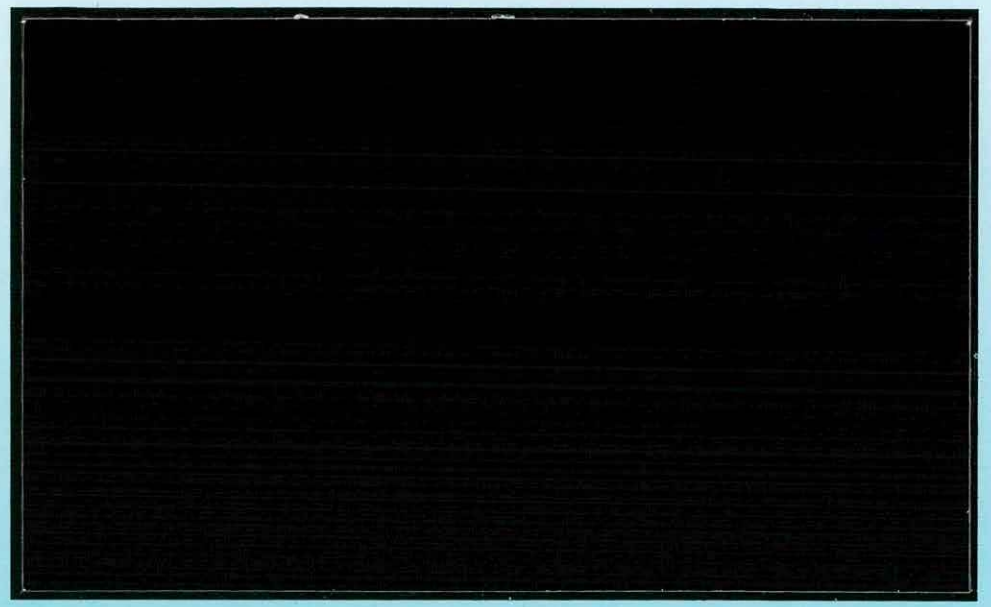
Institute of Terrestrial Ecology



Natural Environment Research Council

ITE has administrative headquarters north and south, and the geographical distribution of its 250 staff in six Research Stations throughout Britain allows efficient use of resources for regional studies and provides an understanding of local ecological and land use characteristics.

ITE NORTH

Edinburgh Research Station
(Admin HQ for ITE North)
Bush Estate
Penicuik
Midlothian EH26 0QB
Tel: 031 445 4343; Telex 72579
Fax: 031 445 3943

Banchory Research Station
Hill of Brathens
Glassel
Banchory
Kincardineshire AB31 4BY
Tel: 0330 823434
Fax: 0330 823303

Merlewood Research Station
Grange-over-Sands
Cumbria LA11 6JU
Tel: 05395 32264
Fax: 05395 34705

ITE SOUTH

Monks Wood
(Admin HQ for ITE South)
Abbots Ripton
Huntingdon
Cambs PE17 2LS
Tel: 048 73 381; Telex 32416
Fax: 048 73 467

Bangor Research Unit
University College of North Wales
Deiniol Road
Bangor
Gwynedd LL57 2UP
Tel: 0248 370045
Fax: 0248 355365

Furzebrook Research Station
Wareham
Dorset BH20 5AS
Tel: 0929 551518
Fax: 0929 551087

The ITE Research Marketing Officers for ITE North and South are based at Edinburgh and Monks Wood, respectively.

2ND DRAFT

INSTITUTE OF TERRESTRIAL ECOLOGY
(NATURAL ENVIRONMENT RESEARCH COUNCIL)

Project T02051m5
DOE/NERC contract PECD 7/2/100 (F3CR05/D6)

Countryside Survey 1990

Database documentation

Report to the Department of the Environment

D C Howard, M T Furse and C J Barr.

Institute of Terrestrial Ecology
Merlewood Research Station
Grange-over-Sands
Cumbria
LA11 6JU

April 1995

Contents

Countryside Survey 1990 databases

Introduction

Original data from the Countryside Survey 1990 (CS1990) is held by the collecting Natural Environment Research Council (NERC) agencies at three sites; ITE Merlewood (field survey of landscape features), ITE Monks Wood (satellite imagery Land Cover Map) and IFE River Laboratory (freshwater invertebrates in running water). The different sites use different computer systems (both hardware and software) to store data and also use individual storage formats. The storage facilities and structures are described by Work Program Output codes (WPO)

The sites are are connected to the Joint Academic NETwork (JANET), which allows information to be passed without transfer to disks or tapes. The sites are also all supported by NERC Computer Services, who can assist in the transfer of data into different formats as required. Agreement between the Department of the Environment and NERC has to be obtained before any data collected during CS1990 can be released.

