GRADE

Scoping a Geospatial Repository for Academic Deposit and Extraction

JISC DEVELOPMENT PROGRAMMES

Project

Project Acronym	GRADE	**Project ID**	
Project Title	Scoping a Geospatial Repository for Academic Deposit and Extraction		
Start Date	1 June 2005	**End Date**	19 April 2007
Lead Institution	EDINA		
Project Director	David Medyckyj-Scott, EDINA		
Project Manager & contact details	Anne Robertson, EDINA University of Edinburgh Data Library, Main Library Building, George Square, Edinburgh EH8 9LJ Tel 0131 651 3874 email a.m.robertson@ed.ac.uk		
Partner Institutions	University of Edinburgh and University of Southampton		
Project Web URL	www.edina.ac.uk/projects/grade		
Programme Name (and number)	*Digital Repositories*		
Programme Manager	Neil Jacobs		

Document

Document Title	Designing a licensing strategy for sharing and re-use of geospatial data in the academic sector.		
Reporting Period	*1st Feb – April 2007*		
Author(s) & project role	Charlotte Waelde AHRC Research Centre for Studies in Intellectual Property and Technology Law, School of Law, University of Edinburgh Mags McGinley, Digital Curation Centre, University of Edinburgh		
Date	March 2007	**Filename**	GRADE Digital Rights Issues WP 3.doc
URL	*http://edina.ac.uk/projects/grade/*		
Access	☐ Project and JISC internal		√☐ General dissemination

Document History

Version	Date	Comments
Final	March 2007	Comments incorporated throughout 2006

The GRADE Project (http://edina.ed.ac.uk/projects/grade) is one of a cluster of projects in the Digital Repositories Programme funded by the Joint Information Services Committee (www.jisc.ac.uk) of HEFCE investigating the interactions between data and institutional (publications) repositories, support for scientific lifecycle, storage and access requirements.

The JISC is bringing together a programme of work relating to digital repositories. Its aim is to bring together people and practices from across various domains (research, learning, information services, institutional policy, management and administration, records management, and so on) to ensure the maximum degree of coordination in the development of digital repositories, in terms of their technical and social (including business) aspects.

Within this context, GRADE is investigating the technical and cultural issues around the reuse of geospatial data within the JISC IE in the context of media –centric, informal and institutional repositories.

GRADE Work Package 3 aims to develop a clear understanding of digital rights issues for created geospatial data respecting, where applicable, the licensing conditions of any source geospatial data and to develop a conceptual framework for resolving those described rights management issues raised in relation to repositories.

Geospatial material created in the education sector can be highly complex, incorporating data created elsewhere either as found, or customised to fit the particular need of the academic or lecturer. The downstream rights can become very complex, as it is necessary to ensure that permissions have been gained to reuse or repurpose the data, and it is usually essential that correct attribution is made. There are currently concerns and confusion over the assertion of IPR and copyright of created geospatial data particularly where third party data are included.

This report considers a licensing strategy for the sharing and re-use of geospatial data within the UK research and education sector.

Geospatial databases and the research and education sector in the UK
Designing a licensing strategy for sharing and re-use of data.

Introduction

Geographic information has been said to mean *'information that can be related to a location defined in terms of point, area, volume on or of the earth, at a specific point in time, particularly information on natural phenomena, cultural and human resources'*[1]. The definition of geospatial information (or data) is broader and is said to be '*more precise in many … contexts than 'geographic,' because geospatial information is often used in ways that do not involve a graphic representation, or map, of the information'*.[2] The term 'geospatial data' will be used in this report.

The purpose of this report is to consider the acquisition, manipulation and use of geospatial data within the research and education sector in the UK (Higher and Further Education (HFE)) and to analyse which intellectual property rights might reside in a database containing geospatial data. The intention is to suggest what needs to be taken into account in constructing a licensing framework for geospatial databases where the content is used by researchers and teachers within the HE Community for non-commercial research and teaching purposes.

As will be described below, there are a variety of ways in which geospatial data may be collected. The focus of this report is on the rights subsisting in data collated by way of the Global Positioning System as supplemented by data collected using field survey techniques.

The stages in collecting, processing and use of geospatial data in HFE.

There are three distinct stages within HFE where data are used for non-commercial research and teaching: collection, processing and use of geospatial data.

Stage 1 – Collection

1 www.gipanel.org.uk/gipanel/docs/GIPanelMinutesSept05approved.pdf

2 www.opengeospatial.org/resources/?page=glossary#G

Developing the geospatial database

An organisation may obtain geospatial data from a variety of sources.

1. Survey

Surveyors and researchers gather data using field survey techniques. Data is captured either by traditional pen and paper or more commonly with an electronic distance measurer (Figure 1).

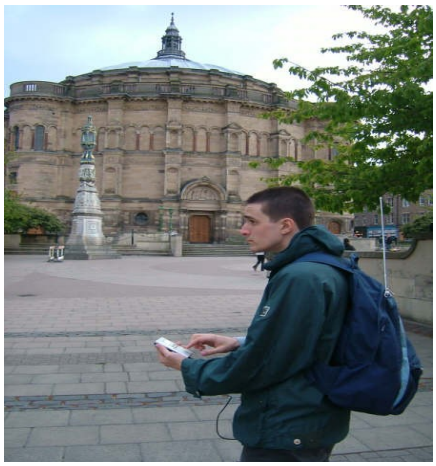


Figure 1.

2. Digitising

Existing maps may also be digitised. Digitising captures coordinates through the use of an electronic cursor. The data are captured when the cursor moves over a particular point of the map. The data captured using this process are updated by surveying and other digital capture techniques.

3. Photogrammetry

Photogrammetry is a process which enables objects to be measured and/or digitally captured from photographs. Aerial photographs are one type of source data, although this is commonly supplemented with satellite imagery. Data are then captured through the process of digitising as described above. Satellite imagery involves the capture of data from many wavelengths of electromagnetic radiation (e.g. infra-red).

4. GPS

More recently still, mechanisms for satellite based positioning and navigation (commonly referred to as the Global Positioning System or GPS) are used to capture geospatial data.

This is generated by satellites orbiting the earth producing signals captured by GPS receivers. Measurements are captured in a global reference system, World Geodetic System 1984 (WGS84), which can in turn be converted to local coordinate systems (e.g. OSGB36), to ensure compatibility with local geospatial data.[3]

Stage 2 – Processing

EDINA National Datacentre (EDINA)

Much geospatial data used in HFE is licensed from Ordnance Survey (OS), an organisation which has been providing mapping services for over 200 years, and which is now a Government trading fund.[4] When this is the case, then EDINA acts as the hub to which data are first licensed. These are then used within the research, teaching and learning community. The relationship between OS and HFE is managed through a matrix of licence agreements. The Secretary of State (acting through Ordnance Survey) contracts with HEFCE (acting on behalf of the funding bodies). HEFCE appoints EDINA to supply the service. Other education institutions who wish to subscribe to the service to enable their researchers to gain access to OS data must enter into a sub-licence with HEFCE. End users (the researchers and teachers) who wish to use the data within those Institutions gain access to the data held by EDINA via Athens authentication. The researchers and teachers must also enter into an End-User Agreement.

There are many other suppliers of geospatial data to HFE. Where these are supported centrally by EDINA (such as British Geological Survey, SeaZone and Landmark) then similar agreements are negotiated. Educational Institutions are free to enter their own arrangements with commercial geospatial data suppliers (such as LaserScan, Environment Agency, Forestry Commission, Dotted Eyes, XYZ Mapping) in which case the terms of the agreements will be determined as between the Institution and the data provider.

