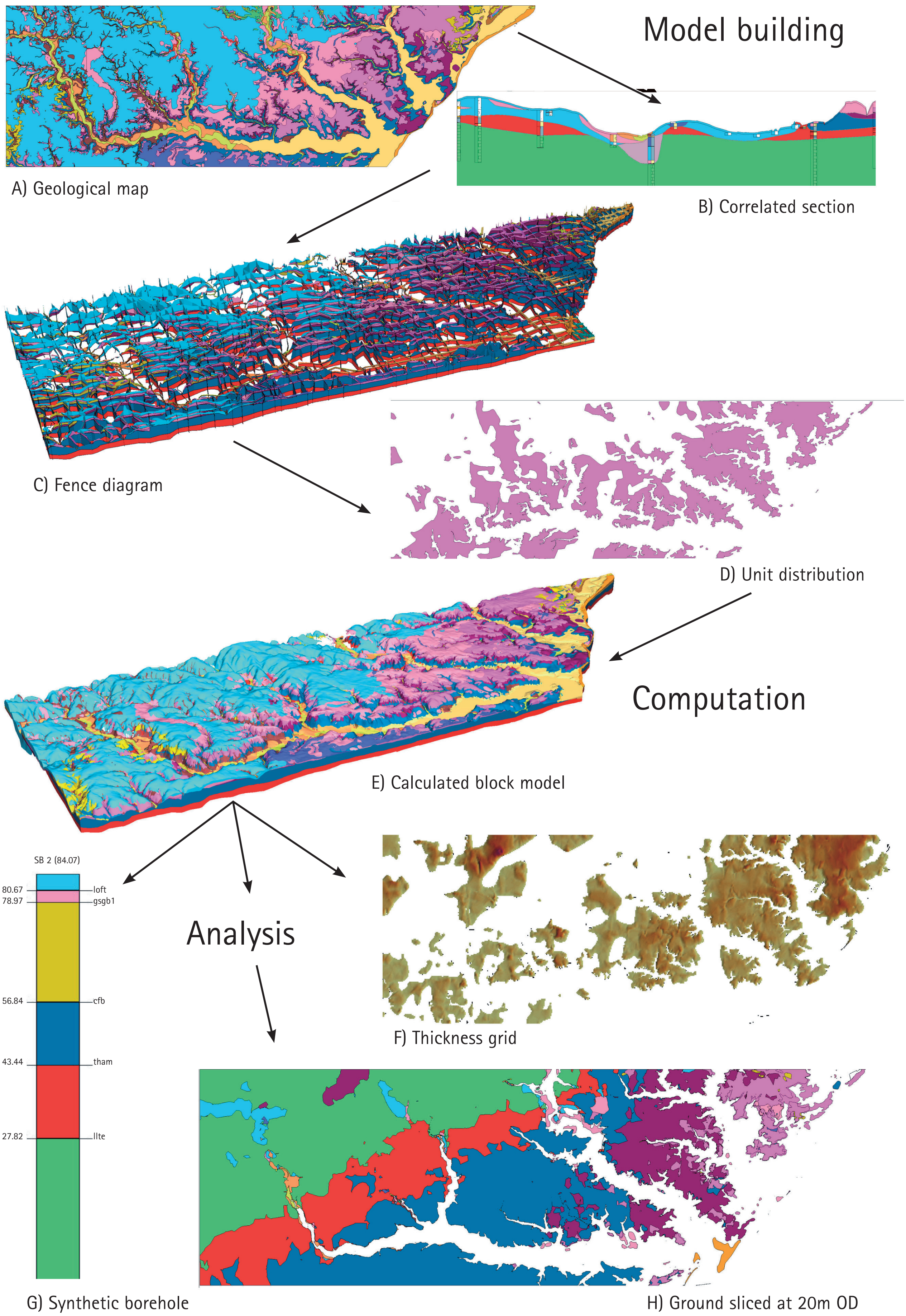


Figure 2: The GSI3D modelling workflow (from Kessler, H and Mathers, S J. 2006).

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

The BGS have developed groundwater models that more closely represent the structure of hydrogeological systems, producing flexible models which can both conform to aquifer geometry and simulate processes at different scales. In collaboration with the University of Birmingham and the Environment Agency (EA), the BGS have developed the ZOOMQ3D[7] as a discrete process model that is able to effectively model flow in a saturated groundwater system.

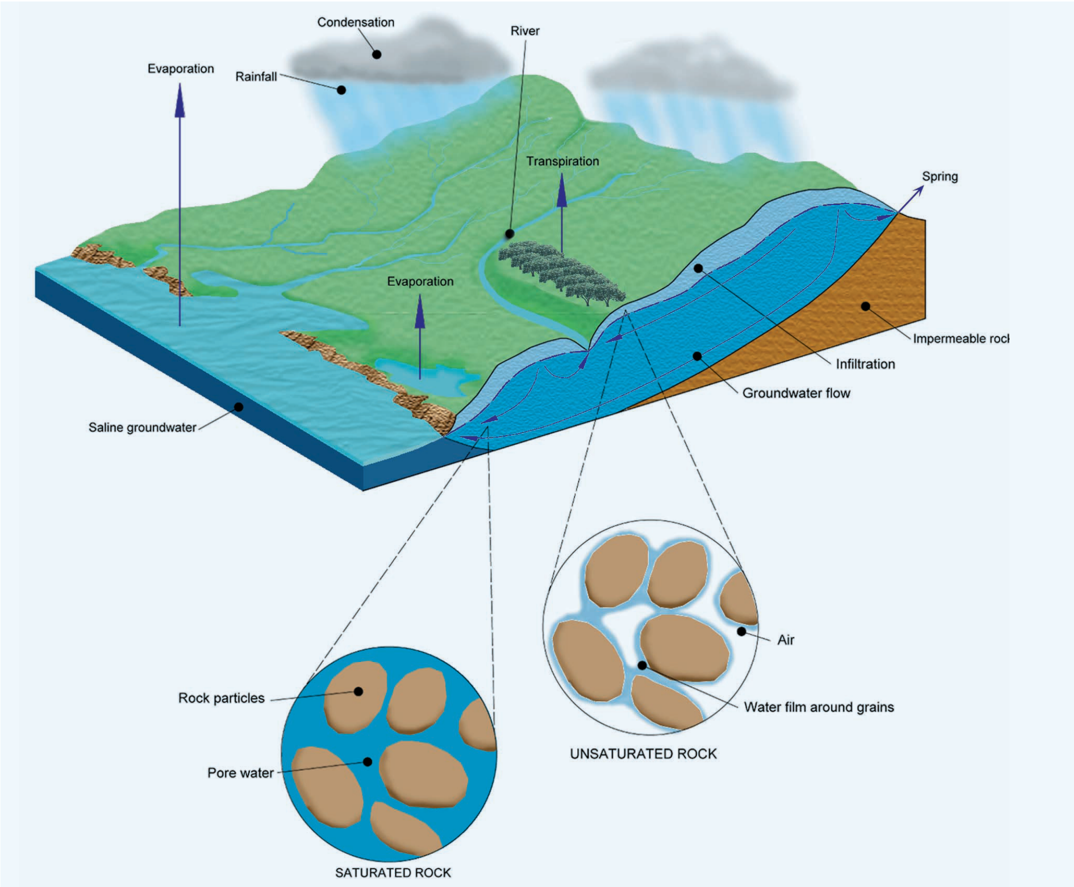


Figure 3: The hydrological cycle.

2.4 Linked process models

When a number of discrete processes have been successfully modelled, an expert can create new knowledge by taking the outputs of these models and making an assessment of all or part of the system. In the case of water\groundwater an expert 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 the expert role in the above process. The only way was to replace the existing models with a single new model that attempted to replicate the functions of the existing discrete process models. This is a slow and expensive process which 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 required key parameter. This approach has a number of advantages:

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

[7] http://www.oomodels.info/pmwiki/pmwiki.php/Main/OOModelsHome

3 Science case for DREAM

The scientific problem that DREAM is addressing has been well articulated by Reitsma and Albrecht (2005 and 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 trans-disciplinary semantics, or even explicit domain specific semantics, reduces the ability of linked models to create real understanding.

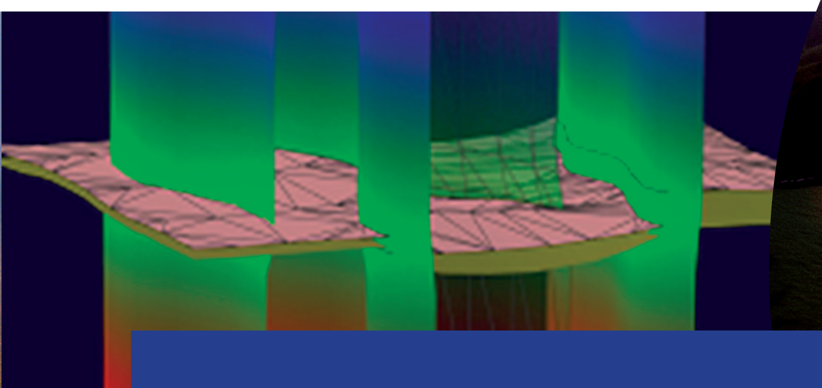
The science case for the DREAM project is based upon:

- The analysis of a series of internal and external drivers;
- A portfolio of anticipated benefits which will be realized by the projects;
- A range of risks that will be mitigated; and
- A series of user stories that describe their aspirations.

3.1 Drivers

The key drivers are:

1. Next Generation Science for Planet Earth–NERC Strategy 2007-2012.
2. Applied geoscience for our changing Earth-BGS Strategy 2009-2014.
3. BGS delivery plan 2010-2014.
4. The increasing need to develop an holistic approach to environmental models covering all aspects of the environmental sciences rather than geoscience in isolation.
5. The expectation that the BGS' income from NERC will decline over the period of the current BGS strategy.
6. The changes introduced by NERC under the Funding Allocation and Budget (FAB) project and the need to generate a greater percentage of the BGS' overall income through grants won as part of collaborative teams.
7. European environmental legislation including infrastructure for spatial information Europe (INSPIRE).



3.2 Benefits and risks

The following benefit portfolio and risk inventory has been developed:

- Through an analysis of the drivers;
- Through analysis of the challenges (page 9);
- Through discussions with potential users, both internal and external, of the system and its outputs;
- Through the input of many people in the start-up workshop;
- From the reports of visits and conferences attended during the study;
- A review of lessons learned in previous BGS technology lead projects;
- A SWOT analysis;
- Users stories that capture the expectations of potential users;
- Discussions with potential collaborators.

3.2.1 Benefits

1. Progress the BGS strategy for 2009-2014.
2. Progress the NERC strategy for 2007-2012.
3. DREAM will increase the impact of NERC on policy development at national, European and international level.
4. DREAM outputs will lead to the development of better tools for Earth System Science approaches to environmental modelling.
5. DREAM will promote a holistic Earth Systems Science approach to environmental modelling rather than a discipline specific approach as at present.
6. Promote trans-disciplinary collaborative working.
7. DREAM will increase the opportunity for the BGS to be involved in national, European and international collaboration.
8. DREAM will leverage new funding opportunities for NERC and BGS.
9. DREAM will provide the opportunity for more accurate quantifications of hazard predictions.
10. DREAM will provide the opportunity for more accurate quantification of resource estimation.

3.2.2 Risks

1. The BGS will fail to win an increasing share of its income from collaborative grant applications.
2. The DREAM project begins designing solutions and commits to specific technologies before the questions DREAM should be answering are fully defined.
3. Cultural differences between scientific disciplines could hinder or even prevent effective modelling of interacting processes.
4. Other Environmental Modelling Platforms become more widely accepted by the wider modelling community than the solutions selected by DREAM therefore making the project outputs of less relevance.
5. The platform produced through DREAM is not widely accepted as a 'trusted brand' by users.
6. The interfacing method(s) between modelling tools or data sources are not simple enough. Resulting in frustrated users and low take up.
7. The outputs from DREAM don't accurately quantify uncertainty resulting in lack of user confidence.
8. IPR is not adequately protected; therefore the BGS or any of its partners within the DREAM framework may incur damage to their finances or reputation.
9. Geosciences continue to be ignored by other scientific areas when addressing questions that perhaps should consider geological features or processes.

4 The challenges the DREAM project faces

To achieve the vision there are a range of challenges that have been identified during the DREAM Scoping Study Project for which solutions must be sought during the following implementation project. These challenges are:

- **Software** – Select the most appropriate software methodologies to achieve DREAM project ambitions.
- **Ontology and Semantics** – Linking models also links the concepts and classifications of those disciplines and the language used to describe them. To achieve DREAM 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, that is a system 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** – DREAM 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** – DREAM must promote collaboration between researchers both within and across disciplines.
- **Workflows** – DREAM should reduce the chaotic nature of modelling multidiscipline environmental issues and enable ordered repeatable processes to be put in place.

These challenges are discussed in greater detail below.

4.1 Software

Challenge
To select the most appropriate software methodology to achieve the DREAM project ambitions.

At the heart of the DREAM 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 et al. (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 impacting on the survival of the juvenile fish. The presentation entitled – 'An Integrated Framework for Improved Stream Temperature Predictions to Mitigate Fish Mortality' described 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 Open Modelling Interface (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[8] [9], 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

[8] http://ec.europa.eu/environment/life/
[9] http://www.openmi-life.org/

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 that integration is far from straightforward. The OpenMI Standard has provided an option which 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 (Devroedea et al 2008).

4.2 Ontology and semantics

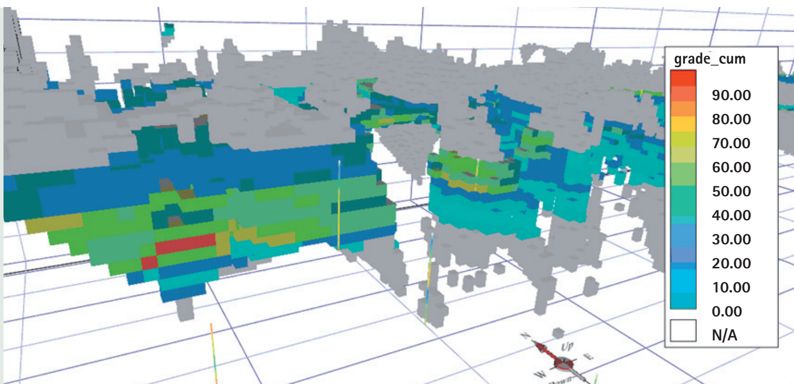
> **Challenge**
> Linking models also links the concepts and classifications of those disciplines, and the language used to describe them. To achieve DREAM goals requires ontological and semantic alignment.

