

# Dataset Acquisition to Support Geoscience

J. R. A. Giles, S. H. Marsh and B. Napier

ERMS ID	D23155
ERMS Title	Dataset Acquisition to Support Geoscience GILES & MARSH
ERMS Version Number	8
Last Saved	20/08/2008 12:55
Word Count	3532

**Abstract:** Environmental scientists are both producers and consumers of data. Numerous studies have shown that significant amounts of scientists' time can be consumed in acquiring, managing and transforming data prior to its use. To facilitate the work of its scientists the British Geological Survey (BGS) has identified a series of national datasets that are required by scientists across the organisation. BGS then seeks to acquire and manage these centrally, and supply them to the scientists in formats that they normally use. Making these datasets readily available helps to:

- Enhance the quality of the science;
- Promote interdisciplinary working; and
- Reduce costs.

The strategy has also enabled the development of advanced, domain-specific visualization tools, which have significantly improved the scientific output while also reducing costs.

## Introduction

A modern geological survey organisation (GSO) such as the British Geological Survey (BGS), requires a wide range of digital and analogue datasets to support the activities of the scientists employed to fulfil its mission. For many years GSOs primarily used datasets that they had compiled internally. Typically, databases would be designed in-house to meet the needs of specific geoscience communities within the GSO. An example of this is provided by geochemists, who analyse stream sediments for a specified suite of chemical elements and visualise the data spatially to understand the distribution of those elements (Johnson *et al.* 2005). In addition to in-house datasets the legislation in some countries provided GSOs with ready access to specified datasets produced by industry. For example, - Geoscience Australia ([www.ga.gov.au](http://www.ga.gov.au)) houses one of the worlds largest collection of petroleum data in its Petroleum Data Repository. This is accessible internally within the GSO and much of the dataset is "Open file" and available on the Internet through The Petroleum Information Management System ([www.ga.gov.au/oracle/npd/](http://www.ga.gov.au/oracle/npd/)).

During the past decade a range of public and private sector organisations has been creating national digital datasets to meet the needs of a variety of customers. Some of these datasets have direct relevance to geoscience and can be readily integrated with the in-house digital datasets that GSOs typically maintain. For example INTERMAP™ Technologies ([www.intermap.com](http://www.intermap.com)) have created a range of regional and national datasets using airborne Interferometric Synthetic Aperture Radar (IFSAR). Under the brand name NEXTMap® Intermap has produced digital surface models and digital terrain models for Britain, large parts of Europe and several states in the USA.

The process that the BGS undertook to identify and acquire specific digital datasets to support the work of its geoscientists is described below.

## Datasets

In 1999 the BGS-geoIDS (BGS Geoscience Integrated Database System) and the SIGMA (System for Integrated Geological Mapping) projects within the BGS established a team to identify the national third-party digital spatial datasets that the BGS was likely to need during the next decade. The team was also tasked to identifier potential suppliers and put in place activities to acquire the datasets under appropriate terms and conditions. The team identified the following digital datasets:

- Elevation
  - Terrain and Elevation Models
  - Topographical Survey Elevation
- Imagery
  - Aerial images
  - Satellite images
- Topography
  - Modern topography
  - Historical topography of various ages

## Elevation

Elevation data are a fundamental tool in any geologist's toolkit; all the more so in Great Britain where there is little exposure of the underlying rock over large parts of the country. Commonly, the geologist must infer what lies at depth from the surface expression of the lithology and structure through their effects upon the topography. In the field, BGS geologists have done this for many years by employing a technique known as feature mapping. This consists of careful surveying of subtle breaks of slope, which can then be related to particular lithologies under the superficial cover. A classic example comes from the Chalk of Southern England (**Fig X**), where up to 12 distinct units can now be recognised by the effects that they have on the surface topography. Discontinuities in these surface features can commonly be related to faulting and other expressions of the underlying geological structure. These features are characterised by changes in elevation that can be seen in digital elevation data just as they can in the field. The data may be generated by digitising contours acquired during topographic survey; from aerial photography, either as a photogrammetric source for these contours or via the photogrammetric generation of a digital elevation model; or from direct measurement techniques, such as spaceborne, airborne or terrestrial radar or lidar sensors. **Table X** details the various elevation data used in the UK by the BGS over the past decade.

Depending on the technique used and the desired application, such data may represent the bare earth (a digital terrain model) or include the elevation of surface features such as vegetation (a digital elevation model or digital surface model). Geologists usually prefer to analyse the former, although the latter might give a more realistic visualisation of landscape when used in combination with aerial photography.