[3] It is understood that the majority of data used to update and supplement the Ordnance Survey compilation of geospatial data now comes from surveyors working in the field and from the GPS system (it is said that an average of 5,000 changes are made every working day to the OS large-scale map data of Great Britain) www.freeourdata.org.uk/ordnancereply.php

[4] For details see http://www.ordnancesurvey.co.uk/oswebsite/

In carrying out their part of the agreement with OS, EDINA goes through a number of complex steps in manipulating the data – a process which adds presentational value. So, for example, when EDINA receives data from OS for one of its products (Landline) it looks like the file represented in Figure 2

```
2100000010001020875018505 0%
1100018310061ORegistrars 6040406 0000000%
1200010123000000%
2100000010001020755016040 0%
1100018410060600ffice 0860415 0000000%
1200010125000000%
2100000010001020670015945 0%
1100018510060ZPH 0860415 0000000%
1200010143000000%
2100000010001021900013570 0%
1100018610001ZTEVIOT PLACE 0860415 0000000%
1200010144017100%
2100000010001025985013670 0%
1100018710061lMcEwan Hall 0860415 0000000%
1200010163000000%
2100000010001029690011865 0%
1100018810090ZTerrace 5021008 0000000%
1200010125012600%
2100000010001035160014750 0%
1100018910000GBristo 5040408 0000000%
1200010143000000%
2100000010001035880013345 0%
1100019010000GSquare 5040408 0000000%
1200010145000000%
2100000010001035880013215 0%
1100019110090ZTerrace 3041005 0000000%
```

Figure 2

Having been loaded onto EDINA servers, the data file is passed through a translator into a GIS database (Figure 3)

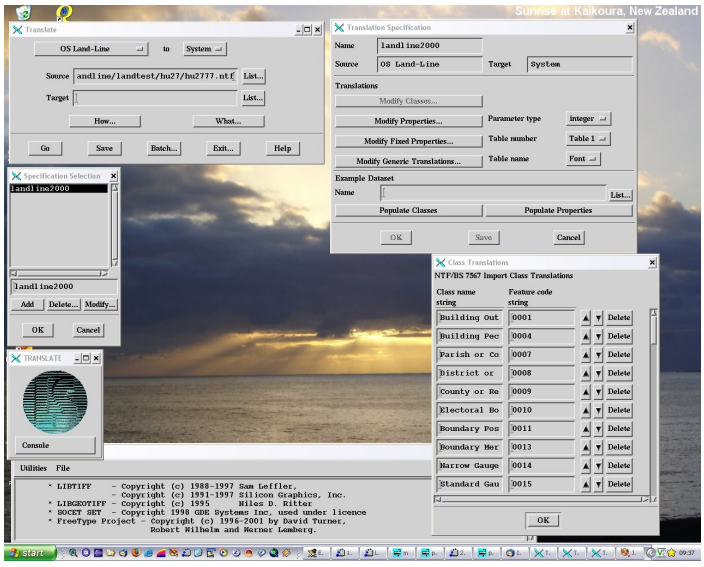


Figure 3

Both automatic and human verification takes place for accuracy (automatic - through the running of the program; human - through brief visual checking of the data). EDINA then creates style rules for symbolisation of the map features prior to the end user 'seeing' the final product on screen for example as depicted in Figure 4.

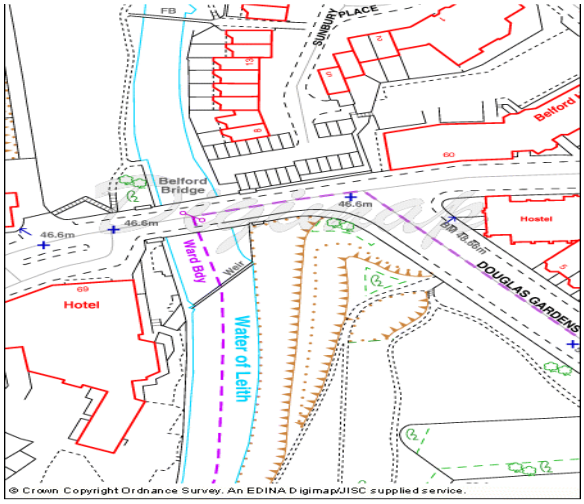


Figure 4.

Stage 3 – Use

The Research and Teaching Process

The third stage is when the data, some of which having been manipulated by EDINA, are accessed and used by the researcher or teacher (the end user). In this process the end user may have a particular goal in mind and manipulate and combine the data with other sources. Or the end user may simply wish to browse the available data and make it available during teaching. In the example below, the end user wished to generate geomorphological field maps and did so using data derived from a variety of sources including OS, Ordnance Survey of Ireland and NASA. Figure 5 depicts one of the stages (interim stage) on the way to producing the final output shown in Figure 6.

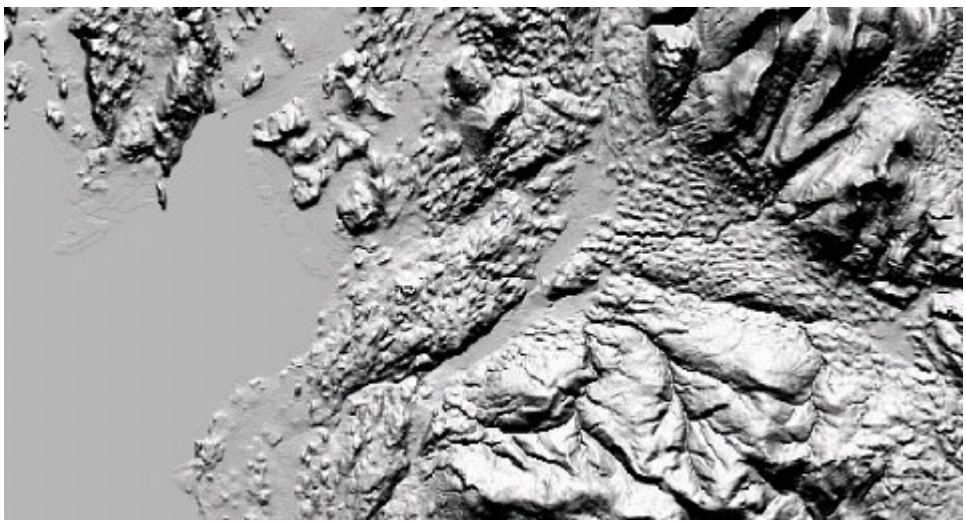


Figure 5 interim stage

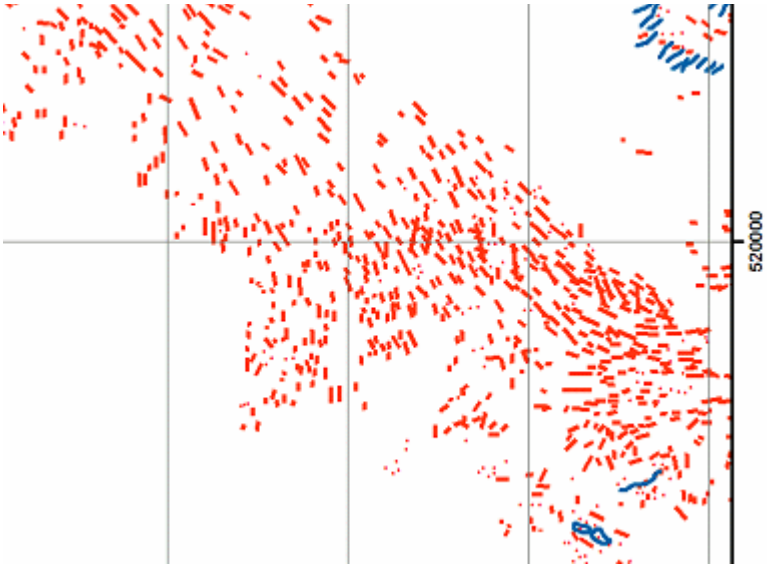


Figure 6 - Final output

A key point is that it is the data that are licensed to HFE. These data are then manipulated by EDINA and/or the end user in accordance with the research or teaching objective.

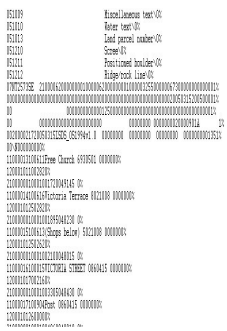
Rights in the geospatial database

Since the introduction of the Database Directive in 1996, there are (for our purposes) two rights that may subsist in a collection of geospatial data: a. copyright in the structure of the collection of geospatial data; and b. the sui generis right in the contents of the database.

Copyright in the structure of a geospatial database

This image of an NTF file (Figure 7) illustrates how geospatial data may be represented within a geospatial database.