Ontology is the branch of metaphysics that deals with the nature of being, whilst semantics is the branch of linguistics concerned with meaning. 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 exists 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 are currently largely transactional. Information is requested and exchanged and there are simple, automated tests to make sure that transactions were completed as anticipated. However, there is little domain knowledge held by either computer in a transaction, neither of which have any significant inference ability, to verify that the transaction was both successful and that knowledge exchanged was correct (Reitsma et al. 2009).

The DREAM project 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 encourages 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. More information is available at http://www.w3.org/TR/owl-features/.

4.3 Scale

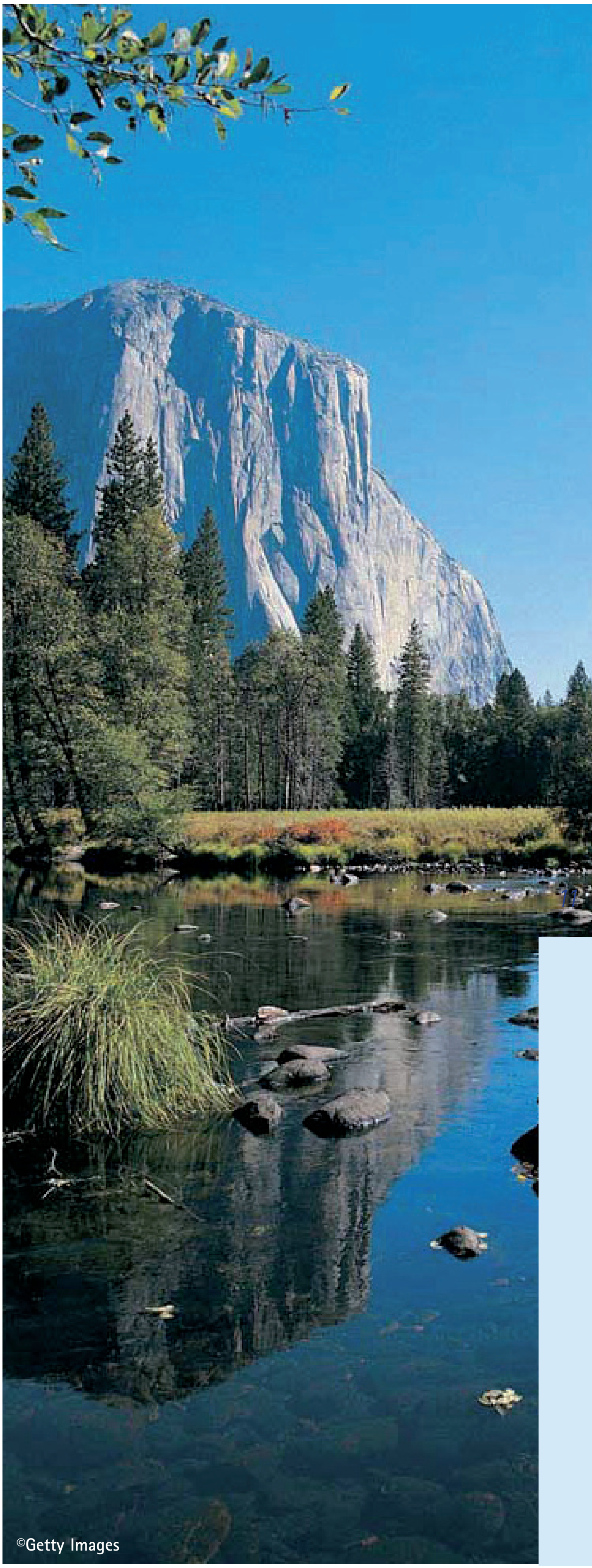
The environment is affected by processes that operate from the micron-scale to the solar system scale and potentially beyond. Studies of aquifers polluted by dense non-aqueous phase liquid (DNAPL) have shown that a model of the behaviour of the pollutant within the pore spaces between the grains of the sedimentary material contribute to remediation of the polluted sites (Goody et al 2002 and Wealthall 2002). At the other end of the scale is space weather that requires monitoring and modelling of the state of the space environment. It requires understanding of the behaviour of energetic particles as well as in 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 man-made technologies on Earth and in space, for example, on satellites and spacecraft, electricity power grids, pipelines, radio and telephone communications and on geophysical exploration.

Challenge
Systems must take into account the range of scales over which environmental processes operate – from microns to the scale of the solar system and beyond.

Solutions that are developed during the DREAM 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 wellbeing of an overcrowded island like Britain. There are two challenges relating to scale:



1 How to develop process models in heterogeneous environments where critical parameters may be at micro scales and also at kilometre scales? An example is 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, measured at the kilometre scale.

2 In geology a common problem is the uneven distribution of the available data. This leads to the requirement to 'upscale' and 'downscale'.

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

   b Down scaling is the reverse problem to up scaling in which limited regional scale information is leveraged to produce a more detailed local scale understanding.

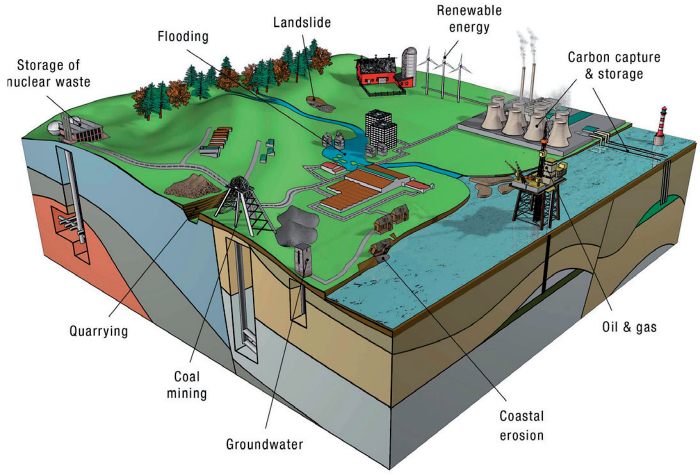
The challenge is to ensure that solutions produced by the DREAM Implementation Project take full account of the range of scales required in environmental modelling and are not restricted to only a limited scale range.

4.4 Uncertainty

> **Challenge**
> 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.

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 then the natural reaction of any scientist is to add complexity to their models. In other words, as data is collected and understanding correspondingly improves then more and different processes can be added to any model. However, as more processes are added then the model requires more parameters, these all have 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 DREAM 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 to date. The limited number of models that have been linked together, to be used as predictive tools, seem to have avoided addressing the issue of combined uncertainty.

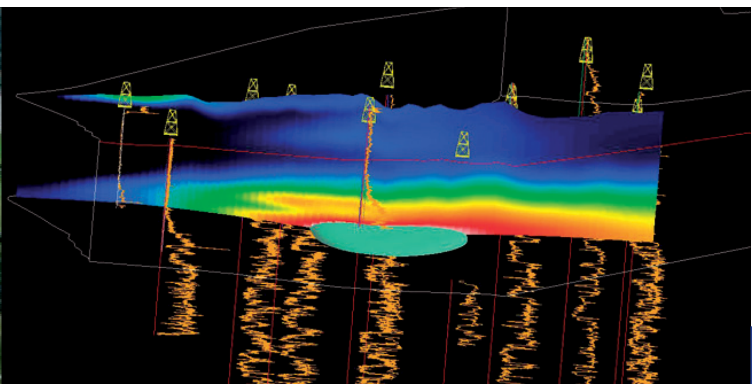
It is the objective of the DREAM project to link together framework and process models to produce a more complete understanding of the natural environment. Without a clear understanding of the uncertainties inherent in the combined models the predictions they produce will have little credibility.

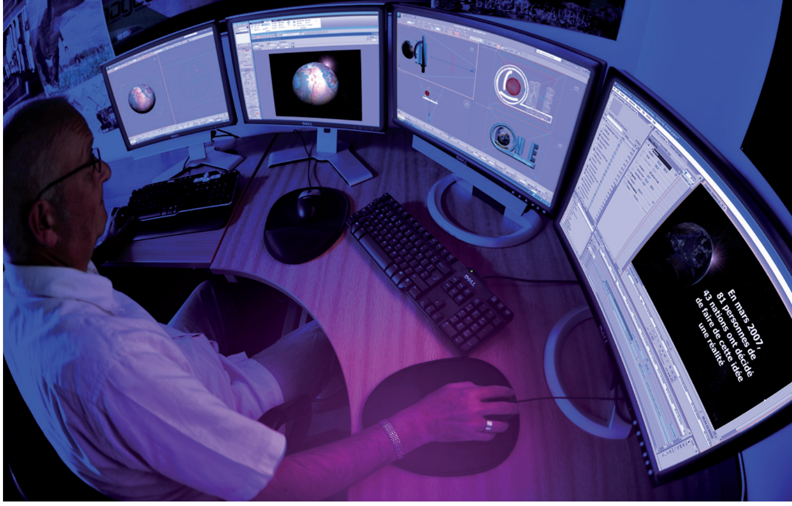
Research is being undertaken into model uncertainty. For example the GoCad Research Group, based at Nancy Universite in France, is becoming increasingly interested in uncertainty. Professor Caumon, Nancy Universite, recognises the success of 3D modelling and its growing importance as a major tool in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

4.5 Heterogeneity

Challenge
Natural systems are heterogeneous, that is a system consisting of multiple components each of which may have considerable internal variation. Modelling Earth systems requires recognition of the inherent complexity.

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 et al. (2003) in the International Association of Hydrological Sciences (IAHS) Science Plan.

Earth systems are made up of many individual processes that are related but can vary independently. The variation may reflect natural cycles, which may be over a short time scale (e.g. the season) or longer term (e.g. orbital forcing and resulting climate change). Time-series data from observations of component processes within Earth systems may not capture the whole 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.

4.6 Data

Challenge
Ready access to well managed data, in appropriate formats, associated with rich metadata is essential for success.

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

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

These eight dimensions naturally fall into two groups. The first group reflects quality and the second is management.

4.6.1 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 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 is a good example. The original data collected in the field is process 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.

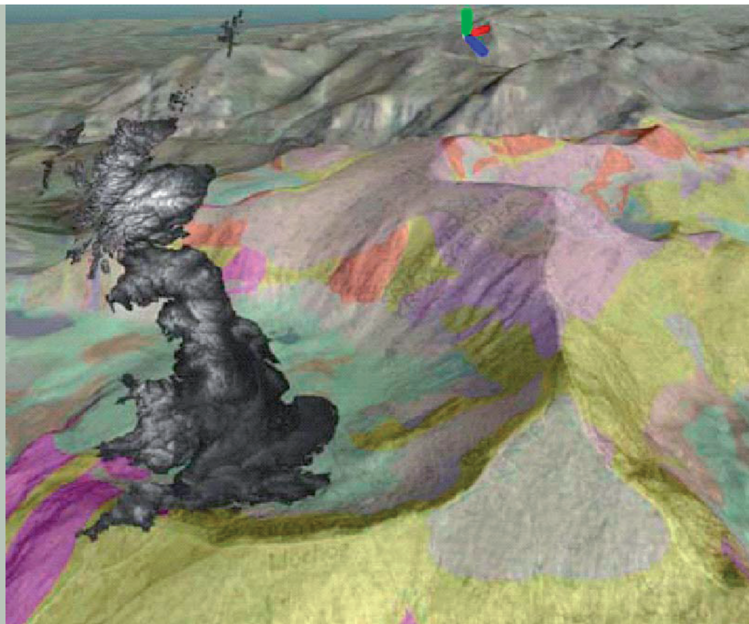
4.6.2 Data management

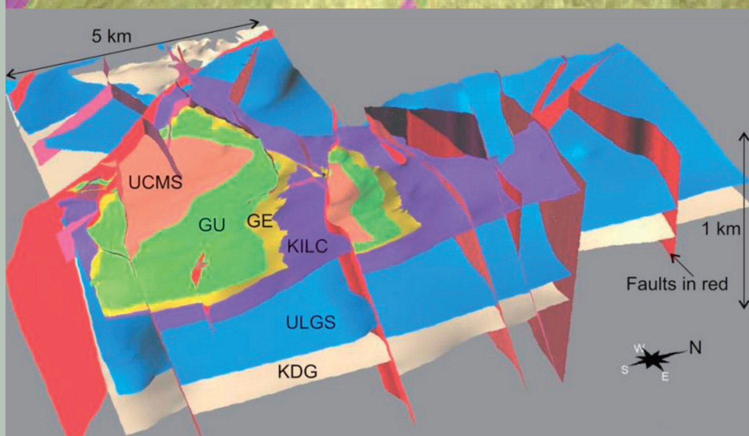
Well managed datasets are those that are easily accessible, contain timely data and are stored in a secure environment.

Scientists spend considerable amounts of time searching for and formatting datasets so that they are usable (see text box). Well managed dataset 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

An internal audit undertaken by Shell International in 2002 revealed that their exploration geoscientists working in 'new frontiers' areas spent as much as 53% of their time finding the relevant data, 24% of their time archiving and documenting the data, and a meagre 23% of their time interpreting and adding-value to those data. Shell responded by setting targets aimed at increasing time spent on interpretation and adding-value to 46% by reducing time spent on finding data to 30%.

Source: NERC Science Information Strategy 2009





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 represent the current state of knowledge, or the state of knowledge at the time of data collection\synthesis is recorded and described.

4.7 Intrusion

> **Challenge**
> Any solution must leverage the investment in existing models rather than attempt to replace them. The challenge is to enhance the best of the existing environmental models by linking them together.

Intrusion is an important concept in relationship to the DREAM project. A single organisation will not succeed if it proposes an approach which assumes that all other organisations will abandon their existing approaches, and the associated investments, and adopt the new approach. It would be too intrusive if the DREAM 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[10]. Numerous environmental models have been produced to aid the study of various aspects of the natural environment. A study by the European Environment Agency (EEA) produced a report called 'Modelling environmental change in Europe: towards a model inventory'[11] . The report looked at more than eighty models that had been recently used in environmental assessments by the EEA. 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 DREAM 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 DREAM Implementation Project take into account the existing range of environmental models and leverage the significant investment, rather than committing considerable resources into trying to replace well established models.