Geologists increasingly use elevation data, how ever it has been acquired, to help them accelerate feature mapping. This has several advantages. It can be done at the desktop and then checked in the field. The data are GIS-ready and their interpretation can be carried out within the standard GIS and digital mapping packages. The sun angle and topographic exaggeration can be adjusted to emphasise subtle features. In fact, several other advanced image analysis techniques have been developed that can at least partially automate the process of interpretation (**Fig X**). Elevation data are also used as the base on which other datasets, such as aerial photography, are visualised in their landscape context as part of the digital mapping process. BGS has purchased national coverage of the NEXTMap<sup>®</sup> dataset from Intermap as a BGS Baseline Dataset.

## Terrain and Elevation Models

Geologists are most often interested in terrain, or bare earth, models, because these allow them to analyse the shape of the Earth's surface rather than those things growing or built upon it. A component of the NEXTMap<sup>®</sup> Britain dataset includes a bare earth model, but this has been generated from the original elevation model that was measured during the survey. This process involves editing the elevation data to remove features such as trees and buildings. Doing this for the entire country requires automation and this is a considerable technical challenge. Consequently, the resulting bare earth model contains residual artefacts related to the features removed. **Figure X** shows a comparison between the terrain and elevation models to illustrate this. Small forest stands can confuse subsequent analysis techniques such as slope angle, giving apparent steep slopes at their margins. Such errors propagate through into derived products such as landslide hazard maps, giving erroneously high hazard values around forests. The BGS has attempted further editing of the NEXTMap<sup>®</sup> Britain dataset, using satellite imagery and photography to map the spatial distribution of vegetation on a national scale and highlight potential problem areas for further editing. This artefact editing problem also affects elevation data extracted from satellite radar interferometry, which works on a similar basis to that from an aircraft, and stereo aerial photography and satellite imagery like stereo ASTER, SPOT or ICONOS data.

There are two other ways to tackle the problem. The first is to use nationally available contour data. National Mapping Agencies commonly use digital photogrammetry to generate contours from stereo aerial photography, as part of their topographic surveying process. This involves the creation of a stereo model that contains an elevation model from the photography, from which contours can be digitised by a skilled analyst, who can place the cursor onto the ground even amongst tree stands. Unfortunately, it has been common practice not to extract and store the elevation model once the contours have been generated, otherwise this would be a valuable source of national elevation data. Instead, the contours can be used to work back to the elevation model by gridding and interpolation. As the contour data have already been generated in a way that avoids the recording of unwanted surface features, such datasets circumvent the artefact problem efficiently. However, such models tend to lack topographic detail due to the degree of interpolation employed, especially in relatively flat ground, where contours can be both poorly constrained and sparse.

The second approach is to use a data acquisition technique that measures the ground surface directly, penetrating the tree canopy. Airborne lidar data have been used effectively for this purpose, even in rainforest areas. The laser in the aircraft is pulsed, so that every point on the ground has multiple measurements, at least some of which penetrate through the canopy and reflect off the land surface. Using such data, it is possible to extract the bare earth model from the last return, the structure of the tree canopy from the intermediate returns and the top of canopy elevation model from the first return within the same dataset. Lidar data are also high resolution in x, y and z, commonly in the range of centimetres rather than metres, and they make a highly suitable dataset for geological terrain analysis. Their only drawback is that they are not yet available on a national basis; as ad-hoc acquisition continues, the coverage in the UK is increasing, to the point where national coverage has become possible to contemplate. Initiatives to pursue this are under discussions and it is likely to occur in the near future.

## Imagery

Aerial photography has long been used by geologists to visualise the landscape in three dimensions, both in the laboratory using stereoscopes and in the field using field pocket stereoscopes or portable stereo glasses. Before the advent of elevation data, this was the main way in which the topography was visualised so that feature mapping could be undertaken. In addition, imagery gives useful clues about lithology and soil type through the colour and texture that could be seen and associated with particular rock types. The patterns made by streams also vary depending upon the lithology and these patterns, together with textures, provide clues about jointing and fracturing. Major topographic lineaments that persist over kilometres are commonly associated with significant faulting or other geological structures.