Figure 7 (NTF file)



It has a literary feel to it – the type of thing that might well be protected by copyright as a table or compilation. Although neither is defined in the Copyright Designs and Patents Act 1988 (CDPA), historically a variety of products have been treated as falling in this category and thus within the parameters of copyright protection. These have included: tables comprising grids of five-letter sequences for a monthly newspaper competition;[5] a list of Stock Exchange prices;[6]a price list;[7] a list of three letter mnemonics,[8]and an electronic circuit diagram[9] to name but a few. Seldom, although it has happened, has a work been denied protection under this head.[10] In *Cramp v Smythson*[11] protection was not accorded to a chronological list of tables in a diary whose making was 'automatic and only required painstaking accuracy'.

To be protected within this category, a table or compilation has to be 'original'.[12] This criterion has historically been satisfied in two ways: by the quantity of labour expended in gathering together the materials for the work, and by the quality of the labour expended in selecting and arranging the material into final form. On the first criterion, the courts have generally found that unfair advantage may not be taken of the labours of

[5] *Express Newspapers v Liverpool Daily Post and Echo* [1985] 1 W.L.R. 1089,
[6] *Exchange Telegraph Co Ltd v Gregory & Co [*1896] 1 Q.B. 147.
[7] *Payen Components South Africa Ltd v Bovis Gaskets C.C.* (1996) 33 I.P.R. 406 (South Africa).
[8] *Microsense Systems Ltd v Control Systems Technology Ltd* noted at [1992] I.P.D. 15006 (the three-letter mnemonics comprised the language code for communication with a pelican crossing controller).
[9] *Anacon Corporation Ltd v Environmental Research Technology Ltd* [1994] F.S.R. 659 (the diagram showed a large amount of writing and symbols and was a list of components together with information as to how they were connected).
[10] *Ladbroke (Football) Ltd v William Hill (Football) Ltd* [1964] 1 W.L.R. 273, at 278. Lord Reid pointed out that cases in which copyright has been denied to a compilation are comparatively few.
[11] [1944] AC 329
[12] CDPA ss 1(1)(a) and 3(1)(a).

another: '*no man is entitled to avail himself of the previous labour of another for the purpose of conveying to the public the same information'.*[13]

Databases and copyright: the current law.

In 1996 the European Directive on the Legal Protection of Databases was enacted.[14] The Directive provides for copyright protection for those databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation.[15] Protection extends to the structure of the database.[16] The sui generis[17] database right gives to the maker of the database (the person who provides the investment necessary for such compilation)[18] exclusive rights to prevent unauthorised extraction and re-utilisation of a substantial part of the contents of the database,[19] rights which may be licensed or sold. As will be discussed below, although a collection of data may fall under the definition of a database, it is only where the relevant investment is expended in the obtaining, verification and/or presentation of the data that the sui generis database right will arise.

In implementing the Directive changes were made to the CDPA. The legislation now provides that a literary work includes *'a table or compilation* ***other*** *[emphasis added] than a database'*.[20] Thus the first port of call is to ascertain whether a collection of geospatial data falls under the definition of a database. Only if it does not will it be necessary to look to the law on literary copyright in tables and compilations.

The ECJ and the Database Directive

---

13 *Scott v Sandford* (1867) L.R. 3 Eq. 723.

14 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases OJ L 77, 27.3.1996, p. 20-28 (hereafter Database Directive).

15 Database Directive Article 3.

16 Database Directive Recital 15 and Article 3.

17 Sui generis is a (post) Latin expression literally meaning 'of its own kind'/genus or unique in its characteristics. The expressioin was effectively created by scholastic philosophy to indicate an idea, an entity or a reality that cannot be included in a wider concept. In the structur 'genus – species' a species that heads its own genus is sui generis. Wikipedia July 2006.

18 Database Directive Article 7.

19 Database Directive Article 8.

20 CDPA s 3(1)(a).

In 2005 the European Court of Justice (ECJ) had the opportunity to rule on the interpretation of a number of provisions of the Directive referred to it in four cases.[21] Of these cases one concerns details of horseracing fixtures,[22] the three others details of football league matches.[23] The ECJ judgements in these cases provide some guidance on the interpretation of aspects of the Database Directive.

Definition of a database

In order to benefit from the regime set out in the Database Directive, a database has to meet the definition set out in Article 1(2). A database should be:

'a collection of independent works, data or other material arranged in a systematic or methodical way and individually accessible by electronic or other means.'[24]

In *Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE* (OPAP), [25] the ECJ ruled that, to be classified as a database there had to be a collection of 'independent materials.' These had to be '*systematically or methodically arranged and individually accessible in one way or another.*'[26] In addition the materials should *be 'separable from one another without their informative, literary, artistic, musical or other value being affected*.' [27]

Thus the data/information/materials need to be

Independent without losing their informative value;

---

[21] *British Horseracing Board v William Hill* C-203/02 [2004] E.C.R. I-10415 [2005] 1 C.M.L.R. 15 [2005] (hereafter BHB) (from the Court of Appeal, England and Wales); *Fixtures Marketing Ltd. v Svenska Spel AB* C-338/02 (hereafter Svenska) (from the Hogsta Domstol, Sweden); *Fixtures Marketing Ltd v OY Veikkaus Ab* C-46/02 (hereafter Veikkaus) (from the Vantaan Darajaoikeus, Finland); and *Fixtures Marketing Ltd. v Organismoa Prognostikon Agnon Podosfairou* (hereafter OPAP) C-444/02 [2004] E.C.R. I-10549 [2005] 1 C.M.L.R. 16 [2005] (from the Monomeles Protodikio Athinion, Greece).
[22] BHB
[23] Svenska; Veikkaus and OPAP.
[24] The wording of the CDPA is similar. 'Database' mean a collection of independent works, data or other material which – a. are arranged in a systematic or methodical way, and b. are individually accessible by electronic or other means. CDPA s3A.
[25] C-444/02 [2004] E.C.R. I-10549 [2005] 1 C.M.L.R. 16 [2005].
[26] OPAP para 30.
[27] OPAP para 29.

Systematically or methodically arranged and individually accessible

Independent materials and informative value

It seems that with a geospatial database the question over whether the data are independent and whether they have independent informative value are two parts of the same question. Can one element of what is represented in the geospatial database, be both independent and have informative value? In the process described above, it is clear that measurements/data etc. are represented relative to all other elements in the database. It might be argued that a single piece of datum, for instance representing information on the height of a mountain, would be meaningless unless there is data present representing other physical features, such as the location of the mountain, the depth of the valley floor, and by virtue of which the location of the mountain can both be calculated and understood. That after all is what a collection of geospatial data are designed to do: to inform relative to all other surrounding elements. But just because that is what is intended taking into account the data as a whole, does not mean that the individual information cannot be informative in and of itself. Indeed, it may be most useful to a certain constituency (those seeking to find out if there is a mountain of a certain height in Scotland). There is no requirement that all need to find the information informative.

Although not specifically asked whether the database compiled by the British Horseracing Board fell within the definition of a database, the ECJ seemed to have no difficulty in accepting that it did. This database contained information on inter *alia 'the pedigrees of some one million horses, and 'pre-race information' on races to be held in the United Kingdom'* including *'the name, place and date of the race concerned, the distance over which the race is to be run, the criteria for eligibility to enter the race, the date by which entries must be received, the entry fee payable and the amount of money the racecourse is to contribute to the prize money for the race'*.[28] If you take one part of that information – say the date on which a particular race is to be run (but not the place or time) then it could be argued that it has no informative value at all – what is the point in knowing a date when the significance is not apparent? Such a deconstructionist approach would result in very few collections of data and information being protected and that is clearly not the intention expressed in the Directive. Recital 17 provides that a database is to include a collection of material such as 'numbers, facts and

[28] BHB para 10

data' clearly contemplating those databases containing such material will fall within the definition. The reference to independent materials may rather be to avoid other complete works, such as films and musical works, from being broken down into constituent parts to satisfy the definition. Indeed these are expressly excluded from protection by Recital 17.

If larger parts of the database were considered as a whole then it becomes more obvious that the criterion is satisfied. So instead of datum, the focus could be on geospatial data sets. For example, geospatial data that contains information pertaining to Scottish Munros. This 'dataset' may contain a number of records with each record having the following information: mountain location, mountain name, mountain height. The dataset could easily fit within the definition of independent materials. Here, the juxtaposition of the mountain name, location and height would certainly have autonomous informative value and be separable from, for instance, the data representing the nature of the remainder of the Scottish Highlands. Reverting to the BHB example, the date, place and time of a race has autonomous value and separable from that same information on other races.