4.8 Standards

> **Challenge**
> The DREAM proposal will have succeeded when its outcomes are recognised as formal International Standards.

There are a wide range of standards that are applicable to the domain of environmental modelling. DREAM should not add to these unless absolutely necessary. The DREAM 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 through to national and international standards. The adoption of this approach will reduce the potential conflict within

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

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

the community and will reduce the risk of having to re-engineer systems at some later date when one standard becomes dominant.

4.9 Visualisation

> **Challenge**
> Environmental models are most easily understood by their users when the output is an easy to interpret 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 decision and policy makers it is essential that DREAM outputs have a clear visual interfaces that are simple to use.

An example of such a system is WaterSim[12]. 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 gauge future water-supply conditions in response to climate change, drought, population growth, 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 documenting, with 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[13].

4.10 Culture change

> **Challenge**
> DREAM must promote collaboration between researchers both within, and across disciplines.

Individuals, small groups of researchers and 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[14] (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 decision and policy makers to choose the appropriate models and to have confidence in the model results.

For DREAM to work there will need to be considerable collaboration and promoting this change is one of the principle challenges for the project. It will require influencing research funders to promote collaboration in grant application and to recognize the importantance of trans-disciplinary research. Communities that use large instruments, such as astronomers and high-energy physicists, have developed means of collaboration

[12] http://watersim.asu.edu
[13] http://www.learner.org/courses/envsci/index.html
[14] http://www.modflow.com

that recognize individual contribution whilst promoting collaboration.

4.11 Solution workflows

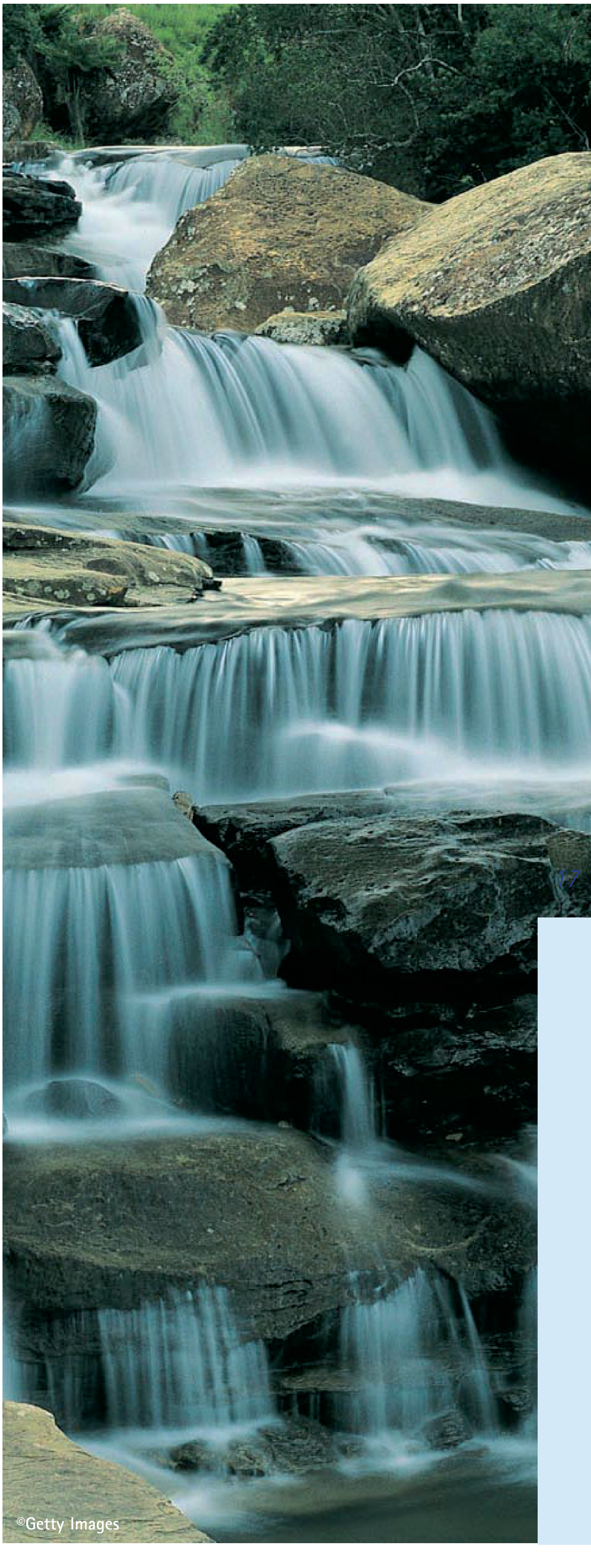
Tackling multidiscipline environmental questions requires individuals from each discipline to contribute information from their area of expertise. When all of the information is combined in the correct sequence the resulting workflow contributes to the solution.

Challenge
DREAM should reduce the chaotic way that multidiscipline environmental issues are modelled, and enable ordered repeatable processes to be put in place

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.

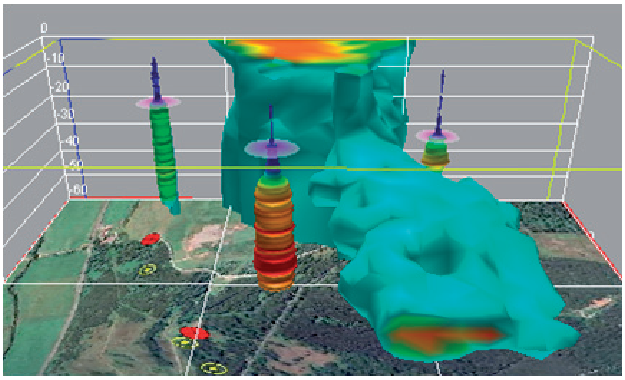
To produce a range of answers based upon a variety of scenarios often requires a significant amount of manual re-processing. Each time a new scenario is modelled there is a danger that the steps taken are inconsistent with previous model runs, leading to solutions or answers that cannot be reliably compared.

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


©Getty Images

5 The approaches we will adopt to deal with the challenges

Each of the challenges identified in the previous section needs to have an approach identified to deal with the issue. These are described below.



5.1 Software approach

The BGS regularly uses two types of models, 3D geological framework models and process models. The purpose of 3D geological framework models is to understand the relationships between stratigraphical units within a 3D volume. They can be created and maintained in a number of commercial products and an increasing number of more open software tools. Process models are used to understand an environmental process such as the behaviour of a pollutant within a groundwater body. In the geosciences a process model's behaviour may be constrained by stratigraphical units modelled by framework models. The main functions of a process model are estimation and prediction.

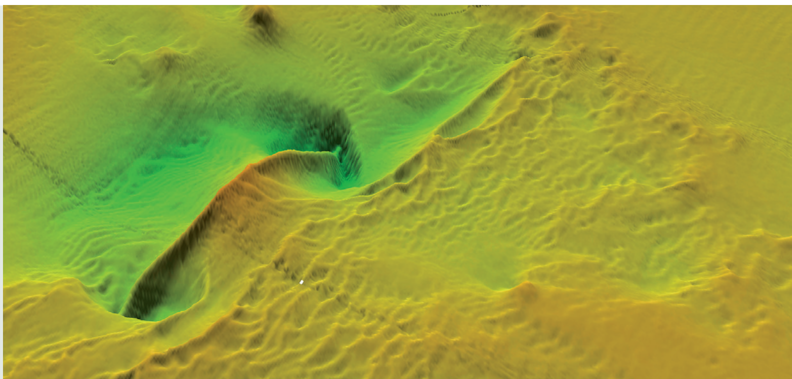
Framework models are often constructed at a range of scales for the same area. This approach provides regional context as well as local detail. The next version of GeoSciML should be able to effectively represent 3D geological models. Process models are more challenging to integrate, as they are often custom designed mathematical models.

There appear to be two approaches in the literature to linking process models:

- Custom written software that hard codes the links between two or more specific models; and
- A model interface standard that allows compliant process models to communicate with other compliant models.

Gregersen et al. (2007) recognises, that the management issues in many sectors of society demand integrated analysis that can be supported by integrated modelling. They also recognise that all-inclusive modelling software is difficult to design and deploy. Gregersen's preferred approach to integrated modelling is the linkage of individual models or model components that address specific domains. The recommendation of the DREAM scoping project is that the BGS adopts the second approach as articulated by Gregersen. The OpenMI Association[15] have developed a standard that acts as glue that can link existing compliant models together so that they become components of a larger model. Gregerson et al. (2005 and 2007) describe how OpenMI provides a standardized interface to define, describe and transfer data on a time basis between software components that run simultaneously. This supports systems where feedback between the modelled processes is necessary in order to achieve physically sound results. The OpenMI allows the linking of models with different spatial and temporal

[15] http://www.openmi.org



representations; for example, linking river models and groundwater models, where the river model typically uses a one-dimensional grid and a short timestep, and the groundwater model uses a 2D or 3D grid and a longer timestep. The OpenMI is designed to accommodate the easy migration of existing modelling systems, since their re-implementation may not be economically feasible due to the large investments that have been put into the development and testing of these systems.

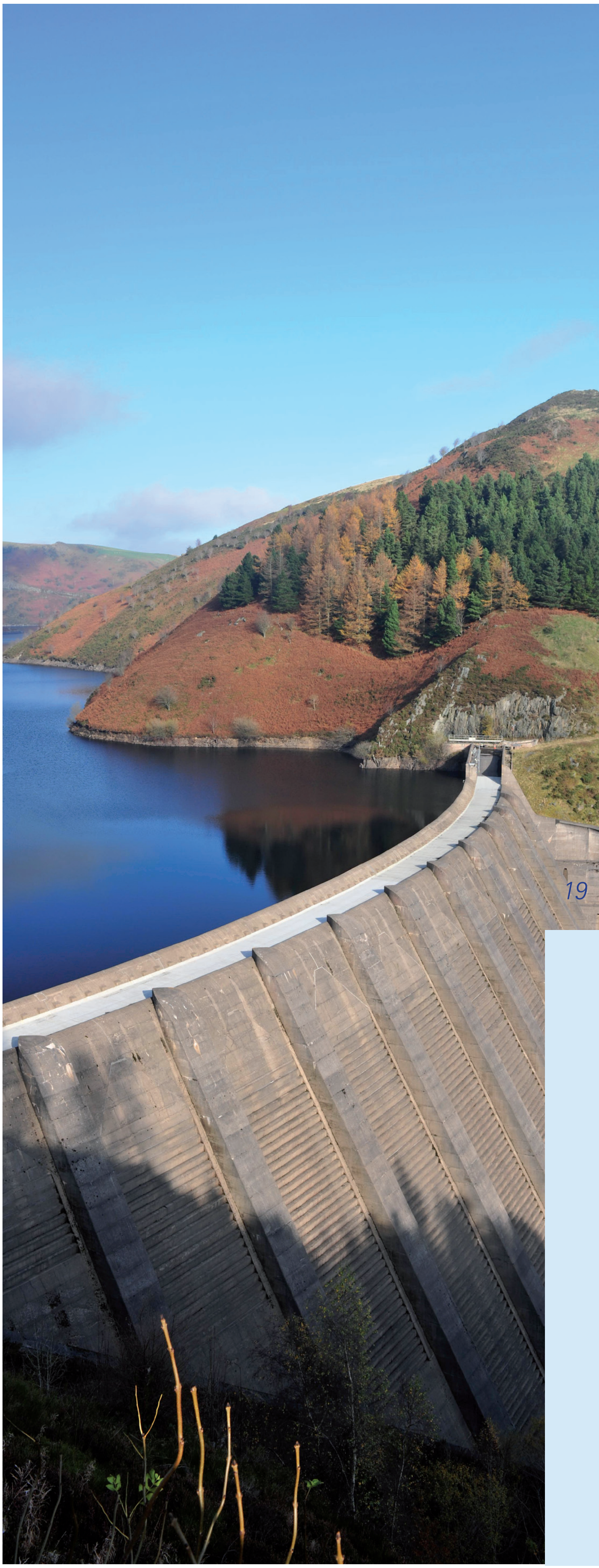
At the recommendation of the scoping study project the BGS has become a member of the OpenMI Association and has partnered in a proposal to the EU to further the development of OpenMI.

5.2 Ontology and semantics approaches

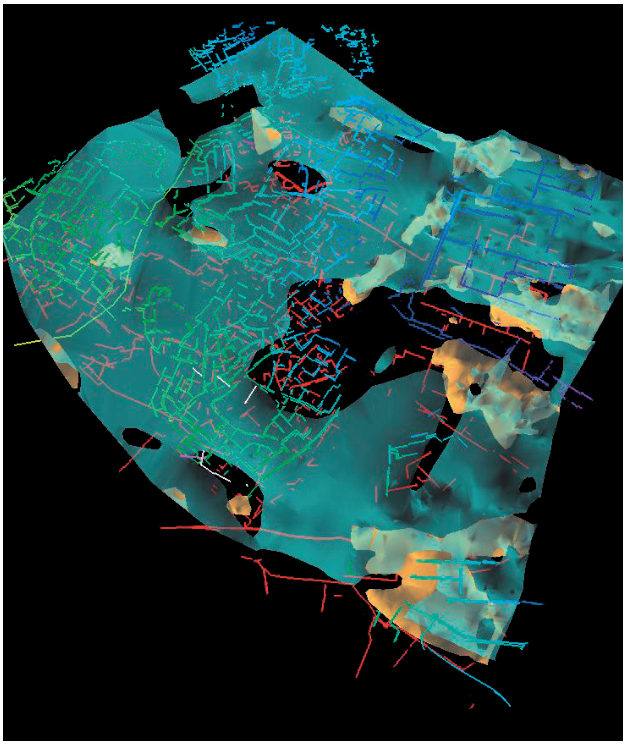
The DREAM project requirements include semantic interoperability between models that are linked together to create a more comprehensive understanding of the environment. The World Wide Web Consortium (W3C), under the title of Semantic Web, has developed the most mature framework for semantic interoperability. The Semantic Web uses standard formats to integrate and combine resources from diverse sources. This allows a person, or a machine, to follow a trail of links through multiple data sources. One of the results of this is to enable natural language querying. Three simple things achieve this:

- Unique identifiers for all resources (URI — Uniform Resource Identifiers);
- A standard way of describing resources (RDF — Resource Description Format);
- Ontologies.