The advances in computer processing power and storage in recent years have revolutionised the use of such data in geological mapping, and digital imagery is now a key dataset in the digital mapping workflow. The change started with the advent of satellite imagery in the 1970s and Landsat has become an important reconnaissance tool for geologists, which still has its place today, particularly in poorly mapped, well-exposed terrain (**Fig X**). But in the UK it was the widespread availability of digital stereo aerial photography that began to see imagery take its place in the digital workflow. The geological survey has invested in national coverage from UK Perspectives and Getmapping, including monoscopic orthophotography for draping on elevation data and full stereo photography for more advanced analysis. This aerial photography underpins many mapping projects. Geologists interpret the landscape before going in the field and use the field time for checking and to investigate challenging or interesting areas. This generates both an economic and scientific return on the investment in the data. The national Landsat and aerial photography coverage form two further BGS Baseline Datasets.

## Topography

Topography is the essential spatial back drop to GSOs outputs. Geological maps are ultimately of little use unless they relate to an underpinning topography. It is also important in data acquisition as individual observations are referenced by the spatial framework provided by topographic mapping. The relationship between geological mapping and the topographic survey is so fundamental that the founder of the

Geological Survey in Great Britain, Sir Henry De la Beche, was originally funded in 1832 by the Board of the Ordnance “*to cover the cost of geologically colouring the topographical maps of the Trigonometrical Survey*”.

### **Modern Topography**

In Great Britain (GB) the principal supplier of modern topographic mapping is the Ordnance Survey (OS). The OS produces a wide range of mapping outputs at a range of scales down to 1:1250. The most important scales for most BGS geological mapping are 1:50 000 and 1:10 000. These scales are readily available in raster format, which is ideal for back drops and locating observations using global positioning systems. The MasterMap product provides mapping at scales to 1:1250. It is used occasionally for specific tasks but the cost of national coverage prevents its routine use by the BGS.

### **Historical Topographical Mapping**

In Great Britain the Ordnance Survey was founded in 1747 and it has been publishing maps since that date. By the 1840s systematic surveying had commenced at 1:10560-scale for Great Britain. Many areas have been resurveyed repeatedly, producing a series of editions showing the changing landscape. This serial snapshot of the landscape of Great Britain provides valuable information to geoscientists. Information about the location of mine entrances, quarries and other excavations can be derived from historical topographical mapping. They also provide valuable information about anthropogenic landscaping. For example, consider a late 19<sup>th</sup> century mine that produced a waste heap that was subsequently modified after closure of the mine, with associated tree planting to stabilise the slopes. Planners of subsequent developments need to be aware that the visible wooded hill is in fact composed of mine waste, with potential slope stability issues and the possibility that it contains toxic minerals or high-acidity materials and drainage.

## **Management and Delivery of Datasets**

Once the datasets have been acquired they need to be managed rigorously within a controlled environment. The principal issues are:

- Licensing and intellectual property rights
- Long-term storage and digital preservation;
- Datasets limitations; and
- Communication.

### **Licensing and Intellectual Property Rights**

Acquired datasets commonly come with complex licensing agreements. It is essential that the terms and conditions of licences are understood and communicated to data users in a way that they can understand. Internal monitoring systems are required to ensure that research outputs, derived datasets and information products do not infringe third-party intellectual property rights. In the most difficult cases licensing agreements for a given product vary over time as the policies of the supplier change. This means that a dataset used in the creation of a specific research output might no longer be available for its continued use and exploitation, so the research output must be withdrawn from use. BGS has found it simplest to attempt to acquire in-perpetuity

licences for a given dataset in exchange for a one-off payment. This simplifies licence management and means that research outputs are more long-lived and robust.

## **Long-term Storage and Digital Preservation**

National digital datasets can be large, comprising multi-terabytes of data. When such data are being acquired it is essential that the related storage issues are considered during the acquisition process. Does the organisation have the storage and compute power to manipulate the datasets? Digital data formats change over relatively short timescales. Plans need to be in place to ensure that the dataset continues to be available, even if the original delivery format becomes obsolete. This is not simply a process of progressively migrating datasets to the current appropriate file format. The organisation must have a clear understanding of any information losses that might take place during progressive file format changes.