Database descriptions

WPO B

Air photography
Site: ITE Merlewood
Contact: M K Gillespie
Database: ORACLE
Machine address: MWUE at ITE Merlewood
Ancillary programs:

Aerial photographs are stored in a fire retardent store room. Details of the number of photographs, colour or black and white are stored in a data table. The structure is described below.

Table name: **CS1990_AP** - aerial photograph information *463 records*

SERIES_NO	number(4)	sample square reference number
COLOUR	char(1)	colour or black/white (C/B)
SCALE	number(5)	scale of photograph
NUMBER	number(1)	number of photographs
DATE	date	date flown

WPO D1 & D2

Field survey: planning and survey

Site: ITE Merlewood
Contact: M K Gillespie
Database: ORACLE
Machine address: MWUE at ITE Merlewood
Ancillary programs:

The land cover maps used to record information in the field are stored in a fire retardent store room at ITE Merlewood, along with the associated quadrat recording sheets. A table listing the squares visited is held in the database and is detailed below. To identify information recorded on the sheets see section D3 below.

Table name **SURVEYSQ_LC** - survey sites *508 records*

SERIES_NO	number(4)	sample square reference number
OLD_LC	number(2)	land class pre-1990 (sample)
NEW_LC	number(2)	land class in all squares ITE Land Classification
EASTING	number(3)	easting coordinate at 1 km level using OS National Grid
NORTHING	number(4)	northing coordinate at 1 km level using OS National Grid
COUNTRY	char(3)	country of sample square (dominant area if border)
COUNTY	number(2)	county of sample square (dominant area if border)
LCGR	number(1)	land class group

Details of survey teams and time spent on each sample square are stored in:

Table name: **FABCOVER_90** - details of survey visit *508 records*

SERIES_NO	char(8)	sample square reference number
GRID_REF	number(7)	OS National Grid reference at 1 km resolution (ie EEENNN)
START_DATE	char(9)	date surveyors arrived on square
FINI_DATE	char(9)	data surveyors left site
SURVEYORS	char(60)	surveyors

WPO D3

Data preparation and storage

Site: ITE Merlewood
Contact: J W Watkins
Database: ORACLE
Machine address:
Ancillary programs: ARC/Info, bespoke FORTRAN 77 routines and Countryside Information System

Details of the locations for all the botanical quadrats and descriptions of surrounding land cover are held in:

Table name: **QU90TXT** - quadrat descriptions *11577 records*

SQR	number(4)	sample square reference number
TYP	char(1)	quadrat type ie. X, Y, H, R, V, S, W
REP	number(1)	replicate number
CAT	char(20)	vegetation category
LU	char(80)	land use
PH	char(70)	physiography
SL	char(26)	slope
ASP	char(31)	aspect
SH	char(20)	shading
GR	char(22)	grazing
COMMENTS	char(60)	description

Species lists from each quadrat is held in :

Table name: **QU90_SPECIES** - species presence and cover *217513 records*

SERIES_NO	number(4)	sample square reference number
TYPE	char(1)	quadrat type ie X, Y, H, R, V, S, W
REP	number(1)	replicate number
SPNO	number(4)	species code
Q	number(1)	quadrat nest
C1	number(3)	percent cover in first nest
C2	number(3)	percent cover over whole quadrat

An equivalent table exists for 1978 data, but was not collected as part of CS1990. The species are recorded using a reference code, translation into full species name and the category of species (ie whether it was used in analysis or not) is held in SP_NAMES, the table was not generated as part of CS1990

Table name: **SP_NAMES** - details of botanical species *1531 records*

SP_NO	number(4)	species code
SP_GENUS	char(20)	genus
SP_SPECIES	char(60)	species
SP_TYPE	char(2)	species type
OLD_SP_NO	number(4)	species code in previous projects
SP_CAT	number(1)	species category
SP_AMAL	number(4)	species in aggregates

The land cover element of the field survey is stored in a number of related tables. The information is also stored as coverages in ARC/Info a Geographical Information System (GIS). Each square has coverages (digital maps) for each theme and each year and these are summarised and aggregated into one coverage containing all the information held for the square in both 1984 and 1990. Only the 1990 coverages were recorded as part of CS1990.