A URI is a string of characters used to identify a name or the location of a resource on the Internet, or both. URI is the high level term that has sub-types of Uniform Resource Locator (URL) and Uniform Resource Name (URN). A URL is a sub-set of a URI used as a means to specifies where an identified resource is available and the mechanism for retrieving it. The most familiar URL is a web address such as http://www.wikipedia.org/, which identifies the resource that is the international home page of Wikipedia. A URN is a sub-set of a URI used as a means of uniquely identifying a resource. URNs are intended to serve as persistent, location-independent resource identifiers. A prime example is an ISBN (International Standard Book Number). This URN can be used to identify a specific edition of Shakespeare's Romeo and Juliet without



indicating where the resource may be obtained. This framework allows a resource to be uniquely identified and its location described. The resource that is identified by a URI can be described in a standard way using RDF. At the heart of RDF is the 'graph' which acts as a representation of the resource description. In many ways it is similar to an entity-relationship diagram used in classical concept modelling familiar to databases designers. This approach makes it easy to produce a series of logical statements about the resources in the form of a subject-predicate-object expression. These expressions are known as triples in RDF terminology. In a RDF triple the subject is the resource being described. The object is some property of the subject. The predicate both identifies the type of object property and also expresses the relationship between the subject and the object. For example: the MALACHITE specimen with URN (xyz) has a COLOUR whose value is BRIGHT GREEN. The subject is the 'malachite specimen with URN (xyz)', the predicate is the statement 'has a colour' and an object describing the colour of the subject 'bright green'. Many other RDF triples could be created to describe the malachite specimen.



The power of RDF becomes evident when a number of RDF triples are available. To demonstrate this; an animal feed manufacturer may ask where they can obtain a copper dietary supplement for animal feed-mixes. One answer is 'Doar-Na Eilot, Israel', which is obtained when a search engine navigates a machine-readable version of the following RDF triples.

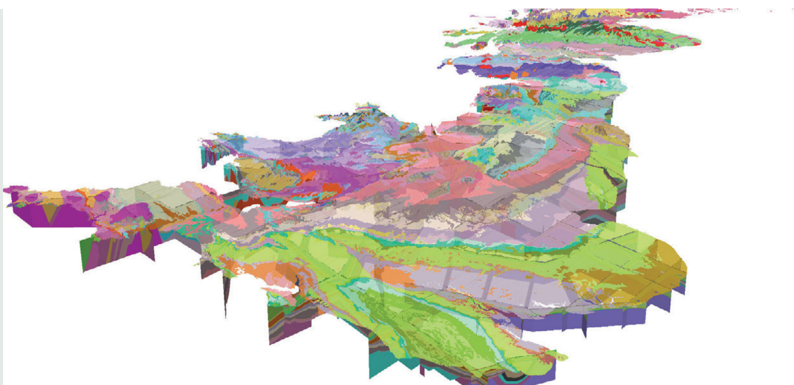
- The TIMNA MINE with URN (abc) has a PRODUCT whose value is BASIC COPPER CARBINATE with a URN (mno).
- BASIC COPPER CARBONATE with a URN (mno) has a USE whose value is COPPER DIETARY SUPPLIMENT FOR ANIMAL FEED-MIXES.
- The TIMNA MINE with URN (abc) has a LOCATION whose value is DOAR-NA EILOT, ISRAEL

These three RDF triples need not be located on the same machine or even on the same continent.

The similarity between RDF 'graphs' and entity relationship modelling means that it is technically relatively trivial to produce RDF triples from relational databases such as Oracle®. Oracle 11g® has direct support for semantic technologies make the production of RDF triples straightforward, provided the desired information is already stored within the database.

Over the past 25 years the BGS has invested considerable resources in developing vocabularies that have been turned into dictionaries for use in our information systems. The vocabularies are now available publicly as downloads and web services, see:

http://www.bgs.ac.uk/data/vocabularies/home.html



In addition to the vocabularies there are a number of mature ontologies. The principle ontologies managed by the BGS are the Lexicon of Named Rock Units and the Rock Classification Scheme. The BGS is part of both the GeoSciML community and the Commission for the Management and Application of Geoscience Information (CGI) Multi-Lingual Thesaurus Working Group. The GeoSciML community is working with the Geoscience Concept Definition Working Group of the Commission for the Management and Application of Geoscience Information to promote semantic interoperability within the geosciences. This Concept Definition Working Group has published a range of vocabularies in 2008 and has a SimpleLithology under review and will be published during 2010. The CGI Multi-Lingual Thesaurus Working Group aims to enable the global exchange of geoscience information, with the help of a common multilingual core vocabulary, by developing and expanding the Multilingual Thesaurus of Geosciences. The aim is that the thesaurus will be compatible with commonly accepted international standards, easily accessible and free of charge.

Many of the basic components are in place to develop the semantic framework that will be required. However, the development of ontologies for the geoscience disciplines has been limited to keyword lists for classification. The potential of ontologies and the semantic web have yet to be tapped for scientific modelling and simulation (Reitsma and Albrecht 2006). This is a real opportunity for the BGS to show leadership in an area which is essential.

5.3 Scale, uncertainty and heterogeneity approaches

Scale, uncertainty and heterogeneity are closely related issues. It is essential that a clear understanding of them is developed in relation to discrete and linked process models. This work is a significant piece of research in its own right. The approach will be to work with an appropriate university department and seek funding for one or more research students to address the problem and recommend an approach.

5.4 Data approach

The principles of data quality and data management are well understood and documented. The mature international metadata standards provide mechanisms for recording the necessary data quality information. The issue is ensuring that the necessary information about a dataset is accurately recorded in order to ensure that the quality is understood. There are two threads to this: (i) a management process and (ii) an organisational culture change.

The management process required to improve data quality is largely covered in the NERC Science Information Strategy (2009) which will be implemented across

NERC over the next few years and will, hopefully, set a benchmark for others to follow.

It must be acknowledged that within in some disciplines, especially the analytical disciplines, there is a robust culture of recording quality information. However, culture change is required in many areas to ensure that all scientists systematically capture and record the quality data within the metadata. This is a persistent and intransigent in some disciplines and must be tackled. The approach is:

- To systemically explain why the quality data is necessary;
- To create systems that automatically capture as much quality metadata as possible;
- To encourage and support the establishment of peer reviewed data journals;
- To develop system of dataset peer review; and
- To encourage peer-to-peer challenging of scientists who do not provide quality information for the data that they produce.

5.5 Intrusion approach

The approach here is at the same time both the simplest and the most challenging. The approach must be humility. It must be recognised that there are numerous environmental models in existence. Each has something to contribute to the overall picture of the environment. In addition there will be a number of solutions available to link models together. History teaches us that when an issue becomes apparent to one individual or group it will have also become apparent to many other groups. Bryson (2003) notes that while Charles Darwin was pondering the theory of evolution, so was Alfred Russel Wallace; a well documented story. More surprisingly both were beaten to it by a Scot called Patrick Matthew, who had published his idea of natural selection while Darwin was still on the HMS Beagle.

We need to learn the lessons from others. Tapscott and Williams (2006) describe how International Business Machines (IBM) was spending large sums unsuccessfully competing with the emerging Linux operating system.

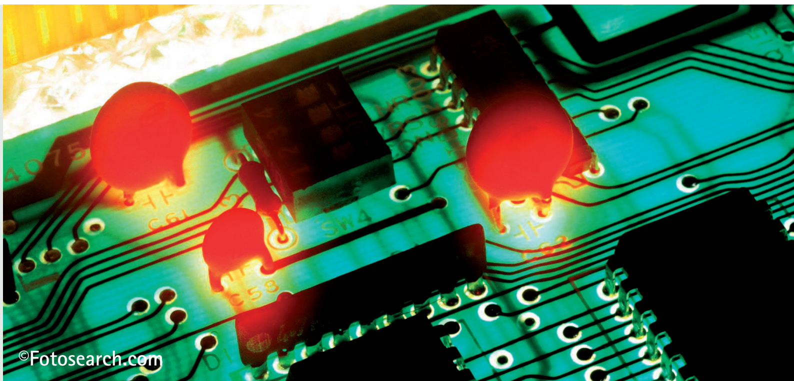
In 1998 they decided that they would support Linux and invest in the development.

IBM engaged with the community at two levels, firstly by showing leadership, and secondly by doing the less glamorous activities in the Linux workflow such as promoting standards and producing documentation. In this way they typically gained community buy-in and respect.

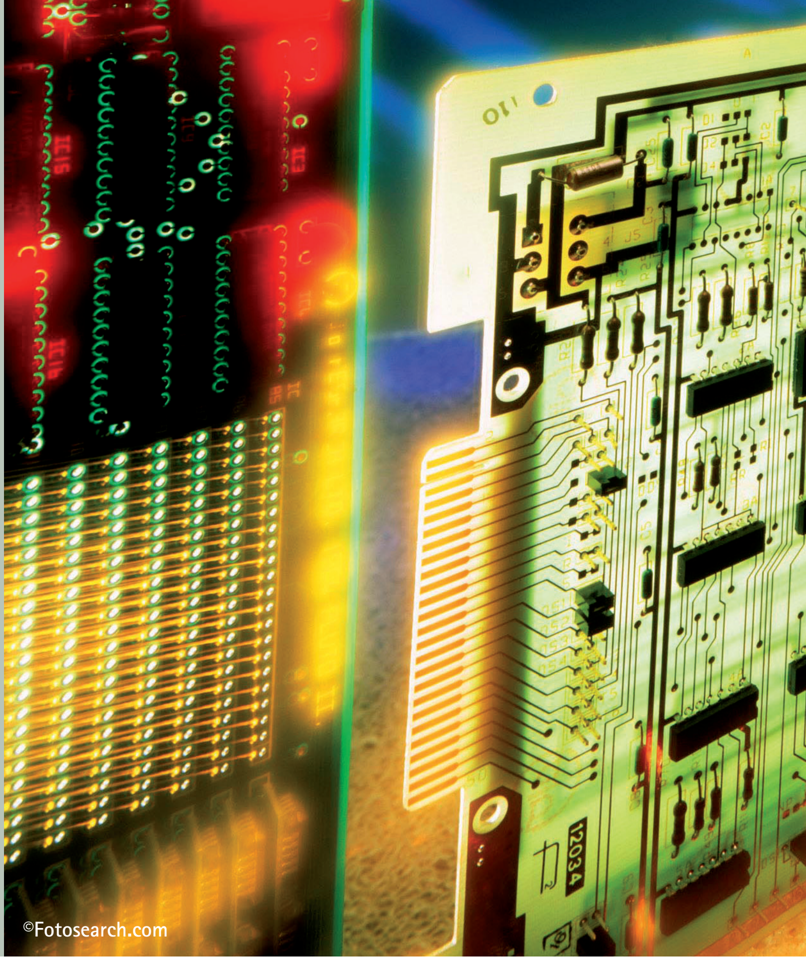
The BGS' recent experience with OneGeology suggests that such community engagement is possible and by so doing the activity becomes a focus of funding. By its formal launch in at the International Geological Congress in Oslo in September 2008 OneGeology had attracted major funding from both the EU eContentPlus Programme and the USA National Science Foundation as well as a best practice network from over 100 countries.

The BGS could adopt this model by taking a leading in the developing community and at the same time taking on the less glamorous actives such as:

- Leading the bids and managing EU projects;
- Leading on grant writing;
- Developing standards and documentation for the system;
- Acting as the facilitator to build partnerships; and
- Ensuring that all BGS IPR in this project is truly 'Open'.



"IBM spend about $100 million per year supporting Linux development. If the Linux community puts in $1 Billion of effort, and even half of that is useful to IBM customers, the company gets $500 million of software development for an investment of $100 million. 'Linux gives us a viable platform uniquely tailored to our needs for 20% of the cost of a propriety operating system" says Joel Caley (IBM VP.)



The BGS is also trialling the EMP on the Cross-cutting projects that are running in parallel to DREAM.

5.6 Standards approach

Our object is to create a community that will contribute to the development of an open environmental modelling platform. The BGS aims to be one of the leaders in this community. Such a development has to be standards driven and where the needed standards are absent, the BGS must take a lead in their development. We should learn from the Global Spatial Data Infrastructure (GSDI) community. It is clear that the success of this community of taking desktop GIS to a GSDI has been based around the development of a large number of standards. The key series is the ISO 19000 series of standards along with the wide range of W3C standards. The development of standards is not glamorous work as it involves numerous technical meetings, both at the national and the international level. The BGS should learn the lesson from IBM Corporation who contributed to the development of Linux by developing many of the unglamorous elements of the system such as producing robust documentation. In the process IBM developed a world leading understanding of the system and a matching reputation.

5.7 Visualisation approach

The models used in the sub-surface may come from different developers, represent processes from different domains, be based on different concepts, and have different spatial and temporal resolutions (Sivapalan et al. 2003; Reitsma and Albrecht 2005 and 2006). Data standards such as GeoSciML aims to deliver a common conceptual data model on the nature and structure of the geoscience information, to which data held in individual databases can be mapped and consequently transferred between users. Using common practices and languages like those described above will facilitate the development of a subsurface information system, which will aim to inform planners, regulators and other decision makers through 3D and four dimensional (4D) models of the combined surface and subsurface environment. Dissemination of subsurface data and models will be via commonly used tools such as Nederlands Instituut Voor Toegepaste Geowetenschappen's (TNO) DINOLoket Web service (van Wees 2003) and software like the Subsurface Viewer, which allows the integration of artificial subsurface and surface infrastructure with natural environment features such as geological structures (Figure 4). The means of disseminating data and information just described is only the beginning of meeting the needs of planners, regulators and other decision makers by visualizing data in the context of the real world. There is an identifiable need for a comprehensive multi-dimensional subsurface management system forming the basis for spatial, volumetric, temporal decision making in the same way as today's GIS systems are used for two dimensional spatial planning, insurance risk assessment, or emergency planning. It is vital that this system is not developed in isolation from the real end-users and also that the system is able to deal with the wide variety of subsurface models that exist in the Geological Survey Organisations (GSO) across the world.