Arranged in a systematic or methodical way and individually accessible

The next part of the definition of a database requires that the materials in the database be 'arranged in a systematic or methodical way'. In dealing with this requirement the ECJ in OPAP stated:

'While it is not necessary for the systematic or methodical arrangement to be physically apparent… that condition implies that the collection should be contained in a fixed base, of some sort, and include technical means such as electronic, electromagnetic or electro-optical processes … or other means, such as an index, a table of contents, or a particular plan or method of classification, to allow the retrieval of any independent material contained within it.'[29]

[29] In OPAP the ECJ was satisfied that the criteria for systematic arrangement were met by the arrangement of data according to 'dates, times and names of teams in those various football matches'.

Looking to the way in which data are arranged in a geospatial database in forms and fields as discussed above (e.g. mountain location, height and name), it would appear that this requirement would be met in a geospatial database. The data are contained in a fixed base at each of the stages discussed above. Further, there would appear no difficulty in fulfilling the requirement that the data be individually accessible and capable of being retrieved.[30] A geospatial database includes sophisticated means by which the data can be accessed and retrieved whether individually or as datasets.

Summary

A collection of geospatial data falls under the definition of a database in the Database Directive

Does a geospatial database attract copyright protection?

Copyright subsists in a database where, by virtue of the selection and arrangement of the contents, it represents the authors own intellectual creation.[31] According to Recital 15 of the Directive, this protection extends to the structure of the database.

A number of points need to be considered:

The test of the authors intellectual creation;

Whether any of the types of labour considered relevant for the originality requirement in the law relating to tables and compilations might be applicable to database copyright;

The scope of database copyright.

Authors intellectual creation

This originality test of 'authors own intellectual creation' now to be used in determining whether copyright subsists in a database represents a change to the law. Recital 15 of the Directive states: '*Whereas the criteria used to determine whether a database should be protected by copyright should be defined to the fact that the selection or the arrangement of the contents of the database*

[30] T. Aplin, '*The ECJ Elucidates the Database Right*', I.P.Q. 2005, 2, 204-221. *'a work will be "individually accessible" if it is possible to search for the work (whether by keyword, alphabetical arrangement or otherwise) and perceive it distinctly, even if the work is accessed alongside other material.'*

[31] CDPA s 3A(2).

*is the author's own intellectual creation; whereas such protection should cover the structure of the database'; a*nd Recital 16 of the Directive : '*Whereas no criterion other than originality in the sense of the author's intellectual creation should be applied to determine the eligibility of the database for copyright protection, and in particular no aesthetic or qualitative criteria should be applied'*. This standard for database copyright is one that is familiar from civilian legal systems, and is generally thought by most commentators to be higher than the traditional British test of skill labour and effort. The suggestion is that the originality must now in some way represent the personality of the author. [32]

The courts have had little opportunity to elaborate on the test albeit that it was mentioned in passing in the Chancery Division in *British Horseracing Board v William Hill*[33] where Laddie J, referring to Recitals 15 and 16 of the Directive said that *'…for copyright to subsist, it must be shown not only that there is a relevant collection of information but that it is also original. Although there is no requirement to demonstrate aesthetic or qualitative criteria, there must be a quantitative baseline of originality before protection is acquired.'*.[34]

Types of labour: Table and compilation case law

The cases in this area showed two types of labour relevant to showing that a table or compilation was original and thus worthy of protection: quantity and quality.

Quantity

The quantity of labour required to achieve protection was considered in *Autospin Oil Seals Ltd v Beehive Spinning*[35]. This case concerned *inter alia* the making of seals in accordance with instructions in the form of a table. It was acknowledged that the type of literary work was a compilation. On the skill and effort required to justify protection Laddie J said that *'it is not the mere form of words or notation used which justifies copyright protection for a compilation, it is the author's skill and effort expended in gathering together the information which it*

[32] Bently & Sherman, (2004) Intellectual Property Law (2nd ed.) (Oxford University Press: Oxford) (hereafter Bently and Sherman) p. 103.
[33] *British Horseracing Board Ltd v William Hill Organisation Ltd* [2001] 2 C.M.L.R. 12, [2001] R.P.C.
[34] ibid para 28
[35] [1995] RPC 683 at 698

contains'. In other words the focus here was on the labour that was expended *prior* to the expression of the information in the compilation that was considered relevant.
Thus the 'physical effort of writing down names and addresses to produce a street directory' would not justify copyright protection in the compilation. However, 'the effort and skill expended in finding out who lives at which address in which roads'[36] would. It is the painstaking labour which has been expended in assembling the facts that has been considered as worthy of protection.[37]

Quality

It has been suggested that under the qualitative criterion the originality requirement may be met by the way the material is selected or arranged[38] as distinct from the labour that is expended in ascertaining the information. There may have been much labour in deciding what to be included in the compilation, and in gathering those materials together, but the requisite originality criterion is only met once those materials are selected and then arranged in the compilation. So in *Ladbroke (Football) Ltd. v. William Hill (Football) Ltd.* [39] which concerned copyright in a football pools coupon, a distinction was drawn between the labour needed to calculate the odds from that needed to express the resultant document. *'[W]hen all the hard work has been done in deciding upon the wagers to be offered, there still remains the further distinct task, requiring considerable skill, labour and judgment (though of a different kind) in the way in which the chosen wagers are expressed and presented to the eye of the customer'.*[40] As can be seen this is skill labour and judgement of a quite different kind to the type of pre-expressive labour considered in *Autospin*.

Database Copyright

So what then of database copyright? A preliminary point may be made. As will be discussed below, the sui generis database right subsists where there has been substantial investment in the obtaining, verification and/or presentation of the contents of the

[36] Kelly v Morris (1866) LR 1 Eq 697
[37] Laddie, Prescott, Vitoria, Speck, Lane, The Modern Law of Copyright and Designs (2000) (Hereafter The Modern Law of Copyright and Designs) para 3.88. Bently & Sherman term this 'pre-expressive', para 3.1.1.
[38] Bently & Sherman, para 3.1.5.
[39] [1964] 1 W.L.R. 273
[40] Ibid p. 281. See also *Elanco Products Limited v Mandops (Agrochemical Specialists) Limited* [1980] RPC 213, [1979] FSR 46.

database. The purpose of this protection seems akin to the quantitative labour test used to determine the subsistence of copyright. That investment is now protected by way of the sui generis right (so long as the contents are collected and not created - see below). Logically therefore this same labour should not count towards the subsistence of copyright in the database. If it were otherwise, there would be dual protection for the same investment. Although overlapping protection is certainly not unknown in the field of IP, it seems that the policy with databases is to separate the two.[41]

So if the 'collection' element is removed as a criterion for originality and thus database copyright, focus must be on selection and arrangement of the materials in the database – the qualitative or expressive part of the test.

It may be difficult to argue that a geospatial database exhibits the necessary intellectual creativity in the selection or arrangement of the material in the database to satisfy the originality requirement. On the matter of selection: certainly there will be decisions to be made in deciding what data should be captured. But these would seem to relate to the type of geospatial database to be compiled, rather than to the data collected to populate that database. For example, a choice might be made to compile a geospatial database showing the location of all Scottish Munros. That would include the level of detail to be incorporated (such as surrounding lochs, roads, houses). Once the scope had been determined the data would be collected and used to fill the database. Certainly the surveyor may be highly skilled in deciding what should be captured, but that is according to predefined specifications which flow from the decision concerning the scope of the database. Such selection is quite unlike the creative intellectual effort required to decide what should be included in, for example, an anthology. For a geospatial database the selection would appear more apt to fall under the 'pre-expressive' category required to satisfy the quantitative test for originality in non-database compilations – in other words the industrious or painstaking labour needed to gather together the information for inclusion in the compilation. It seems also problematic to argue that there is sufficient originality in the arrangement of the data to meet the test. A geospatial database is designed to represent accurately what is on the ground. It would seem that no intellectual creativity of the kind needed for the subsistence of database copyright is needed to arrange a Scottish Munro, a building or a house in its proper place.

[41] This is apparent from the First evaluation of Directive 96/9/EC on the legal protection of databases Brussels, 12 December 2005.