Data are divided into two groups, those recorded for squares visited in both 1984 and 1990 (table CS1990) and the squares first surveyed in 1990 (CS1990A):

Table name: **CS1990** field survey land cover data for 1984 and 1990 squares *62003 records*

SERIES_NO	number(4)	sample square reference number
AREA	number(12)	area in square metres
PERIMETER	number(12)	perimeter length around parcel
HS_ID	number(6)	link value to ARC/Info
DEF84	number(3)	Definition code for 1984 survey
S_PH	char(3)	physiology code for 1984
S_AG	char(3)	agriculture code for 1984
S_FO	char(3)	forestry code for 1984
S_ST	char(3)	structures code for 1984
DEF90	number(3)	Definition code for 1990 survey
P_PH	char(3)	physiology code for 1990
P_AG	char(3)	agriculture code for 1990
P_FO	char(3)	forestry code for 1990
P_ST	char(3)	structures code for 1990
G84	char(1)	aggregated definition code for 1984
G90	char(1)	aggregated definition code for 1990

Table name: **CS1990** - land cover data for squares first surveyed in 1990 *12134 records*

SERIES_NO	number(4)	sample square reference number
AREA	number(12)	area in square metres
PERIMETER	number(12)	perimeter length around parcel
H_ID	number(6)	link value to ARC/Info
DEF90	number(3)	Definition code for 1990 survey
P_PH	char(3)	physiology code for 1990
P_AG	char(3)	agriculture code for 1990
P_FO	char(3)	forestry code for 1990
P_ST	char(3)	structures code for 1990
G90	char(1)	aggregated definition code for 1990

Where either surveyors recorded a mosaic of different Definition categories, or overlap between themes (eg physiography and agriculture both being mapped) caused mosaic Definition codes to be generated, areas were proportionally allocated to the 58 defined terms. Details are held in:

Table name: **CS1990_MOSAIC** - allocation of mosaics to primary codes *9193 records*

SERIES_NO	number(4)	sample square reference number
AREA	number(12)	area in square metres
MOS84	number(3)	mosaic code from 1984
MOS90	number(3)	mosaic code from 1990
D84	number(2)	Definition code from 1984
D90	number(2)	Definition code from 1990
FACTOR	number(3)	weighting factor

The Definition code titles can be found in:

Table name: **DEF_CODE** - Definition codes *58 records*

DEF_CODE	number(2)	Definition code
NAME	char(40)	land cover type

The full list of Field Assessment Booklet (FAB) codes for each parcel, as recorded by the surveyors are stored in individual tables for each theme. The tables hold a single FAB code per record and the position of each value in the list. An equivalent set of tables describe data recorded in 1978 and 1984, although they were neither collected nor generated during CS1990.

Table name: **FA_FAB_1990_PH** - full code list for physiography parcels *6286 records*

FA_SERIES_NO	number(4)	sample square reference number
FA_PARCEL_CODE	char(3)	parcel label
FA_FAB_CODE	number(4)	FAB code
FA_FAB_POSITION	number(2)	position in FAB code list for parcel

Table name: **FA_FAB_1990_AG** - full code list for agriculture parcels *65858 records*

FA_SERIES_NO	number(4)	sample square reference number
FA_PARCEL_CODE	char(3)	parcel label
FA_FAB_CODE	number(4)	FAB code
FA_FAB_POSITION	number(2)	position in FAB code list for parcel

Table name: **FA_FAB_1990_FO** - full code list for forestry parcels *64376 records*

FA_SERIES_NO	number(4)	sample square reference number
FA_PARCEL_CODE	char(3)	parcel label
FA_FAB_CODE	number(4)	FAB code
FA_FAB_POSITION	number(2)	position in FAB code list for parcel

Table name: **FA_FAB_1990_ST** - full code list for man-made structures parcels *14598 records*

FA_SERIES_NO	number(4)	sample square reference number
FA_PARCEL_CODE	char(3)	parcel label
FA_FAB_CODE	number(4)	FAB code
FA_FAB_POSITION	number(2)	position in FAB code list for parcel

Table name: **FA_FAB_1990_BD** - full code list for boundary elements *37371 records*

FA_SERIES_NO	number(4)	sample square reference number
FA_PARCEL_CODE	char(3)	parcel label
FA_FAB_CODE	number(4)	FAB code
FA_FAB_POSITION	number(2)	position in FAB code list for parcel

When surveyors found they could not adequately describe a feature using the standard FAB codes, novel codes were generated. The values refer only to individual thematic coverages for one square and the information is held in a separate table:

Table name: **EXTRA_MAPCODES_90** - surveyor generated FAB codes *5544 records*

SQR	number(4)	sample square reference number
MAP	char(2)	theme (ie PH, AG, FO, ST, BD)
CODE	number(4)	novel FAB code
DESCRIPTION	char(60)	description

The standard FAB codes are stored along with their description in:

Table name: **CD_CODE_1990** - FAB code look up table *555 records*

CD_USE_CODE	char(3)	land use code
CD_NO_CODE	number(4)	FAB code
CD_DESCRIPTION	char(30)	description
PRIMARY	char(1)	type of code (primary or secondary code)

The linear feature data are also derived from ARC/Info, but the comparison of vectors is handled differently to that of polygons. Two styles of figure were produced, first simple vector lengths directly from the digitised coverages. Data are stored in DB1984 and BD_1990_508.

Table name: **BD1984** *1887 records*

SERIES	number(4)	sample square reference number
CODE	number(10)	boundary type code
LENGTH	number(12)	length in metres

Table name: **BD_1990_508** *2405 records*

SERIES	number(4)	sample square reference number
CODE	number(10)	boundary type code
LENGTH	number(12)	length in metres

In order to produce the matrix of change (ie what has changed to what), the vectors had to be converterted into rasters (a thin grid) which could then be simply overlaid. The data are held in:

Table name: **BD_MATRIX_RATIO** *5833 records*

SERIES	number(4)	sample square reference number
CODE	number(10)	code of boundary in 1984 and 1990
L_CORR90	number(12)	length corrected against 1990 figures
L_CORR84	number(12)	length corrected against 1984 figures

In order to identify the elements, three look up tables have to be used, to identify the vector information, relate it to the change codes and finally to present it as descriptive text.

Table name: **NEWBD2_LUT** *28 records*

DESCRIP	char(50)	description
D_SHORT	char(10)	brief description
CODE90	number(10)	1990 code
CODE84	number(10)	1984 code
CODE93	number(10)	1993 code

Table name: **NEWBD2_MATRIX** *675 records*

CODE_90	number(10)	1990 code
CODE_84	number(10)	1984 code
CODE_MATRIX	number(10)	code for both years

Table name: **MATRIX_CODE** *675 records*

DESCRIP	char(22)	description
CODE_MATRIX	number(10)	code for 1984 and 1990

In order to produce national or regional predictions from the field samples, the figures have to be weighted by the area contained in each of the land classes. The data is held in a table called SQUARE which also holds information such as county and altitude (MOD data not for widespread release).

Table name: **SQUARE** - information defining Great Britain *240243 records*

EASTING	number(3)	easting coordinate at 1 km level using OS National Grid
NORTHING	number(4)	northing coordinate at 1 km level using OS National Grid
LAND_CLASS	number(2)	land class from the all squares ITE Land Classification
LC_GROUP	number(1)	landscape type (arable, pastural, marginal upland or upland)
ALTITUDE	number(4)	mean altitude in metres (in 5 m bands MOD data)
COUNTY	number(2)	reference for the county covering the largest part of the square
RURAL	number(3)	rural area (ie minus sea and urban)

The county reference number can be translated into county name using the following look up table:

Table name: **COUNTY** - county codes and names *68 records*

CO_NO	number(2)	county reference number
CO_NAME	char(20)	county name
CO_FG_TOT	number(5)	number of 1 km squares

WPO E1 & E2

Freshwater biology: survey and identification

Site: IFE River Laboratory
Contact: M T Furse
Database: ASCII (soon to be loaded into ORACLE)
Machine address: NWA
Ancillary programs: bespoke basic programs

Data are stored in ASCII format on VAX computer systems. There are three files, holding the details of the site, namely SCO_CS1990.DAT, PCO_CS1990.DAT and INV_CS1990.DAT. Their structures are as follows:

File name: **SCO_CS1990.DAT** - sample register file *722 records*
(Two records per site)

line 1		
SITE CODE	1-8	square sample reference number
NAME	10-36	water-course name
SITE	38-64	site name
GRID REF	66-77	alpha numeric OS National Grid reference
line2		
SPR SAMP	10-20	Spring sample date
SUM SAMP	24-34	Summer sample date
AUT SAMP	38-48	Autumn sample date