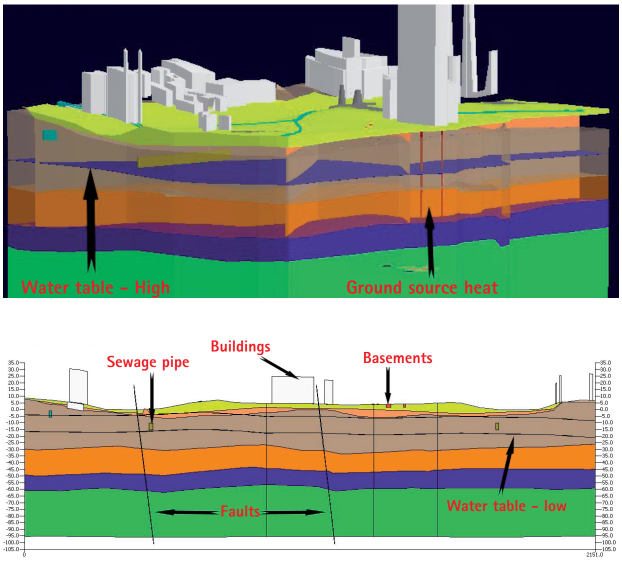


Figure 4: A mock-up model with artificial infrastructure integrated within the geological model visualised in 3D and in cross-section using the Subsurface Viewer.

5.8 Culture

The cultural challenges presented by the task of developing the EMP both within the scientific community and in the wider community are immense. To take the challenges in the scientific community first, then one of the biggest is to change the practice of the scientist working as an individual, without significant cross-organisation collaboration. This is exemplified within organisations where career progression is dependent on publishing in the peer-reviewed literature. If single author publications are required then this will discourage collaborative working. Therefore, a mechanism has to be sought to encourage the sharing of data, models and understanding within the scientific community.

In the wider community where models use by decision-makers of all types from the householder to the politician is to be encouraged, the challenge is different. Routine use of models by decision makers of all types is the ideal, but there is a suspicion of models by the non-scientific community where there is only a partial understanding of models and the process of developing them, and their advantages and disadvantages are not appreciated. This issue can be overcome by improving how models can be accessed and used by decision-makers of all types. The important advance is to provide model results interactively using appropriate visualisation techniques. An example of good practice is the 'Decision Theater' developed at the University of Phoenix (e.g. Gober, 2006). This allows a range of decision-makers to access simulations interactively.

5.9 Solution workflows

The approach recommended by the DREAM project team to mitigate solutions being produced in an ad hoc manner is through the use of Solution Workflows. In this context a Solution Workflow is a description of each component and step taken to produce an answer. A Solution Workflow would be needed each time a new environmental question required models and data sources to be linked or configured differently.

In its simplest form a Solution Workflow describes a repeatable process; this can be achieved by fully documenting each of the required components and the sequence of steps that need to be executed. This should be done in a way that is clear and available to all stakeholders. If the Solution Workflow is suitably documented a potentially chaotic process can be transformed into one which is both standardised and repeatable.

DREAM aims to use technology to design Solution Workflows as configurable systems that allow users to alter parameters and re-run the system to produce a range of outputs.

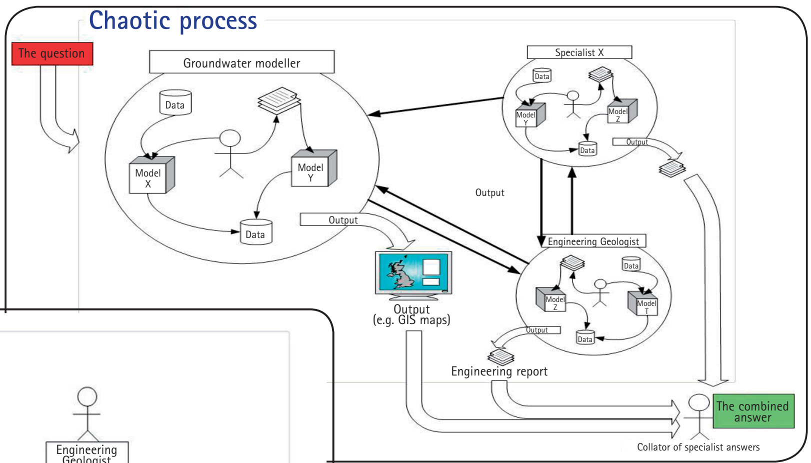
Visual configuration of a workflow:

It is possible to use the OpenMI Configuration Editor to visually link components and subsequently run the linked system (Gregerson et al. 2005).

DREAM will evaluate OpenMI Configuration Editor to assess whether it is a suitable way of designing Solution Workflows.

Repeatable scenario driven processes:

It is also possible to produce tools that enable non-expert users to select scenarios, alter parameters and run a Solution Workflow based on OpenMI interfaces; for example the graphical user interface AM-DSS (Agricom Mozart Decision Support System) (Dirksen et al., 2005). Such a system could be developed to help both the managers and users of a groundwater resource to understand the issues around over exploitation. It could illustrate the balance that needs to be maintained between extraction and recharge.



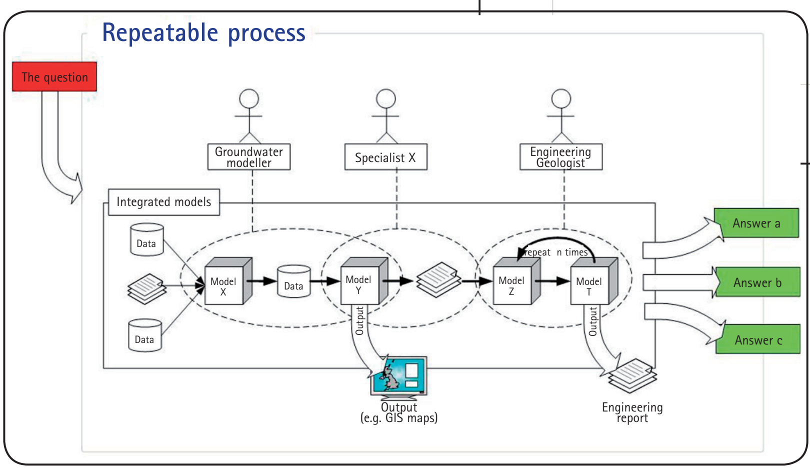


Figure 5: Visual configuration of a workflow.

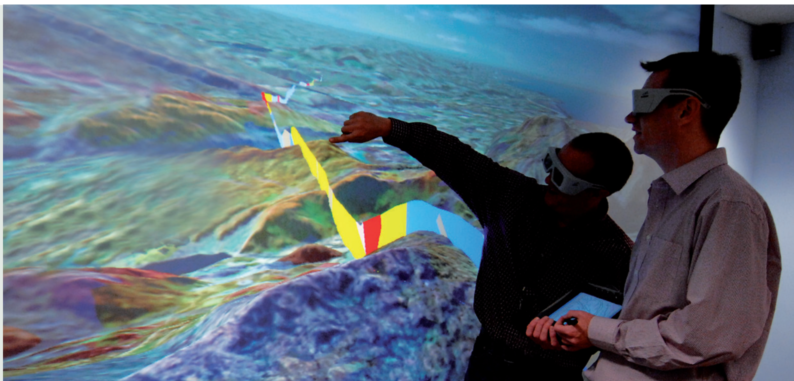
6 What will the project deliver?

The challenges of this project are larger and more far-reaching than the scope and resources of the British Geological Survey. Therefore, any description of the project deliverables must be circumspect. Project success will depend upon the building of an open community that will be successful in both leveraging that communities' existing resources and attract additional funding through research grants, EU project and sponsorship.

6.1 Year 1 deliverables

The deliverables in the first year can be defined with more confidence that those in subsequent years. There are the project set-up deliverables, such as the Project Initiation Document, the first of the test beds to produce a linked model within the London area, a number of communication tasks and finally the planning stages for a future EU proposal.

Product	Outline product description
Project Initiation Document	The first deliverable will be a formal Project Initiation Document (PID) based on the scoping study that will act as the Project Mandate.
Set up project board	Invite suitable internal and external candidates to sit on project board; arrange meetings.
Promoting understanding within the BGS	Contacts with other Cross-cutting project leaders, Team leaders arranging internal dissemination including lunchtime lectures and workshops.
Public version of the scoping study report	A public version of the scoping study report should be printed as both a glossy publication and as an e-book on the website. This is essential for communicating the vision and scope of the project.
Visit report actions completed	During the scoping study a number of visits were made to conferences and organisations. The full visit reports contain a range of actions. It is important that these are followed through and completed.
Toward a Global Geoscience Initiative	Draft a single page proposal to Geological Society of America, Geological Society of London, American Geological Institute, and the USGS for a follow on to IYPE based around DREAM concepts.

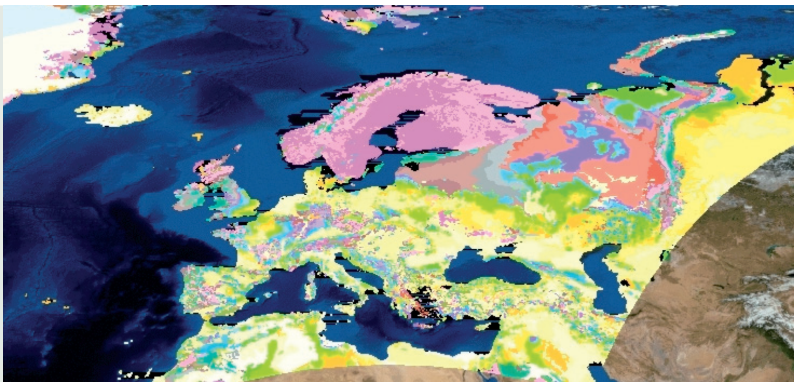


Discussion paper on leading semantic framework development for the geosciences	Produce a discussion paper which will contain a costed proposal and a plan for leading the development of a semantic framework for the geosciences.
Evaluate the CSDMS 'system'	The Community Surface Dynamic Modelling System (CSDMS) undertook a scoping study on the technology available in 2004. This should be evaluated and a report with recommendations produced.
Evaluate EEA decision support system	Edwards Aquifer Authority (EEA), Texas, has a decision support system which provides real time groundwater heads (see http://www.edwardsaquifer.org/pages/J17RealTime.asp). Produce a short report with recommendations.
EarthKnowledge study for US DoE	EarthKnowledge are a partner in a US DoE contract to scope an Earth Modelling System. The final report (due May 2010) should be of interest to DREAM and should be reviewed.
Test Bed 1: London Basin Model	The first test bed will be to develop a linked model of the London basin. This will bring together the hydrogeological models, the hydrological model and the rainfall run-off model. A report will describe the work undertaken, the successes and issues and recommend follow on actions.
National Workshop	Model fusion workshop in the UK to promote the DREAM project vision, promote the London Basin Model, disseminate OpenMI Life experience, engage with NERC theme leaders, Living with Environmental Change programme and other similar programmes.
International Workshop Planning	Start the planning for an international workshop to take place in 2011.
International Conference	Promote the DREAM concepts at American Geophysical Union (AGU), EGU, USGS Modelling Conferences etc.
Geosciences Network (GEON) Links	Arrange for a geoscience IT specialist to attend the GEON cyberinfrastructure Summer Institute for geoscientists at the San Diego Super Computer Centre, California.
Establish links with a university research partner	Scale, uncertainty and heterogeneity in discrete and linked process models are an issue which should be addressed in conjunction with an appropriate university. Links will have to be established and research proposals made.
EU Proposal Planning	Start the planning of an EU proposal to further the DREAM concepts. Possible candidates are water management, regeneration of brown-field sites, waste disposal and nuclear waste storage.
Data Management	A report describing the requirements for data and its quality for modelling within the BGS.
Model Management	A report describing digital preservation issues related to the long-term management of linked models.

6.2 Year 2 & 3 deliverables

The aims of the second and third years are; to build on the successes and lesson from the first year, launch a 2nd test-bed in the Clyde basin, communicate results and build a community to submit an EU bid.

Product	Outline product description
Prototype workflow developed	A report documenting the workflow used in the development of linked models in the London Basin Test Bed.
Promote London Basin Linked Model	Promote the London Basin Linked Model developed in the first year in grey-literature, at conferences and in peer-reviewed literature.
Seek review of the London Basin Linked Model	Seek to get the London Basin Linked Model rigorously peer reviewed by third parties. Produce a report documenting the review findings and the lessons learned.
NERC Strategy Development 2013-2018	Engage with the NERC Strategy development team to ensure that there is an opportunity for multidisciplinary environmental modelling in the 2013-2018 NERC Strategy.
Challenges	Work with an appropriate university department to understand uncertainty, scale and the semantic and ontological issues within linked models.
A semantic framework for the geosciences	Lead the development of a semantic framework for the geosciences based upon plans developed in year 1. This must be based on collaboration with other organisations who will together seek the necessary funding.
Test Bed 2: Clyde Basin Model	The second test bed will be to develop a linked model of the Clyde basin. This will bring together the hydrogeological models, the hydrological model and the rainfall run-off model. In addition the aim will be to integrate the social and economic models to extend the scope of the work. A report will describe the work undertaken, the successes and issues and recommend follow on actions.
Extend London Basin Model	Extend last years London Basin Test Bed by linking additional models from other domains such as social and economic.
EU Proposal	In year 1, planning was done towards creating an EU proposal to study an aspect of the environment and its social and economic impact through using linked models. This proposal should be completed and submitted.



Poverty Alleviation	Lead a community approach to Department For International Development (DFID) to develop a proposal to use linked models as part of the strategy for poverty alleviation.
Visualisation of Linked Models	Individual models in a linked series can be visualised. However the existing standards do not permit the visualisation of the output of linked models. This needs to be addressed, as visualisation is a powerful tool for communication of results\ interactions.
Teaching tool	Work with university partners to develop a teaching tool using linked models. It must have an easy to use interface and convey challenging science in understandable ways. WaterSIM is a good example: http://watersim.asu.edu/ .
Standards Development	Engage with the British Standards Organisation to work out how to progress the outcome of the work towards British and International Standards.
Maintain links with a university research partner	Maintain links with the university research partner working on scale, uncertainty and heterogeneity in discrete and linked process models.