It will be recalled that in *Cramp v Smythson*[42] copyright protection was denied to a series of tables containing factual information. In that case it was said that the making of a list which is 'automatic and only requires painstaking accuracy' would not, of itself, be original.

'One of the essential qualities of such tables is that they should be accurate, so that there is no question of variation in what is stated. The sun does in fact rise, and the moon set, at times which have been calculated, and the utmost that a table can do on such a subject is to state the result accurately. There is so far no room for taste or judgment'. [43]

This, it seems, is quintessentially the function and the purpose of a geospatial database: that it should be accurate so that there is no question of variation on what is stated. It would seem odd to accord copyright protection to a work under the higher standard of originality for database copyright when a similar work was not accorded copyright protection under the lower level.[44]

Scope of Database Copyright

The Database Directive makes it clear that where copyright protection subsists in the selection and arrangement of the contents then it is the structure of the database that is protected.[45] It is also equally clear that copyright protection does not extend to the contents of the database[46]. What then is meant by structure in this context? It would seem that terminology differs. The Directive refers to 'structure' and to 'selection and arrangement'; Laddie Prescott and Vitoria to the 'order' of the material.[47] Whatever term is used, the key issue seems to be originality in the way in which the material is presented or arranged. Returning to the example of the anthology: once the scope of the collection had been determined, the contents would be arranged according to the plan. Here it would be the arrangement of the content according to the chosen theme that

[42] [1944] AC 329
[43] Ibid Viscount Simon L.C.
[44] Cramp v Smythson [1944] AC 329
[45] Database Directive Recital 15. T. Aplin, (2005) *Copyright Law in the Digital Society: the Challenge of Multimedia* (Hart Publishing: Oxford), (hereafter Aplin), p 118.
[46] Database Directive Article 3.2
[47] The Modern Law of Copyright and Designs 3.86

would be protected. Originality, in other words, is reflected in the way in the anthology is structured.

Under the old law some cases suggested that where copyright subsisted in a compilation, it was sufficient for infringement if a substantial part was copied even if it was re-arranged. An example was where names and addresses were copied from one diary, re-arranged and published in another[48]. But here it tends to be the labour in gathering together the content that has been appropriated. As has been explained in The Modern Law of Copyright and Designs[49] where a compilation results from labour expended in assembling facts, or in the skill judgment and knowledge involved in selecting those things which are to be included in the compilation, or both, *'the copyright in such a work may be infringed by appropriating an undue amount of the material, although the language employed be different or the order of the material be altered'*. Importantly however, where, '*the originality resides in the order of the material, and the effect of the rearrangement is to destroy this, so that [the third party] is no longer appropriating a substantial part of the author's work, the process is legitimate.'* [50]

An examination of the scope of database copyright fortifies the conclusion that database copyright does not subsist in a collection of geospatial data. As argued above, a geospatial database is designed to represent what exists. The arrangement of the data flows naturally from its inclusion in the database. Certainly the contents of a database may be more or less inclusive but that does not detract from the fact that the way they are structured/arranged is designed to reflect reality. To argue that the structure (arrangement) of a geospatial database was protected by copyright would be to give a monopoly to the first-comer in that structure. That has not been the position with the directory cases which have made clear that because copyright protection might have been accorded to a certain type of directory (e.g. a list of names and addresses) that did not prevent a third party from gathering together the same information to produce the same type of directory. The originality related to the labour expended in gathering the information not to its presentation. All the more so it would seem for a database of geospatial data: there is only one way in which the information may be presented – it is designed to represent accurately what exists. A third party could gather the same

[48] *Waterlow Publishers Ltd. v Rose* (1989) 17 IPR 493; see also *Waterlow Directories Ltd. v Reed Information Services Ltd.* [1992] FSR 409; *Independent Television Publications Ltd v Time Out Ltd* [1984] FSR 64.
[49] Para 3.88
[50] Para 3.90

information and present the information in the same way: that should not infringe the way the data is presented in the first database. If the second comer takes a substantial part of the contents of the existing database that is a different matter and is now is a matter for the sui generis right, and not for copyright. [51]

Summary

No database copyright subsists in the structure of a geospatial database.

Sui Generis database right

Does a geospatial database qualify for the sui generis right of extraction and re-utilisation?

As discussed above, a collection of geospatial data falls under the definition of a database. However it does not thereby mean that it will also attract the sui generis right provided for in the Database Directive.[52] For the right to subsist there has to be have been investment in obtaining, verifying or presenting the contents of a database consisting in the deployment of financial resources, and/or the expending of time, effort and energy.[53] When considering whether the right subsists, the emphasis that the ECJ placed on the purpose of the sui generis right should be borne in mind. It is to *'promote the establishment of storage and processing systems for existing information'.*[54] The focus should therefore not be on the data per se, but the investment in the system used to process the data.

Obtaining

[51] Interestingly in *British Horseracing Board v William Hill* [2001] R.P.C. 31 BHB apparently asserted that it also had copyright in its database but said it was content to pursue the action on the base of the database right. (Para 23). Similarly when a dispute arose between BHB and Attheraces concerning the price of the supply of data, William Hill made several assertions (to Attheraces) that it owned copyright in the database but failed to present arguments to back its case. *Attheraces Ltd v The British Horse Racing Board* [2005] EWHC 3015 (Ch), Para 271. It would seem that BHB may have thought that it was on rather weak ground in this assertion. See also the comments of the Court of Appeal in this case [2007] EWCA Civ 38 paras 252-258.

[52] In each of the cases considered by the ECJ the material fell within the definition of a database. In BHB the ECJ gave a strong steer to the effect that the database did not qualify for the sui generis right because of the absence of relevant investment.

[53] Database Directive Recital 40; Article 7.

[54] BHB para 31

Despite the ECJs attempts to clarify what amounts to the relevant investment in 'obtaining' data for the purposes of the subsistence of the database right, questions do remain. The ECJ said the relevant investment *'must… be understood to refer to the resources used to seek out existing independent materials and collect them in the database and not [emphasis added] to resources used for the creation as such of independent materials'*.[55] So the materials to be placed in the database must already exist as independent materials. Any investment expended in creation of materials will not count towards the subsistence of the sui generis right. Only when those materials exist, and thereafter investment is expended in the collection of those materials, will this criterion be met.

But what is collected and what is created? In BHB the database in question comprised inter alia information on over one million horses, and in particular pre race information on races held in the UK. The latter information included the name, place and date of the race concerned, the distance over which the race is to be run, the criteria for eligibility to enter the race, the date by which entries must be received, the entry fee payable and the amount of money the racecourse was to contribute to the prize money. When trying to find the line between the creation and the collation of the data, the ECJ said that the investment in the selection of the horses admitted to run in the race concerned related to the *creation* of the data which made up the lists for those races.[56] BHB had expended resources to *establish (emphasis added)* the date, the time, the place and/or name of the race and this was investment in the *creation* of materials contained in the BHB database and not in their collation.[57] It would seem that if the data does not exist as such, then any investment expended in establishing what those data are amounts to creation of the information and not mere collation. Thus the investment by BHB was not relevant when considering the criterion of obtaining for the subsistence of the database right.

So what of the data in a geospatial database? Is that created or collated? Is the material established for the purposes of incorporating it into a geospatial database? Or is it already created merely awaiting collation?

There can be no doubt that geospatial data exists – a point that might argue in favour of a finding that the data are collated and not created. The location and height of Ben

[55] OPAP para 40.
[56] BHB para 38.
[57] BHB para 80

Nevis (a Munro) exist whether or not that information is recorded in any particular form. It is the case that operations are carried out on those data, but many of these are carried out relative to other data within the database. Thus, for example, the information relating to any particular object needs to be represented in relation to all other objects within the database; a unique identifying number may be allocated. But that is to ensure that the data are represented as it should be within the database and relative to other data. Arguably it is not done in order to create or establish the information.[58]

It could also be argued that taking the measurements are taken for the first time and entering them in a database is creation rather than collation of data under the BHB ruling and thus would not meet the relevant investment for the subsistence of the sui generis right. However, a number of distinctions can be drawn as between the type of data 'created' for the purposes of the sui generis right in BHB compared with collection of geospatial data. That geospatial data exists whether it has been collected or not compares with the data concerning horseracing created by BHB. There the horse races did not exist, or at least not in the form that they took for the purposes of British racing. The raison d'etre for BHB was to create the data for the purposes of horse racing. Here the emphasis the ECJ placed on the rationale for the sui generis right should also be recalled: it is to *'promote the establishment of storage and processing systems for existing information'*.[59] By contrast with horse racing, it seems that the expenditure incurred in the collection of geospatial data is intimately bound up with the establishment of the geospatial database – the processing system.