## **Dataset Limitations**

No dataset is perfect. Each has its own limitations. Nevertheless, the temptation is to assume that digital datasets are perfect. Tarter (1992) has noted that "...the myth of machine infallibility seems to create a demand for higher standards of quality for machine readable data than for traditionally distributed information." Similarly Peritz (1986) has suggested that "...the presumption of trustworthiness (of digital data) simply carries too much weight..." The reality is that data are not perfect, and dataset limitations need to be understood and documented. The aim of the documentation is to ensure that a potential user can assess whether a given dataset is fit for its intended purpose. For example, there is a predisposition to assume that digital raster images of historical Ordnance Survey 1:10,560 topographical maps have a similar accuracy to their modern 1:10,000 counterparts. However, there are considerable differences:

- 1) The original 1:10,560 maps were paper prints and they are up to 150 years old. Unless stored in a perfect records management environment for their entire life, such paper maps can become distorted to varying degrees over time. The maps might therefore be spatially inaccurate before scanning.
- 2) Different generations of maps will have been surveyed using different methods and/or instruments. The same geographical object will not necessarily be in the same spatial location on subsequent editions.
- 3) The Ordnance Survey 1:10,560 maps are maps, not plans. They include cartographic generalisations that affect the spatial representation and location of geographical objects.
- 4) Geographical objects change over 100 years. Buildings are extended or demolished and rebuilt, changing their footprint in the process. Bridges and roads may be widened.

The cumulative result is that, potentially, a location determined from an historical map might be tens of metres from its correct location.

Other dataset limitations arise from the nature of the instrument collecting the data and the platform upon which the instrument was mounted. Where the platform is an aircraft and the instrument is IFSAR, a number of 'IFSAR Artefacts' may be identified in the final dataset. Several known artefacts can persist despite processing of the data during and after acquisition. One type of artefact is 'Motion Ripples', which are caused by atmospheric turbulence preventing the aircraft from maintaining

level flight during data acquisition. They appear as height ripples in the elevation data and as dark bands in the imagery at rightangles to the direction of the aircraft's flight path. Most motion ripples are eliminated by processing, but some might persist into the final dataset (Intermap Technologies 2007). The other issues primarily relate to the recorded height coming from the top of the tree canopy and from buildings in built-up areas; research is underway to improve the removal of such artefacts.

## **Communication**

The principal way to communicate information about a dataset is through its associated metadata. A rich, well maintained metadata entry can enhance user-understanding of the dataset. It is the dataset custodian's responsibility to develop and maintain metadata to meet the needs of users seeking to re-use and re-purpose the data. The profile and significance of metadata have risen in many countries in recent years following the introduction of new laws, including data protection and freedom of information legislation. Within the European Union (EU), directives have been issued relating to the use of public sector information and a common spatial infrastructure. These are being transposed into national laws by EU member states. All this legislation requires or implies that accurate, well maintained metadata are in place.

## **The Sum is Greater than the Parts**

One of the underlying motivations for the selection of the BGS national baseline datasets was the synergy between them. Elevation data can be analysed in their own right, but come to life when they are used as the backdrop for aerial imagery. Such imagery provides a unique view of the Earth's surface from above, but is far easier to interpret when draped over an elevation dataset to create a virtual, immersive environment than it is when using traditional stereo analysis techniques. This synergy extends the utility of the data from the specialist analyst to geologists (and indeed other scientists, professionals and the public) in general. Other datasets can be visualised far more clearly when viewed in their landscape context within this virtual environment.

Another synergy between elevation data and imagery or other raster datasets involves the generation of synthetic stereo from ortho-corrected mono imagery. This approach allows other digital geoscience datasets, such as the BGS national digital geology and geochemical or geophysical datasets, to be viewed in stereo. Viewing geological map data in a new perspective against the actual 3D model highlights any inconsistencies within the conceptual 3D model that underlay the original mapping, and allows for rapid correction or the targeting of field surveys to update maps in problem areas.

Elevation data have many applications in their own right, but really come into play as an underpinning dataset supporting the processing, display, interrogation and analysis of other geoscience data. This can include: their use to orthorectify other remotely sensed imagery; as a base on which to display 2D geological maps in 2.5 or 3D; or more complex algorithms within a model or GIS that take elevation, or a derivative like slope, as one input. The integration of elevation data into other geoscience workflows delivers substantial synergy. Satellite imagery can be placed in their real-world position and features extracted that can go straight into a GIS. Geological lines drawn in a pre-digital era can be seen in 3D and obvious errors corrected, releasing

the potential of older datasets. Complex problems that may require a wide range of input parameters often have elevation, or a derivative, as a common thread. The latter point was illustrated during the first half of this decade when, one after the other, a series of Integrated Global Observing Strategies identified improved global DEMs as a high priority. They covered not only the geosciences (Marsh (ed), 2004) but also coastal observations, the water cycle, land observations and the cryosphere. Elevation data are one dataset that can pay dividends right across the environmental sciences.