6.3 Year 4+ deliverables

The final years of the project will be about capturing and disseminating the knowledge created. It will also be about influencing the future direction of the BGS strategy for the second half of this decade. Projects started by the community should be brought to a close. The standards work should be progressed but will continue beyond the life of the cross-cutting project.

Product	Outline product description
EU project leadership and participation	On the assumption that the project will have been successful in negotiating an EU project effort will be required to lead and participate
The BGS Strategy Development	The BGS strategy for 2015-2020 will be in development and the project needs to influence this strategy in light of project outcomes.
Standards development	Progress the establishment of standards related to the work
Web-based Visualisation of linked models	Extend the linked model visualisation to work on the Internet
Knowledge capture and disseminate it	Systematic capture of the knowledge created during the project and dissemination to other staff and partners.
Workflow documentation	Document the workflows developed.
Peer-reviewed publications	Publish a series of high-impact peer-reviewed publications.
Final Report	Produce final project report and publish widely.
Follow-on Actions	Document follow on actions required at post-project.
Project Closure	Formally close project according to the BGS' 'Project Closure Procedure'

7 Critical success factors—how do we know if we have succeeded?

- Sizeable Best Practice Network or open community actively supporting the DREAM/Open Environmental Modelling Platform(OEMP) initiative and its underlying aims and ideals.
- Secured funding from project collaborators and other EU/Worldwide funding vehicles.
- Increasing amount of developed environmental models made OPENMI compliant to be used in linked systems at runtime to facilitate multidisciplinary collaborative science.
- Significant web presence including range of information to differing target audiences as well as documentation, discussion forms, standards developments, best practice, examples etc.
- Considerable exposure of DREAM/OEMP initiative and peer-reviewed test cases in papers, articles and conferences.
- Several clear examples of linked models thoroughly peer-reviewed and used to demonstrate the power of linked model multidisciplinary collaborative science to answer pressing questions of the day. Examples must cross disciplines and bring together data/models from diverse specialism's (including geology, water, environmental science, social science etc).
- Increasing number of EU/Worldwide standardised vocabularies and ontology's created and in use by multiple collaborators from differing disciplines.
- Aspects of 'linked model' visualisation tackled and advances made. This is a critical development in the ability to provide solutions to decision makers rather than solutions for scientists.

- Increasing amount of standardised and harmonized scientific data made available (including metadata and quality information) for modelling.
- Community perception of minimal intrusion into their existing defined modelling processes and flows.

8 References

Athanasiadis, I N and Janssen, S. 2008. Semantic mediation for environmental model components integration, Information Technologies in Environmental Engineering, Vol 1, Jan 2008.

Barkwith, A. 2010. DAEM Model Inventory. British Geological Survey Open Report, OR/10/010. 74pp.

British Geological Survey. 2009. British Geological Strategy 2009–2014. Applied geosciences for our changing Earth. (Nottingham: British Geological Survey)

Bryson, B. 2003. A Short History of Nearly Everything (London: Black Swan)

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 Trans. AGU, 90(52), Fall Meet. Suppl., Abstract IN34A-03.

Chatfield, C. 1995. Model Uncertainty, Data Mining and Statistical Inference. Journal of the Royal Statistical Society. Series A. 158. Part 3. 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.

Dirksen, P W; Blind, M W; Bomhof, T and Srikrisnhudu Nagandla. 2005. Proof of Concept of OpenMI for Visual DSS Development. In Zerger, A and Argent, R M. (eds) MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2005, pp. 184-189. ISBN: 0-9758400-2-9. http://www.mssanz.org.au/modsim05/papers/ascough_1.pdf

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

Feineman, D R. 1992, Data Management: Yesterday, Today and Tomorrow. Presentation to the PETEX '92 Conference, 25 November 1992, London

Gober, P. 2006. Metropolitan Phoenix: Place Making and Community Building in the Desert. Philadelphia: University of Pennsylvania Press, 233 pp.

Gooddy, D C; Bloomfield, J P, Harrold, 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 vol. 59 no. 3-4,: pp. 247-265

Gregersen, J B; Gijsbers, P J A and Westen, S J P. 2007. OpenMI : Open Modelling Interface. Journal of Hydroinformatics, 9 (3), 175-191.

Gregersen, J B; Gijsbers, P J A; Westen, S J P and Blind, M. 2005. OpenMI: The Essential Concepts and their Implications for Legacy Software. Advances in Geosciences, 4, 37-44, 2005.

Kessler, H and Mathers, S.J. 2006. The past, present and future of 3D Geology in BGS. Open University Geological Society Journal 27(2), 13-15.

Kessler, H; Mathers, S J & 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, Volume 35, 1311–1321.

Loudon, T V. 2009. Four interacting aspects of a geological survey knowledge system. Computers & Geosciences, Volume 35, Issue 4. 700-705.

Natural Environment Research Council. 2007. Natural Environment Research Council Strategy 2007-2014. Next Generation Science for Planet Earth. (Swindon: Natural Environment Research Council)

Natural Environment Research Council. 2009. Natural Environmental Research Council Science Information Strategy. (Swindon: Natural Environment Research Council)

OpenMI-Life project. 2010. Final report—Task C: The Pinios River Use Cases, unpublished report

Oreskes, N. 2003. The role of quantitative models in science, In Models in Ecosystem Science, ed. Canham, C D; Cole, J J and Lauenroth, W K. pp. 13-31, Princeton: Princeton University Press, US.

Reitsma, F and Albrecht, J. 2005. 'Modeling with the Semantic Web in the Geosciences,' IEEE Intelligent Systems, vol. 20, no. 2, pp. 86-88, Mar./Apr. 2005, doi:10.1109/MIS.2005.32

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

Reitsma, F; Laxton, J; Ballard, S; Kuhn, W and Abdelmoty A. 2009. Semantics, ontologies and eScience for the geosciences, Computers & Geosciences, Volume 35, Issue 4, Geoscience Knowledge Representation in Cyberinfrastructure, April 2009, Pages 706-709, ISSN 0098-3004, 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, 48(6) pp857-880.

Tapscott D, and Williams A D. 2006. Wikinomics—How Mass Collaboration Changes Everything (London: Atlantic Books)

van Wees, J D; Versseput, R; Allard, R; Simmelink E; Pagnier H; Ritsema, I and Meijers, P. 2003. Dissemination and visualisation of Earth system models for the Dutch subsurface in New Paradigms in subsurface Predictions, Edited by Rosenbaum, M S and Turner, A K. Lecture Notes in Earth Sciences, vol. 99, p.225

Wealthall, G P. 2002. Predicting DNAPL source zones in fractured rock. PhD thesis, Dept of Civil & Structural Engineering, University of Sheffield, UK.

Appendix 1: Project mandate

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 3D 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[16].

[16] British Geological Survey Strategy 2009-2014. Applied geoscience for our changing Earth. (Nottingham: British Geological Survey)

Appendix 2: Concept note

Both the NERC Strategy and LWEC are encouraging environmental science communities to work more closely with each other to model and understand the environment and to make predictions upon which policy can be developed. To do this the NERC and LWEC need to change behaviours within the environmental science community and one driver for this would be the creation of an open environmental modelling platform developed by the community as a true platform for participation. The BGS Strategy 'Applied Geoscience for our changing Earth' embraces the culture of collaboration and includes the development of such a modelling platform as a major initiative, together with crosscutting projects that would be used as test-beds for the validation of the platform and exemplars for the development of process models that would use the platform.

Various parts of the environmental science community are building modelling platforms to meet the specific requirements of their own communities and projects. For example the British Geological Survey has been developing GSI3D[17] to enable the systematic production of geological framework models[18]. This application is being used by a multinational geoscience community as evidenced by the attendees and speakers at the 2nd International Conference held in September 2008[19]. The BGS now needs to add parameters to the modelled 3D volumes to enrich the models and improve our understanding of processes in order to make decisions about our changing environment.

[17] http://en.wikipedia.org/wiki/GSI3D

[18] Kessler, H; Mathers, S J and Sobisch, H-G. 2008. The capture and dissemination of integrated 3D geospatial knowledge at the British Geological Survey using GSI3D software and methodology. Computers & Geosciences. Available on-line: http://dx.doi.org/10.1016/j.cageo.2008.04.005

[19] http://www.bgs.ac.uk/science/3Dmodelling/gsi3dMeeting2008.html

Furthermore we have to better understand the purpose and limitations of these models and begin to communicate their uncertainty. It is becoming clear that to do so we need to work with other environmental science communities who are, or have developed independent modelling systems. Another example might be the Joint UK Land Environment Simulator (JULES), being developed for community use by the Centre for Ecology and Hydrology (CEH) and the Meteorological Office (Met Office). At the same time for instance, to understand groundwater. as a sustainable natural resource, we need to create linkages between the 3D geological models and the groundwater models such as the ZOOM family of groundwater models[20]. In turn we need to integrate with hydrological and meteorological models and understand predicted changes in precipitation over time. There are many other scenarios, including other variables, like ecology, socio-economics, and so on that will require linkage to these types of environmental predictive models. These models will be required for the science community and then for practitioners in governments and their regulatory agencies to provide the decision making capability to enable us to live with environmental change.

©Jupiter Images

Currently all the communities' modelling systems have a number of common issues:

- They tend to be disciplined bases supporting only one element of the environmental science community;
- Little consideration is given at the design stage to interfacing with other environmental science communities;
- If there is a need to interface with other parts of the environmental science community the modelling platforms may need significant redesign or complex interchange formats have to be considered;
- Often large parts of the software code produced are proprietary; and
- Environmental Data Centres find it difficult to provide web services that meet the needs of a broad breadth of the environmental modelling community.

Initiatives such as GeoSciML[21] and OpenMI[22] try to bridge the gap between the different modelling communities by creating common standards on data formats and ontologie as well as publishing input/output (I/O) formats of different modelling software.

The platform would be:

- Open;
- Interdisciplinary;
- Designed with interoperability at its core;
- Standards based, building on the Open Geospatial Consortium (OGC) International Standards Organisation (ISO) standards and others;
- Community owned and built;
- Free to use; and
- Enable environmental data centres to provide appropriate web services.

Industry experience suggests that there are real advantages to open platforms. IBM was spending large sums unsuccessfully competing with the emerging Linux

[20] http://www.oomodels.info/pmwiki/pmwiki.php/Main/OOModelsHome

[21] http://www.geosciml.org/

[22] http://www.openmi.org/reloaded/

operating system. In 1998 they decided that they would support Linux and invest in the development.

"IBM spend about $100 million per year supporting Linux development. If the Linux community puts in $1 Billion of effort, and even half of that is useful to IBM customers, the company gets $500 million of software development for an investment of $100 million. 'Linux gives us a viable platform uniquely tailored to our needs for 20% of the cost of a propriety operating system" says Joel Caley (IBM VP)[23].

IBM engaged with the community at two levels, firstly by showing leadership, and secondly by doing the less glamorous activities in the Linux workflow such as promoting standards and producing documentation. In this way they gain community buy-in. The NERC could adopt this model by leading the community and encouraging the developers within NERC to do some of the less glamorous activities essential to a successful platform.

The BGS' recent experience with OneGeology suggests that such community engagement is possible and by so doing the activity becomes a focus of funding. By its formal launch in at the International Geological Congress in Oslo in September 2008 OneGeology has attracted major funding from both the EU eContentPlus Programme and the USA National Science Foundation.

The BGS is keen to progress this idea with a six month international scoping study commencing in April 2009. The benefits of the scoping study would be:

- Demonstrate NERC leadership;

[23] Tapscott, D and Williams A D. 2006. Wikinomics – How Mass Collaboration Changes Everything Atlantic Books London.

- Build community engagement;
- Articulate the vision and objectives of the open environmental modelling platform;
- Identify any existing parallel initiatives developing around the world and seek to engage with them;
- Identify appropriate international standards under which the platform should develop;
- Recommend leadership framework under which the development would be steered; and
- Identify components that need accelerated development to create a 'critical mass' that others can build upon.

The appendix will contain a glossary of terms. Vic Loudon has supplied a starting list of terms, will use some, alter some and delete some.

Appendix 3: Glossary of terms

Abstraction: reducing the information content of a concept or an observable phenomenon, typically in order to retain only salient information, relevant for a particular purpose (see Scale, Generalisation).

Complex system: A complex, emergent system has many adjacent parts that may interact according to simple rules without central control. Feedback mechanisms may result in effect not being proportional to cause and the linear equations of physics do not apply.

Cyber-: Derived from the Greek word meaning 'steersman' through cybernetics (the study of control mechanisms and feedback systems in animals and machines), used as a prefix (equivalent to e-) to indicate the electronic or computer based version of a conventional product or service.

Cybergeology: A view of geology as part of a knowledge system re-based on the developing cyber-infrastructure.

Cyber-environment: Aspects of the cyberinfrastructure assembled to meet requirements relevant to a particular field of enquiry, aiming to maintain global compatibility while providing access through interfaces that match users' working practices.

Cyberinfrastructure: An integrated assemblage of computing, information and communication facilities, deploying the combined capacity of multiple sites to provide a framework to underpin research and discovery, typically with broad access and end-to-end coordination. According to Wikipedia, the term cyberinfrastructure "was used by a United States National Science Foundation (NSF) blue-ribbon committee in 2003 in response to the question: how can NSF, as the nation's premier agency funding basic research, remove existing barriers to the rapid evolution of high performance computing, making it truly usable by all the nation's scientists, engineers, scholars, and citizens? The NSF use of the term focuses on the integrated assemblage of these information technologies with one another. Cyberinfrastructure is also called e-Science; in particular, the United Kingdom has a major e-Science initiative."

Data: A collection of observations, measurements or other information about a set of variables, generally in the form of numbers, words, or images.

Discrete Process Model: A process model that simulates a single or limited set of processes within a single model.

e-book: (also eBook, ebook) a book published in electronic form or a digital version of a paper book, readable by computer possibly with interactive content that includes hyperlinks and multimedia (1.2).

e-Science: See cyberinfrastructure.