A further factor that may count towards the data being considered collected rather than collates is that, unlike the position in BHB where it was only BHB who was in a position to establish the information on horse races, it is open to anyone to collate geospatial data. Considerations of financial resources aside, geospatial data from, for example, the GPS system[60] can be acquired by anyone.[61]

---

[58] For a discussion concerning 'official' databases in light of the Court of Appeal determination of the BHB case see S. Kon, & T. Heide, *BHB/William Hill – Europe's Feist* 2006 EIPR, 60-66.

[59] BHB para 31

[60] It would appear that the majority of information from the GPS system is freely available en.wikipedia.org/wiki/Global_Positioning_System

[61] It may be that the controllers of the satellites filter what data are available within the GPS system. For example, in the interests of national security, certain data may be filtered out.

On the subject of financial resources, it should be noted that the Directive specifically states that the resources to be expended in the collation of data can be financial.[62] In other words, paying to buy (or license) the data could meet this criterion. This has led a number of commentators to note the apparent incongruity that might arise where one body (such as BHB) who engaged in creation of data and where the right thus did not exist would be in a worse position than a third party who 'bought' the data from BHB – thus expending the necessary financial resources. But if this is to be an important factor, then and to the extent that a payment is made by the maker of a database to a third party (this would not include a subcontractor[63], but would include a third party who assumed the risk of gathering the data – for example from the GPS system) so the maker of a database would be expending the relevant resources.

Verification

The Directive also talks of investment in the verification of the contents of a database as being relevant for the subsistence of the sui generis right.

In relation to this point the ECJ has said the expression *'investment in … the … verification … of the contents of a database must be understood to refer to the resources used, with a view to ensuring the reliability of the information contained in those database, to monitor the accuracy of the materials collected when the database was created and during its operation'.*[64]

It would seem that any relevant investment in verification of the data must take place at the point at which the data enters the database (the accuracy of the materials collected when the database was created) and once the data are in the database (during its operation) rather than verification in the course of creating data i.e. verification establishing whether the data are correct in the first place (is Ben Nevis (a Munro) really there?). So when might investment in relevant verification take place for geospatial data? Looking to the processes of collection of data, when a surveyor is out in the field, she may take measurements of a particular building. Verification that takes place at the point of collection of the data (when the measurements are taken – are they correct?) should

[62] Database Directive Recital 40.
[63] Database Directive Recital 41.
[64] OPAP para 43

not count for relevant investment. However, once the data have been collected, there may be special features built into the database that would check that particular entry against other entries within the database – a process that might be carried out automatically (through the operation of a program in which investment has been expended) or manually – through the operation of certain choices made by an individual manipulating the software. So, for example, if the height of Ben Nevis is (erroneously) entered, through either automatic or manual processes it could be verified that such a measurement is impossible, thus enabling the correct figure to be re-checked and amended accordingly. The relevant investment for the verification of the data would be the processes of checking – and not the establishment of the correct measurement when it was discovered the first was wrong.

It may be that substantial investment has been expended in developing software that checks the accuracy of the data when entered into the database or once in there. At this point is should be stressed that the Directive makes it clear that the term 'database' does *not* extend to computer programs used in the making or operation of a database[65]. As computer programs have their own framework of protection in the Computer Programs Directive[66] it would appear this, again, is intended to avoid cumulation of protection. However there appears no reason that would preclude either the investment necessary in developing a computer program that would support the verification of the data, or indeed the investment necessary in operating a program which enabled data to be verified, from being counted towards the investment necessary for the subsistence of the sui generis right.[67] It is just that the protection does not extend or apply to the computer program per se.

Presentation

On the relevant investment concerning the presentation of the contents, the ECJ in *OPAP*, *Svenska*, and *Veikkaus* said that this referred to:

[65] Database Directive Recital 23 and Article 1(3)
[66] Council Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programs
[67] See e.g. Aplin, p 70 where the author questions whether investment in a computer program as such would be counted as relevant investment (in the context of presentation) but notes that if none of the investment could be counted it would be difficult to envisage what kind of investment could be applied to presentation of the contents that did not relate to the design of the underlying software.

'... the resources used for the purpose of giving the database its function of processing information, that is to say those used for the systematic or methodical arrangement of the materials contained in those database and the organisation of their individual accessibility.'[68]

It appears that the relevant investment is that which is linked to the presentation features that are integral to the processing system as such, and not the presentation features that would count towards the 'look' of the end product. Thus, in a geospatial database, the relevant 'presentation' investment would relate to the way in which the data were arranged within the database itself (in the fields within the geospatial database for example). What it would not appear to cover is the investment needed to present the data in its final form – for instance in the form of the map shown at Figure 3 above.

Such a conclusion would make sense in that any skill, labour and effort[69] that goes towards the final presentation of the geospatial data in the form of the final product (such as a map – see Figure 3) will be concerned with copyright that may subsist in that work. It can thus be separated from, and does not overlap with, the sui generis database right which is solely concerned with the investment needed in developing the means by which the information can be processed.

Summary

There are arguments for saying that a collection of geospatial data within a database meets the definition of a database within the Database Directive and that relevant investment is expended in verification and presentation of the data to qualify for the sui generis right. Whether geospatial data is 'obtained' within the meaning of the Database Directive is open to debate. However, and as is clear from the wording of the Directive, investment is not required in all three areas – one, or more is sufficient.[70] It may, however, have an impact on the scope of protection as discussed below.

---

[68] OPAP para 43; Svenska para 27; Veikkaus para 37

[69] *Ladbroke v William Hill* [1964] 1 All E R 465 at 469.

[70] Database Directive Article 7.1. '…investment in either the obtaining, verification or presentation of the contents…'.

Extraction and re-utilisation

Where the sui generis database right exists, then the maker of the database has the right to prevent the extraction and/or re-utilisation of the whole or a substantial part of the contents of the database without authorisation.

The Database Directive defines extraction as the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form,[71] and re-utilisation to mean any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission.[72] In other words, if a substantial part of the contents of a geospatial database is printed out, placed on a CD Rom or other medium that will infringe the right of extraction, and if a substantial part of the contents is distributed to the public, for instance over the Internet, that will infringe the right of re-utilisation.

Interestingly nothing is said about the manipulation of the data so that, if the data are combined with other data and made available in a completely different form (for example a map) whether that would infringe the right. Looking to the definition it seems it would infringe the extraction right if it used a substantial part of the database. To manipulate the data they have to be extracted from the database. But would making available of the data in a completely different visual form (e.g. a map) infringe the re-utilisation right? It is not the contents (or a substantial part) of the database that are made available, but rather an image (e.g. Figure 3 above) of what is represented by the data. This might perhaps be thought of in a different context. In an example of a database of houses for sale; so long as there has been substantial investment in the database, so the extraction and re-utilisation of a substantial part will infringe the right.[73] Re-utilisation would normally mean re-utilisation of the images and details of the houses as they have originally been depicted – albeit that they may be arranged differently. The re-

[71] Database Directive Article 7.2.(a).
[72] Database Directive Article 7.2.(b).
[73] Although whether there would be substantial investment for the sui generis right to subsist is debatable. In the Netherlands the Court of Appeal decided that the websites of real estate agents did not show substantial investment sufficient for subsistence of the database right. *Zoekallehuizen.nl v. NVM* 4 July 2006 – Court of Appeal (Gerechtshof) Arnhem.

arrangement would make no difference to the infringement of the right. But what if the images were completely re-arranged so that they all looked quite different? It seems that even here the re-arrangement should make no difference. What matters is whether a substantial part of the original database has been extracted and/or re-used.

Questions on the extent of the right were raised in BHB where the ECJ said that, as acts of unauthorised extraction and re-utilisation from a source other than the database concerned *'are liable … to prejudice the investment of the maker of the database, … direct access to the database was not a prerequisite*' for infringement of the rights.[74] Further, while the sui generis right does not extend to cover consultation of a database,[75] the consent of the maker of the database to consultation does not entail exhaustion of the sui generis right. Thus, it does not matter whether the data are extracted or re-utilised directly from the original geospatial database, or through the medium of a third party. If a substantial part of the contents of a protected base are in issue, then the rights of extraction and re-utilisation will be infringed no matter the source of the data.