## **Applications**

The applications of elevation data to the geosciences are many and varied. In the foundations of our science, the basic geological mapping requires the topography to be mapped first, because the geology exerts control on the overlying landscape and an understanding of the latter helps reveal the former. In countries like the UK, large areas must be mapped without seeing a single exposure and the established technique, feature mapping, is essentially a detailed topographic analysis. Digital topographic data have transformed this process in the past decade, allowing much of the critical information to be captured in the office. Fieldwork then focuses on the challenges and areas of particular geological interest. This has allowed BGS to re-engineer its mapping process into a fully integrated, digital workflow, with consequent efficiency gains, standardisation of approach and related cost savings. Beyond this, elevation data have application in many other geoscience disciplines, at the least as a backdrop for other data. An important application area is geohazards, where slope and aspect are one of the key controls on ground instability. In pollution studies, the source-pathway – receptor model relies on topography to help determine the pathway and likely area for receptors. Applications exist in mineral, especially aggregate, resources and groundwater management. In fact, it is hard to think of a geoscience discipline where these data do not apply.

### **Case Study – Virtual Field Reconnaissance using GeoVisionary**

BGS geoscientists have routinely used digital elevation models as part of their work for many years, but accessing and visualising the data was often time-consuming and restricted by technology limitations to either small areas or low-resolution representations. In order to make full use of BGS's new high-resolution baseline datasets (principally NextMap Britain 5m DTM and DSM from Intermap Technologies, and aerial photography from UKP/Getmapping) a project was started in late 2006 - Virtual Field Reconnaissance – that aimed to create an environment in which to visualise and interact with all of these data. The project built on BGS links with Virtualis Ltd., a British-based virtual reality company that had previously been commissioned to install an immersive 3-Dimensional Visualisation Facility at two BGS sites and produce custom geological visualisation software.

The result of this collaboration is GeoVisionary, a software system that has built-in seamless, streaming of multi-resolution levels of data, merging data such as existing digital geological maps, aerial photography, satellite imagery, field-slips, historical topographic maps, and subsurface 3D models, cross-sections and boreholes. The system allows teams of geologists to undertake a virtual survey of an area in the office



before commencing fieldwork, building an understanding of the terrain, which leads to a better interpretation of the geological structure. This initial assessment allows surveyors to effectively target fieldwork in areas where surveying is most required. On completion of fieldwork, surveyors can check their individual field interpretations together in the virtual landscape. This team approach allows colleagues to work together on both pre- and post-fieldwork studies, better enabling communication, so increasing operational efficiency and enhancing scientific understanding. Whilst a variety of data can be visualised in the system, the elevation data are fundamental to its successful operation; they provide geoscience information in their own right but are also the backcloth against which other data are displayed. They also provide the top surface from which the subsurface models are generated and hung.

## Conclusions

The increasing variety and improved availability of national digital datasets are helping to provide exciting new tools for geoscientists. These data products, used in combination with innovative software, provide new ways for geoscientists to perceive and interpret the landscape. Elevation data in particular are central to this because they can both be interpreted as a geoscience dataset in their own right and used to both process and visualise other datasets to best advantage.

Benefits from these datasets and systems include integrated digital workflows and shared 3D models of the surface and subsurface geology, This has led to efficiency gains and cost savings in mapping programmes, whilst at the same time improving teamwork and standardising approaches to mapping. However, for the potential scientific benefits of this technological development to be achieved, a range of non-scientific issues must be addressed. These include:

- Licensing and Intellectual Property Rights management;
- Management of the digital datasets;
- Digital preservation of the dataset;
- Understanding the limitations of the dataset;
- Communication of the above to geoscientists.

It is incumbent upon GSOs to make sure that they create and maintain efficient and well-supported information management systems to deal with these issues.

## References

Intermap Technologies 2007. Product Handbook & Quick Start Guide Standard Edition 4.2, Denver USA

Johnson, C. C; Breward, N; Ander, E. L. Ault, L. 2005. GBASE : baseline geochemical mapping of Great Britain and Northern Ireland : In: Geochemistry: exploration, environment, analysis Vol. 5 pt/no 4 p. 347-357

Peritz, R., 1986, Computer Data and Reliability: North west University Law Review, v 80, p. 960.

Tarter, B., 1992, Information Liability: New Interpretations for the Electronic Age: Computer/Law Journal, v. XI, p. 484.