Earth systems science: The unified study of the physical, chemical and biological components, processes and their interactions that determine states and changes in the planet. See Earth System Science Partnership (ESSP)

Feedback: The process whereby part of the output of a system is returned to an input control mechanism that regulates its further output.

Filtering: A process that selectively enhances or reduces specified components of the information stream.

Framework: A logical structure and guidelines giving a broad overview for classifying and organizing complex information, within which detail can be added as required.

Framework Model: A framework model is a tool to allow scientists to integrate desperate empirical observations into a coherent whole.

Geology: The study of the planet Earth, the materials of which it is made, the processes that act on these materials, the products formed, and the history of the planet and its life forms since its origin. Geology is defined in the AGI Glossary as: "The study of the planet Earth—the materials of which it is made, the processes that act on these materials, the products formed, and the history of the planet and its life forms since its origin. Geology considers the physical forces that act on the Earth, the chemistry of its constituent materials, and the biology of its past inhabitants as revealed by fossils . . . The knowledge thus obtained is placed in the service of man—to aid in the discovery of minerals and fuels of value in the Earth's crust, to identify geologically stable sites for major structures, and to provide foreknowledge of some of the dangers associated with the mobile force of a dynamic Earth." [Bates, R.L., Jackson, J.A. (eds.), 1980. Glossary of Geology (2nd edn.) American Geological Institute, Falls Church, VA.]

Holistic: A view of a system that emphasises its properties and interrelationships acting as a whole, as opposed to the reductionist approach of studying its components in isolation as distinct entities.

Interoperability: Interoperability of information: the ability of concepts, terms or models from various sources to work together, by meeting standards that enable sharing and reuse of information.

Interface: An interface is the shared boundary between systems or parts of a system, or the means of interaction across the boundary that makes joint operation possible.

Linked Process Model: A linked process model is made up of two or more discrete process models that are integrated at run-time through the use of a model linking standard.

Metadata: Metadata is a description of data that is structured to assist the user or computer to find, manage, control and understand the data.

Metamodel: A metamodel is a description of the organisation and function of a model, to assist the user or computer to find, manage, control and understand its contents.

Model: A formalised representation giving a simplified view of aspects of the real (or of an imaginary) world relevant to the purposes in hand.

Objects: Representations of real-world or conceptual things or entities of interest in a particular context.

Ontology: A formal representation and shared vocabulary describing concepts, entities and relationships in a domain of knowledge, typically providing a more detailed and rigorous machine-readable specification than a thesaurus or taxonomy.

Paradigm: The set of common beliefs and agreements shared between scientists about how problems should be understood and addressed [Kuhn, 1962].

Predictive: The ability of a model, for example, to predict values of a property or variable where they are not known directly by observation or measurement.

Resolution: In Geographical Information Systems, regarded as the minimum distance between two points on a map or image that can be distinguished by eye or other sensor.

Semantic Web: Berners-Lee et al. (2001)[24], described the Semantic Web as an extension woven into the structure of the existing Web, in which information is given well-defined meaning, improving the ability of computers and people to work in cooperation.

Simulation: Imitation of aspects of internal processes of a system and their results; usually to visualise, statistically compare with, or predict real-world occurrences.

System: A set of interacting parts that function as a whole. The systems approach involves study of linkages or interfaces between the component activities.

Uniform Resource Identifier (URI): is a string of characters used to identify a name or the location of a resource on the Internet, or both. URI is the high level term which has sub-types of Uniform Resource Locator (URL) and Uniform Resource Name (URN).

Uniform Resource Locator (URL): A Uniform Resource Locator is a sub-set of a Uniform Resource Identifier used as a means specifies where an identified resource is available and the mechanism for retrieving it. The most familiar URL is a web address such as http://www.wikipedia.org/ which identifies the resource which is the international home page of Wikipedia.

Uniform Resource Name (URN): A Uniform Resource Name is a sub-set of a URI used as a means of uniquely identifying a resource. URNs are intended to serve as persistent, location-independent resource identifiers. For example an ISBN (International Standard Book Number) is a URN that can be used to identify a specific edition of Shakespeare's Romeo and Juliet without indicating where the resource may be obtained.

Variable: A quantity that can assume any of a set of values.

Visualisation: Transforming quantitative data (including the results of interpolation) into sensory information—images that the eye and brain can interpret and visualise.

Workflow: The representation of a process or procedure in terms of a sequence of operations to be carried out to complete a task.

[24] Berners-Lee, T., Hendler, J. and Lassila, O., 2001 (May). The Semantic Web. Scientific American, 284 (3). http://www.scientificamerican.com/article.cfm?id=the-semantic-web

Appendix 4: Abbreviations

Abbreviations used in the report body are explained in the table below.

Abbreviation	Meanings
2D	Two dimensional
3D	Three dimensional
4D	Four dimensional
AGU	American Geophysical Union
BGS	British Geological Survey
CGI	Commission for the Management and Application of Geoscience Information
CSDMS	Community Surface Dynamics Modelling System
DEM	Digital Elevator Model
DFID	Department For International Development
DNAPL	Dense Non-Aqueous Phase Liquid
DREAM	Data and Research for Environmental Applications and Modelling
DTM	Digital Terrain Model
EA	Environment Agency
EMP	Environmental Modelling Platform
EN	English Nature
EU	European Union
FAB	Funding Allocation and Budget
GEON	Geosciences Network
GIS	Geographic Information System
GSA	Geological Society of America
GSDI	Global Spatial Data Infrastructure
GSI3D	Geological Surveying and Investigation System 3D
GSO	Geological Survey Organisation
IAHS	International Association of Hydrological Sciences
IBM	International Business Machines
INSPIRE	Infrastructure for Spatial Information Europe
ISO	International Standards Organisation
LWEC	Living with Environmental Change
NERC	Natural Environment Research Council
OEMP	Open Environmental Modelling Platform
OGC	Open Geospatial Consortium
OpenMI	Open Modelling Interface
OWL	Web Ontology Language
PID	Project Initiation Documentation
TNO	Nederlands Instituut voor Toegepaste Geowetenschappen
URI	Uniform Resource Identifiers
URL	Uniform Resource Locator
URN	Uniform Resource Name
USGS	United States Geological Survey
W3C	World Wide Web Consortium

Appendix 5: Description of OpenMI

The OpenMI can be described at two levels[25].

At the users level, the OpenMI provides a standard interface, which allows models to exchange data with each other and other modelling tools on a time step by time step basis as they run. It thus facilitates the modelling of process interactions. The models may come from different suppliers, represent processes from different domains, be based on different concepts, have different spatial and temporal resolutions and have different spatial representations including no spatial representation. A useful analogy is to consider the OpenMI as the modelling equivalent of a USB cable.

At the IT level, OpenMI standard is a software component interface definition for the computational core (the engine) of the computational models in the water domain. Model components that comply with this standard can, without any programming, be configured to exchange data during computation (at run-time). This means that combined systems can be created, based on OpenMI-compliant models from different providers, thus enabling the modeller to use those models that are best suited to a particular project. The standard supports two-way links where the involved models mutually depend on calculation results from each other. Linked models may run asynchronously with respect to timesteps, and data represented on different geometries (grids) can be exchanged seamlessly.

The OpenMI standard is defined by a set of software interfaces that a compliant model or component must implement. These interfaces are available both in C# and Java.

For more information please see:

www.openmi.org/reloaded/about/documents-publications/A_OpenMI_Scope.pdf

Overview of the OpenMI

This section describes the main user requirements for model linking, the functions currently provided by the OpenMI interface and the scenarios against which the standard has been tested.

Original requirements for the OpenMI

To be useful, the OpenMI must be able to link any models whose interactions need to be simulated in meeting the requirements of the Water Framework Directive. Against this background, the key requirements of the OpenMI are perceived to be:

[25] http://www.openmi.org/reloaded/about/what-is-openmi.php

- Linking models from different domains (hydraulics, hydrology, ecology, water quality, economics etc.) and environments (atmospheric, freshwater, marine, terrestrial, urban, rural etc.)
- Linking models based on different modelling concepts (deterministic, stochastic etc.)
- Linking models of different dimensionality (0, 1, 2, 3D)
- Linking models working at different scales (e.g. a regional climate model to a catchment runoff model)
- Linking models operating at different temporal resolutions (e.g. hourly to monthly or even annual)
- Linking models operating with different spatial representations (e.g. networks, grids, polygons)
- Handling feedback
- Handling iteration
- The ability to save its state and revert to a previous state
- Linking models using different projections, units and categorizations
- Linking models to other data sources (e.g. databases, user interfaces, instruments)
- Linking new and existing (legacy) models with the minimum of re-engineering and without requiring unreasonably high level IT skills
- Low barrier to entry
- Not impairing performance, especially of large models
- Being based on proven and available technologies (and, in particular, the architecture must be component-based and multi-layered)
- Linking models running on different platforms (e.g. Windows, Unix and Linux)
- Being 'open' (the interface specification should be placed in the public domain)
- Allowing components to be developed using at least the following programming languages: C/C++, C#, Fortran, Delphi/Pascal, Java and Visual Basic

The remainder of this section shows how these requirements have been met.

Use cases

To check that the requirements were correctly expressed and to ease the development of an architecture for the OpenMI, a range of scenarios or 'use cases' were identified. Some of the initial cases drawn the hydraulic domain are shown below. Since then, the use cases have been drawn from a much wider range of domains.

- Connect two 1D hydrodynamic river models.
- Connect a 1D hydrodynamic model with a water quality transport model.
- Connect a 1D river model with a 3D groundwater model.

- Connect a 1D hydrodynamic river model to vegetation and habitat models.
- Connect a 3D coastal model to a 1D river model.
- Connect a 2D polygon-based root zone model to a 3D regular grid groundwater model.
- Calibrate a rainfall runoff model linked to a hydrodynamic sewerage model.
- Model the propagation of uncertainty through a chain of models.
- Use different units of measurement for the data to be exchanged between models.
- Connect to an agent-based model.

This list has been further extended during the OpenMI-Life project, where it is being evaluated by the competent authorities in Belgium, the Netherlands and Greece against operational water management problems. In the US, it is being extended into a much wider computing environment.

Terminology

A number of terms are used when describing the OpenMI standard.

As shown in Figure 6, the term model application encompasses all parts of the modelling system software that is installed on a computer: for example Mike11, PHABSIM and InfoWorks-RS.

Typically, such systems consist of a user interface and an engine. Usually, the engine is a generic representation of a process and this is where the calculations for simulating or modelling that process take place. The user supplies information through the user interface and this is converted into the input data for the engine.

The data describes a specific scenario in which the process is to be simulated: for example the flow in the river Rhine during a time of extreme rainfall. The user runs the engine by selecting an option or pressing a button on the user interface. The engine reads the input, performs the calculations and outputs the results to files or displays.

When an engine has read its input it becomes a model. For example, an engine may represent the generic process of water flowing in an open channel. When it has read in the data describing the channel network of the Rhine, along with any boundary conditions and rainfall data, it becomes a model of the Rhine in the scenario to be simulated.

If the code for an engine can be instantiated separately and has a well-defined interface through which it can accept and provide data, then it is an engine component. (The engine's interface is the part of the code that handles the transfer of data to and from the engine; it should not be confused with the user interface, which is the part of the application that the user sees.) The key to enabling models to exchange data lies in standardizing the design of the engine interface. When an engine component implements such a standard interface, it becomes a linkable component. An engine that implements the OpenMI interface is called OpenMI compliant.

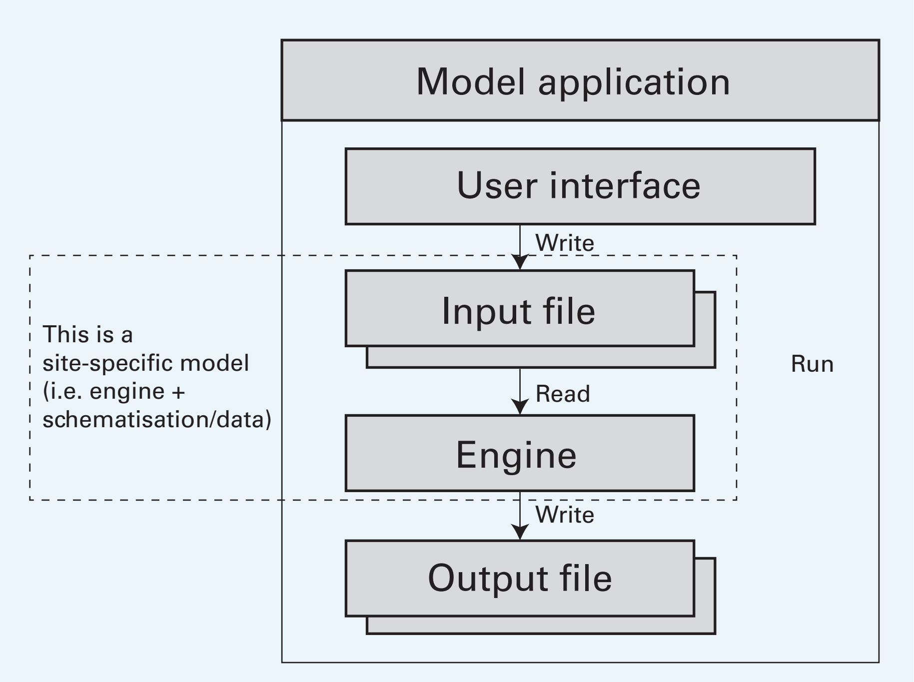


Figure 6: The general structure of a model application.

The OpenMI standard interface

The OpenMI defines a standard interface that has three functions:

- Model definition: To allow other linkable components to find out what items this model can exchange in terms of quantities simulated and the locations at which the quantities are simulated.
- Configuration: To define what will be exchanged when two models have been linked for a specific purpose.
- Run-time operation: To enable the model to accept or provide data at run time.

Figure 7 shows two model applications whose engines have been made OpenMI-compliant. Their overall structure remains unchanged but each engine is now a component with an OpenMI interface and one component can now get values from another.