Insubstantial/Substantial part

A key question in determining the strength of the right is to determine what amounts to a substantial part of the contents of the database. Here there are two tests: quantitative and qualitative. Either one of these will suffice (i.e. either quantitative *or* qualitative) although both may be satisfied.

Quantitative part

The ECJ has said that a substantial part evaluated *quantitatively* refers to the volume of data extracted from the database and must be assessed in relation to the volume of the contents of the whole of those database. If a user extracts and/or re-utilises a quantitatively significant part of the contents of a database whose creation required the deployment of substantial resources, the investment in the extracted or re-utilised part is proportionately equally substantial.[76] This seems as if the part extracted must be judged

[74] BHB para 53.
[75] BHB para 54 where the European Court of Justice said that mere consultation of a database is not extraction of the database. – but did not further elaborate on what was meant by consultation.
[76] BHB para 70.

by the size of the database as a whole. Indeed, in BHB despite having suggested that the investment was in the creation of the data per se, the ECJ went on to comment that the extraction by William Hill of the names of the horses running in a particular race, the date the time and/or name of the race and the name of the racecourse did not constitute a substantial part evaluated quantitatively – being only '*a very small proportion'* of the whole of the database.

So how much is substantial? The ECJ did not quote a figure or percentage. However when considering the test for when the repeated and systematic extraction and/or re-utilisation of insubstantial parts of the contents of the database[77] would infringe the sui generis right, the ECJ said that this measure

'prohibits acts of extraction…which could lead to the reconstruction of the database as a whole, or at the very least a substantial part of it…whether those acts were carried out with a view to the creation of another database or in the exercise of an activity other than the creation of a database'[78].

If a similar test is used in relation to determining a quantitatively substantial part of a database, it would appear that a quantitative threshold would seldom be reached. Substantial must relate to something over 50% even if it did not result in reconstitution of the database.

A question may also arise as to what makes up the database. It may be that the owner of a geospatial database gathers data into one large database but then splits that into other products. For example, Google Earth not only makes a product available for free, but it also has a number of other products for which it charges.[79] Such a 'splitting' of the contents should not detract from the fact that there needs to be substantial investment in the database for the right to subsist. Thus, if the original database cost, say, £1m, splitting that into ten separate products should not automatically produce ten separate sui generis rights. Rather the question should be as to whether substantial investment has been expended in any one of those products and thus whether the right subsists in that part of the database. The more parts into which a database is split, the less likely it should be that the relevant investment would subsist in one small part. If that were not

[77] Database Directive Article 7(5).
[78] BHB para 87
[79] earth.google.com/

the case, it would be possible for any database maker to 'split' a database into a large number of small databases, and yet claim rights in each – in effect avoiding the 'quantitative' rule.

Qualitative part

A substantial part of the contents of the database is determined not only by a quantitative test, but also by a qualitative analysis which is a much more difficult criterion to operate. A *qualitative* part of the contents of a database refers to '*the scale of the investment in the obtaining, verification or presentation of the contents … regardless of whether that represents a quantitatively substantial part of the general contents of the protected database'*.

The enquiry as to whether the part at issue is substantial refers to the investment in the creation of the database and the prejudice caused to that investment by the act of extracting or re-utilising that part.[80] The savings to the person extracting the data would thus seem irrelevant. The intrinsic value of the data must be ignored in deciding what, qualitatively, is a substantial part of a database. To argue otherwise would be to accept that the data per se were protected. Thus, in BHB it was irrelevant that the data extracted and reutilised by William Hill was essential to the organisation of the horse races organised by BHB and others.[81] However it may be that a quantitatively negligible part of the contents of a database may in fact represent, in terms of obtaining, verification or presentation, significant investment.[82] Determining what might amount to a qualitatively substantial part where there may have been large scale investment is far from easy – particularly where the concern is to avoid protecting the data per se. So, for example, if £10 million is invested in the collation, verification and presentation of the data in a geospatial database which is 1Tbyte, would taking of $1/10^{th}$ of the database in which £1 million pro rata had been invested amount to qualitatively a substantial part? Would £10,000 and $1/1000^{th}$ of the database qualify? What if the database is actually very small and consists of 1000 pieces of data but in which £1 million has been spent on the relevant investment? On this analysis it could be that the datum would be protected and that is clearly not within the thinking of the ECJ. In a very expensive but small database, where is the dividing line between protection of data in which there has been a

[80] BHB para 69.
[81] BHB para 72
[82] BHB para 71.

substantial investment in its obtaining, verification and presentation, and protection of the data per se? Does the level of investment depend on the investment in the database as a whole? Or is there a base level of investment that will be protected? So for example, in BHB the sum spent on the database was said to be in excess of £15million. What about say a database of houses for sale? Or jobs available for researchers? Much less in absolute terms will be invested in these types of databases (if protected at all).[83] Much will depend on the base level of investment protected under the sui generis right (an issue not dealt with by the ECJ which merely pointed out that the investment in the data extracted (the date, time, name of race, name of racecourse) was investment in the creation and not obtaining of the data – and thus irrelevant for the subsistence of the right. Where the database right does subsist, the question to be asked is whether there has been a substantial investment in the part of the database that is extracted. Thus, for example, to determine whether the extraction of one dataset from a geospatial database is a qualitatively substantial part, the investment in that part would need to be considered.

Summary

> **A geospatial database exhibits the requisite investment in [obtaining], verification and presentation of the contents for the sui generis right to subsist.**
>
> **Unauthorised taking and making available of substantial parts of the contents of the database will infringe the right of extraction and re-utilisation**
>
> **Where a substantial part, the right is infringed whether the contents are sourced from the original database, or from a third party**
>
> **Consultation of a database does not infringe the right**
>
> **Uncertainty remains over what amounts to a substantial part of a database evaluated qualitatively.**

Exceptions/Limitations on the Sui Generis Right

Lawful users and an insubstantial part

[83] Op cit fn 73.

The Directive provides that a lawful user of a database has the right to extract and/or re-utilise an insubstantial part of the contents.[84] This is subject to the proviso that any acts by the lawful user must not conflict with the normal exploitation of the database or unreasonably prejudice the legitimate interests of the maker of the database.[85] Any contractual provision seeking to override this measure is null and void.[86] In implementing these provisions into UK law, the Copyright and Rights in Databases Regulations 1997 provides that *'a lawful user of a database… shall be entitled to extract or re-utilise insubstantial parts of the contents of the database for any purpose and where under an agreement a person has a right to use a database, or part of a database,…any term or condition in the agreement shall be void in so far as it purports to prevent that person from extracting or re-utilising insubstantial parts of the contents of the database, or of that part of the database, for any purpose.*'[87]

While it has been argued that the narrowness of the general exceptions do little to curb the potential breadth of the sui generis right, this provision prohibiting the use of contract to prevent a lawful user from making any use of an insubstantial part of a database, or limiting the use that might be made of an insubstantial part, could be important for the use by third parties of geospatial data extracted from a database in which the sui generis right subsists. Although there is no protection for insubstantial parts of a database it will often be the case that to gain access to the contents of a database (or part of a database) the user will have to enter into a licence agreement with the maker of the database. (In the research and education sector most end users access data having entered into an agreement). In that licence agreement the maker may require certain undertakings from the licensee as to the conditions on which the data may be extracted and, more likely, re-utilised. To the extent that these conditions relate to an insubstantial part of the contents of a database they will have no effect and the user will be free to do what she will with the contents. The question remains as to what amounts to an insubstantial part of the database – discussed above.

[84] The Database Directive does not define lawful user but it has been suggested that it refers to a person who lawfully acquires the database – e.g. by way of gift, rental sale or as a licensee. Aplin p177. Where authorised to extract and/or re-utilise only part of the database, then the right applies to only that part. Database Directive Article 8.
[85] Database Directive Article 8.2
[86] Database Directive Article 15
[87] Copyright and Rights in Databases Regulations 1997 (SI 1997/3032) (hereafter Database Regulations) Reg. 19.