Figure 8 illustrates some of the information held in the model definition about the quantities that two models can either accept or provide. The arrow

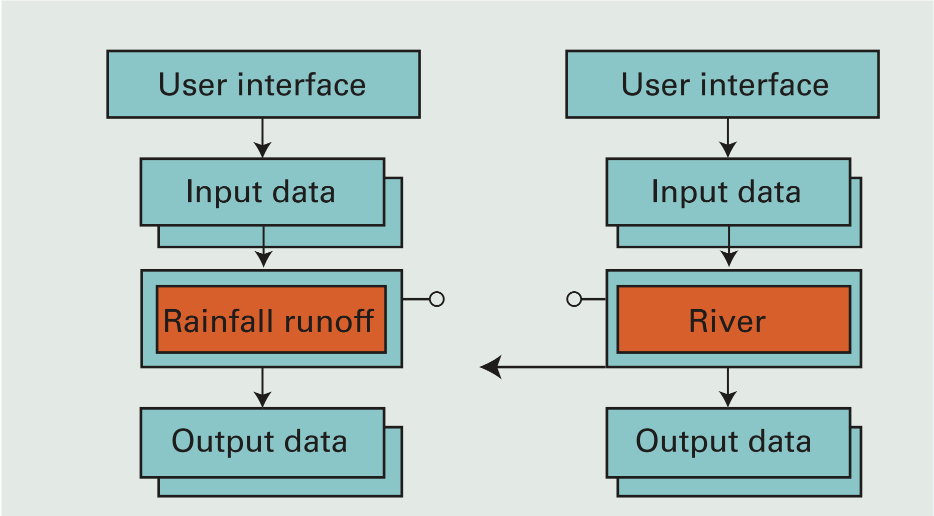
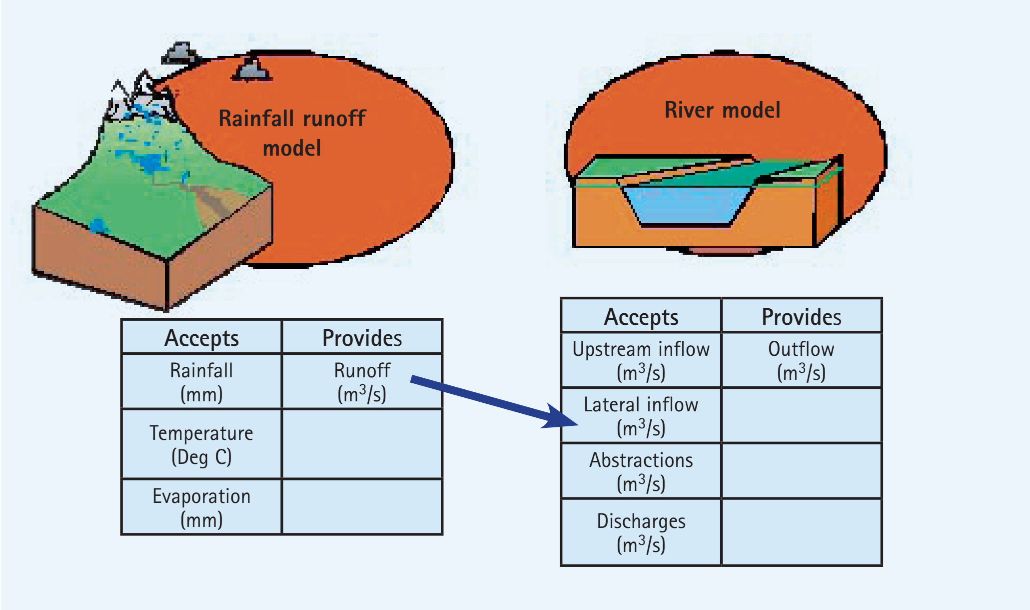


Figure 7: Two applications after migration to the OpenMI standard.

Figure 8: Showing and linking quantities.



represents a link between the two models and indicates that, in this particular case, runoff produced by the Rainfall Runoff Model will be used to represent lateral inflow in the River Model. There is no requirement to harmonize the terminology; the linking process creates the appropriate cross-reference table.

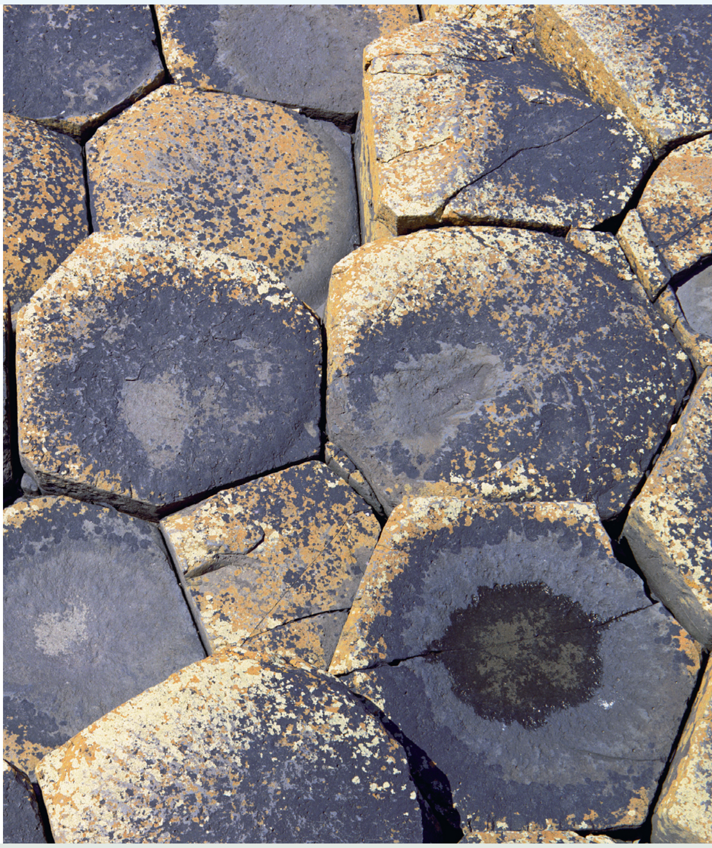


Figure 9 shows the geographical matching of elements in a river model to those in a groundwater model. The river model is a vector model and each element represents a single stretch; the groundwater model is grid-based, each node being an element. Therefore, in order to link the two models, each element in the river model will usually be linked to several elements in the groundwater model. In any non-trivial situation, this will require the matching of thousands of elements and therefore the process is automated.

An interface-based open standard

The OpenMI provides an intelligent mechanism whereby models running simultaneously can exchange data time step by time step. It thus enables process interaction to be represented more accurately than is possible by sequential linkage. It is important to explain that the OpenMI is neither a common data-model specification nor is it an integrated modelling system. It is a standard.

The OpenMI is 'interface-based':

- Its 'standardized' part is defined as a software interface specification.
- This interface acts as a 'contract' between software components.

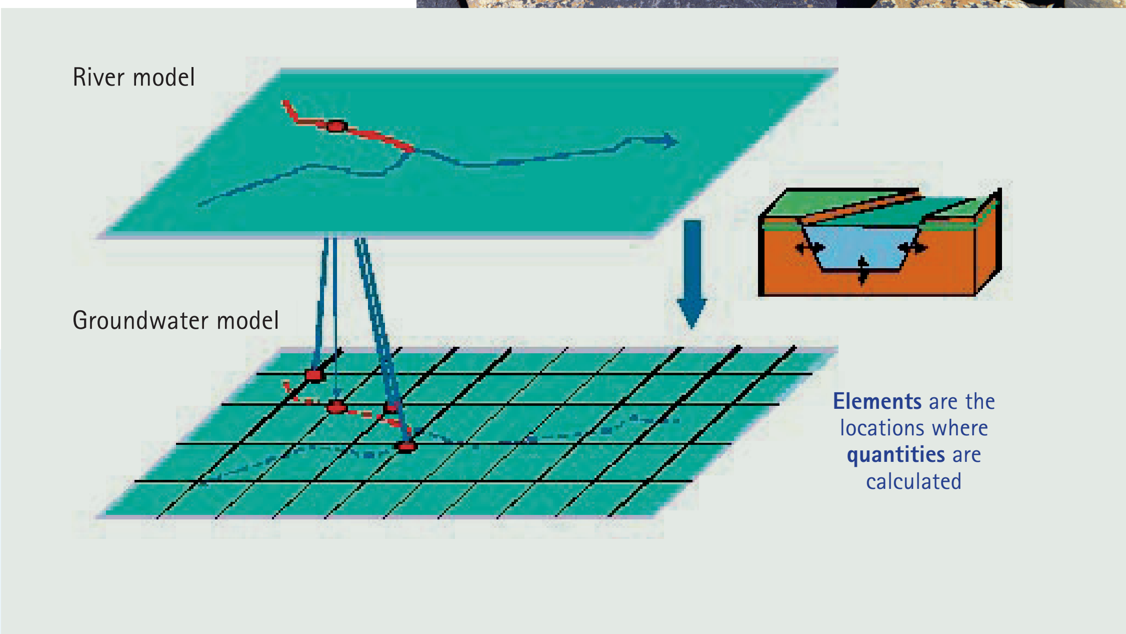


Figure 9: Linking element sets.

- The interface specification is not limited to specific technology platforms or implementations.
- The interface implementation may be limited by the technology supported in a specific release[26].
- By adopting the implemented interface a component becomes an OpenMI-compliant component.

The OpenMI is 'open':
- Its specification is publicly available via the Internet (www.OpenMI.org).
- Its source code is open and available under Lesser GPL licence conditions.
- It enables linkages between different kinds of models, different disciplines and different domains.
- It offers a complete metadata structure to describe the numerical data that can be exchanged in terms of semantics, units, dimensions, spatial and temporal representation and data operations.
- It provides a means to define exactly what is linked, how and when.
- Its default implementation and software utilities are available under an open source software license.

The OpenMI is a 'standard':
- It standardizes the way data transfer is specified and executed.
- It allows any model to talk to any other model (e.g. from a different developer) without the need for co-operation between model developers or close communication between integrators and model developers.
- Its generic nature does not limit itself to a specific domain in the water discipline or even in the environmental discipline.

The OpenMI cannot guarantee that the representation of the process in the component or the link to another component is scientifically valid. That is the responsibility of the modeller, model integrator and user as it should be.

[26] The technology chosen for Release 1.4 is version 2.0 of the .NET framework.

Appendix 6: Review of existing standards

8.1 Geoscience related XML

A range of XML standards have been developed, or are being developed, that are of relevance to the DREAM project. These are briefly described below.

GEOSCIML

The aim of the development of GeoSciML (GeoScience Mark up Language) is to satisfy 'the need of users to be able to access digital geoscience data from a variety of sources, process them using common software packages, and integrate them with data from other disciplines'. In order for there to be interchange of information there has to be agreement, on the nature and structure of the information to be interchanged. If all geoscience data providers stored their data in a common database structure this would be simple, at least within the geoscience domain, but this is far from the case. The solution is to agree a common conceptual data model, to which data held in our existing databases can be mapped. Such a data model needs to identify the objects being described (eg 'faults'), their properties (eg 'displacement') and the relations between objects (eg 'faults are a type of Geologic Structure'). Such a model can be described graphically using Universal Modeling Language (UML).

Having agreed a conceptual data model it needs to be mapped on to an interchange format. The GeoSciML application is a standards-based data format that provides a framework for application-neutral encoding of geoscience thematic data and related spatial data. GeoSciML is based on Geography Markup Language (GML—ISO DIS 19136) for representation of features and geometry, and the OGC Observations and Measurements standard for observational data. Geoscience-specific aspects of the schema are based on a conceptual model for geoscience concepts and include geologic unit, geologic structure, and Earth material from the North America Data Model (NADMC1, 2004), and borehole information from the eXploration and Mining Markup Language (XMML). Development of controlled vocabulary resources for specifying content to realize semantic data interoperability is underway. There are well defined rules for turning the UML conceptual model into an XML Schema which can be used to validate data files. Currently there is also a move to change the serialisation of vocabularies into SKOS to leverage the use of semantic web tools.

Intended uses are for data portals publishing data for customers in GeoSciML, for interchanging data between organizations that use different database implementations and software/systems environments, and in particular for use in geoscience web services. Thus, GeoSciML allows applications to utilize globally distributed geoscience data and information.

Developing such a conceptual data model is a major piece of work and version 2 the scope has been restricted to those geoscience objects which form the main components of a geological map (geological units, faults, contacts, and their defining concepts) as well as boreholes. This scope may be extended in future versions but there will also be cases where other groups are or have designed models and exchange formats for other geoscience domains and the intention is to try to work with these. GroundwaterML is an example of a derived implementation of GeoSciML. It is also the first official collaboration between GeoSciML and an external exchange model group.

MineralOccurrences is an example of an inherited implementation of GeoSciML. It is being developed by the Australian Government Geologists Information Committee (GGIC) as a model to deliver mineral occurrences information as WMS/WFS. Australian State, territory and federal organizations presently govern the model.

Some software vendors have started to support GeoSciML but this isn't extensive and there is still potential for reasonably significant change with future versions. The design methodology is essentially that being adopted for the INSPIRE initiative so this should increase the chances of some interoperability with spatial data for other domains but it's a slow moving complex process. The format is highly self-descriptive and verbose (although also very compressible) which makes it suitable for heterogeneous environments and unforseen useages but not for 'high' data volumes.

WITSML

The Wellsite Information Transfer Standard Markup Language (http://www.witsml.org) is a standard for transfering realtime drilling data between oil and gas drilling service companies and operators. As such it covers a lot of details that are probably not of interest scientifically and some, like boreholes shapes and maybe some of the logging data that a scientist might want to extract.

WATERML

WaterML is a dedicated markup language for water observation data. It has been developed by CUAHSI to enable the ease of exchange of water data for the Hydrologic Information System (his.cuahsi.org) project. This project aims to have a seamless exchange data between the main US water databases, e.g. USGS' NWIS and any potential user. WaterML provides a schema for water observations such as level, flow and water quality at a single geographic location. It has been submitted to the OGC for adoption as an internationally recognised standard.

DE FACTO STANDARDS

There are many more de facto standards in the geospatial domain, the most commonly used standard for interchange of geoscience models are in the form of tab separated ASCII files of ASCII grids.

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

British Geological Survey
Kingsley Dunham Centre,
Keyworth, Nottingham NG12 5GG
Tel 0115 936 3241
Fax 0115 936 3488
email sales@bgs.ac.uk