Lawful users, teaching and non-commercial research

A final point to consider in relation to the sui generis right is the extent of the exceptions/limitations provided for in the legislation. There are a number of permissive exceptions in the Directive, few however of which have been carried into domestic legislation. Of particular note is that fair dealing with a substantial part of the contents does not infringe the right when extracted for the purposes of illustration for teaching or research (and not for any commercial purposes) as long as the source is indicated[88] and that the person making the extraction is a lawful user.[89]

It may seem that as data extracted for these purposed may not be re-utilised (thus inhibiting further dissemination) the exception is of limited utility in the research and teaching sector. However, a number of points are important. On matters of research, it may be that a researcher extracts a substantial part of the database for non-commercial research purposes (which she is perfectly free to do), but that only an insubstantial part of that is eventually re-utilised. This may be particularly so where the data are combined from a number of sources. An example is given in stage 3 of the process described above, where the researcher has combined data from a number of sources. Now it may be (although perhaps unlikely) that the original extraction was substantial. The researcher is quite at liberty to do this within the exception to the right and, so long as a lawful user, needs no permission from the database maker. Once manipulated and combined with the other data, then the final output shown in Figure 3 may contain only an insubstantial part of the contents of any one of the original databases – enabling the researcher to disseminate the output as required. (note that the focus here is on the substantiality of the extraction from the original database – and not the 'importance' of that extraction to the final output.)

On matters of both non-commercial research and teaching, although a substantial part may be extracted, it may not be made available to the *public*. This begs the question as to what is meant by 'the public' in this context. It must be permissible to circulate the data to other researchers, and display the data to students for the purposes for teaching. If these activities were not permitted there would be little point in including the exception within the Directive. Such a conclusion is re-inforced by the obligation in the Directive to

[88] Database Regulations Reg. 20(1)(b).
[89] Database Directive Article 9.

indicate source if data are used for these purposes[90] which can only be of relevance if the data are to be further communicated. The same arguments apply to researchers making data available to other researchers for non-commercial research purposes. The key question would be as to how broad a body of researchers can be included within this circulation.

Summary

> **A lawful user of the database (e.g. the researcher or teacher in an educational institution) may not be prevented from extracting and re-utilising an insubstantial part of the contents of a database for any purposes whatsoever.**
>
> **A researcher or teacher (also a lawful user) may not be prevented from extracting a substantial part of the contents of the database for the purposes of non-commercial research or illustration for teaching. Re-utilisation may only be enjoined if the final product contains a substantial part of the contents**

[90] Database Directive Article 9(b).

Developing a licensing framework.

Following on from the discussion above, the development of any licensing framework for geospatial data within the research sector should take into account the following:

> No database copyright subsists in the structure of a geospatial database.
> A geospatial database exhibits the requisite investment in [obtaining], verification and presentation of the contents for the sui generis right to subsist.
> Unauthorised taking and making available of substantial parts of the contents of the database will infringe the right of extraction and re-utilisation
> Where a substantial part of the contents of a database are extracted or re-utilised, the sui generis right is infringed whether the contents are sourced from the original database, or from a third party
> Consultation of a database does not infringe the database right
> A lawful user of the database (e.g. the researcher or teacher in an educational institution) may not be prevented from extracting and re-utilising an insubstantial part of the contents of a database for any purposes whatsoever.
> A researcher or teacher may not be prevented from extracting a substantial part of the contents of the database for the purposes of non-commercial research or illustration for teaching so long as the source is indicated. Re-utilisation may only be enjoined if the output contains a substantial part of the contents of the protected database

As may have be apparent from the discussion in this report, the greatest uncertainty in the area of database protection lies in the question as to what amounts to a substantial part of a protected database evaluated qualitatively. Such uncertainties at the margins are familiar from, for example, the law of copyright where the parameters of 'fair dealing' are the subject of much anxious debate. In the absence of any court judgement on this matter, the parameters will be for negotiation between the respective parties based on the state of the law as it is currently known.[91]

[91] It may be useful to note the current pressures, sometimes contradictory, upon the re-use of public sector information: See for example - OFT Report: The commercial use of public information (CUPI) December 2006 OFT861; Proposal for a directive of the

European Parliament and of the Council establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) (PE-CONS 3685/2006 – C6-0445/2006 – 2004/0175(COD) approved 12 February 2007; OPSI report on its investigation of a complaint (SO 42/8/4): Intelligent Addressing and Ordnance Survey; freeourdata.org.uk; proposals for the sell-off of public assets, The Times 8 December 2006.

A licensing framework for a repository capable of managing licensed geospatial assets

Parameters (assumptions):

A repository capable of managing licensed geospatial assets (a repository) will be used in the HFE community for consultation, non-commercial research and teaching purposes.

Geospatial data deposited in a repository will come from a variety of sources and is likely to have passed through various stages of manipulation (see use case compendium at http://edina.ac.uk/projects/grade/usecasecompendium.pdf)

The researchers who deposit data in a repository will either have created the data themselves, used data which are not subject to re-use restrictions and/or will be lawful users of the geospatial databases from which extractions of geospatial data are made

So long as a lawful user, those researchers are at liberty to extract an insubstantial amount from the contents of the source database for any purpose (including for deposit in a repository) and where the data are used for non-commercial research and illustration for teaching they may extract a substantial part from the source database

As the research and teaching community interested in geospatial data is limited, and given that only a limited number of researchers will be interested in any particular part of a geospatial database, making available the extractions to other researchers for the purposes of non-commercial research or illustration for teaching would seem not to infringe the re-utilisation right even where those extractions are a substantial part of the source database

Data deposited by researchers may amount to only insubstantial parts of the source database, but if repeated deposits from the same source database are made, then these may, in total, amount to a substantial part of the source database. The Database Directive shields those who use them for the purposes of non-commercial research and illustration for teaching.

Where a researcher or teacher extracts a substantial part of the source database to use for the permitted purposes then the source must be attributed. [This would be easier to manage if there is an obligation to attribute source no matter the size of the deposit]

Where extractions are to be made, a standard access management technology, such as Shibboleth (via the UK Access Management Federation for Education and Research) or Athens would help to ensure that substantial parts of the source databases are used only within HFE and for the permitted purposes

A repository facilitates the work of researchers and teachers but is not itself a lawful user of any of the source databases [it would be legally difficult for a repository capable of managing licensed geospatial assets to be a lawful user as it has no legal existence].

By depositing the contents in a repository, the repository does not thereby 'extract' contents from the original database within the meaning of the Directive

By holding the geospatial data deposited by researchers, a repository thereby re-utilises the data within the meaning of the Directive (i.e. it makes the data available to the public through on-line transmission)

Where only an insubstantial part of the source database is made available, this would not infringe the re-utilisation right.

Where a repository makes a substantial part of the source database available to the public whether the re-utilisation right is infringed will depend upon whether making available to a limited number of researchers and teachers is considered as making available 'to the public'.

Consultation of a database does not infringe the sui generis right. Therefore there would be no difficulty in having a repository 'open' to be consulted by all.

POLICY QUESTION: Is a repository capable of managing licensed geospatial assets going to give any form of 'advice' to researchers as to what amounts to a substantial part of a source database? Or will a repository capable of managing licensed geospatial assets leave this up to researchers?

Incoming data

Where a researcher or teacher deposits outputs which comprise only an insubstantial part of the contents of a source database for which they are a lawful user, no permissions are necessary from the original maker and no attribution in relation to the data is needed.

Where a researcher or teacher deposits outputs which comprise a substantial part of the contents of the source database for which they are a lawful user, no permissions are necessary but the source of the data must be indicated.

There should be no constraints on a researcher depositing data which she has collected during the course of research. If the investment by the researcher has been substantial (as discussed above) then she will own the database right in the deposit. Consideration should be given as to whether a licence should be granted by the researcher in respect of that right.

The data inputted by the researcher/teacher may include a visual element (such as the map described at Figure 3 above) or text explaining the processes undertaken in reaching the final output. Where this is the case, considerations should be given as to whether the work might be licensed under CC or a similar licence.

Outgoing data

To the extent that a substantial part of a source database is to be used, the data may only be extracted by lawful users for the purposes of non-commercial research or illustration for teaching.

Where a substantial part of the source database is extracted the source must be attributed.

Where possible the fruits of research must be re-deposited?