File name: **PCO_CS1990.DAT** - time variant physical file

SITE CODE	1-8	square sample reference number
DAY	10-12	Julian day
YEAR	14-15	year number (starting from 1978 as year 1)
WIDTH	17-21	water width
DEPTH 1	23-25	depth at position 1
DEPTH 2	27-29	depth at position 2
DEPTH 3	31-33	depth at position 3
VELO	35	surface velocity
CLAR	37	water clarity
TIME	39	duration of sample
METHOD	41	sampling method
PAVE	43-44	percent rock pavement
BOLDER	46-48	percent boulder/cobbles
PEBBLES	50-51	percent pebbles/gravel
SAND	54-56	percent sand
SILT	58-60	percent silt/clay
SUBST	62-66	dominant substrate
MACROP	68-70	percent cover by macrophytes
DETRITUS	72	presence of detritus

File name: **INV_CS1990.DAT** - time invariant physical file

SITE CODE	1-8	square sample reference number
DAY	10-12	Julian day
ALTITUDE	14-16	altitude in metres
MAX WIDTH	18-21	maximum width in sample area
MIN WIDTH	23-26	minimum width in sample area
MEAN WIDTH	28-31	mean width in sample area
DEPTH	33	depth category
STABILITY	35	bank stability
TREES	37	dominance of trees on bank
BUSHES	39	dominance of bushes on bank
REEDS	41	dominance of reeds/rushes
PLANTS	43	dominance of low plants
RIP_COVER	45	dominance of other riparian cover type
SHADE	47	extent of shading
URBAN	49	urbanisation by sample area
ARABLE	51	arable land use by sample site
GRASS	53	grassland by sample site
HEATH	55	heath/moor neighbouring sample site
DECID	57	deciduous woodland by sample site
CONIF	59	coniferous woodland by sample site
WEED CUTTING	61	evidence of weed cutting
DREDGE	63	evidence of dredging
MAINTAIN	65	evidence of bank maintenance
STRAIGHT	67	evidence of course straightening
ORGANIC	69	evidence of organic pollution
INORG	71	evidence of inorganic chemical pollution
PHYSICAL	73	evidence of physical pollution
BRIDGE	75	bridge in neighbourhood
WEIR	77	weir in neighbourhood
INFLUENCE	79	other influences
DISCHARGE	81-82	discharge category
SOURCE	84-88	distance to source in km
SLOPE	90-94	slope

The fourth file holds species lists in coded form. Data are stored in sequential series, with each site initially being identified by an eight character site reference. The first two characters define the land class, the next four the square reference number (series no), the next is a 'C' defining the project and finally an '0' indicating it is a non-seasonal survey. Following lines contain an eight character species identifiers, with up to eight species records on a line. A hierarchical structure permits identification at different taxonomic levels. The first two characters define the phyllum, class or order, the next two define the family, sub-family or tribe, the next two the genus and finally the last two identify the species. The terminator used after each site is -1.

File name: **DCO_CS1990.DAT** - species data file
(two types of line, type 2 repeated as necesary per site)

line 1		
SITE CODE	1-8	sample square site reference
line 2 (repeated)		
SPECIES	1-8	species code, -1 if terminator or blank
SPECIES	10-17	species code, -1 if terminator or blank
SPECIES	19-26	species code, -1 if terminator or blank
SPECIES	28-35	species code, -1 if terminator or blank
SPECIES	37-44	species code, -1 if terminator or blank
SPECIES	46-53	species code, -1 if terminator or blank
SPECIES	55-62	species code, -1 if terminator or blank
SPECIES	64-74	species code, -1 if terminator or blank

A look up table allows translation of the codes to full species name (TAXAUTH.DAT).

WPO F1 & F2

Soil survey: England & Wales and Scotland

Site: ITE Merlewood
Contact: J W Watkins
Database: ORACLE
Machine address: MWUE at ITE Merlewood
Ancillary programs: ARC/Info

The soils data recorded by the Soils Survey Land Resource Centre (SSLRC) and the Macaulay Land Use Research Centre (MLURI) are held on paper maps in the fire retardent storeroom and as digitised coverages in ARC/Info. The areal information derived from the GIS is held in:

Table name: **SOIL_GB** - soils data for surveyed squares *records*

SERIES_NO	number(3)	sample square reference number
AREA	number(12)	area in square metres
PERIMETER	number(12)	perimeter length for parcel in metres
SOIL_SERIES	number(4)	soil series number

the soil type can be identified from:

Table name **SOIL_SERIES** - soil series number look up table *records*

SOIL_SERIES	number(4)	soil series number
DESCRIPTION	char(60)	description

Internal relationships

The datasets held in the ORACLE database are all held in the same account on the same platform (MWUE). ORACLE is a fully relational database management system which is designed to allow information in different tables to be combined and interrogated in different ways. There are no internal links between tables, they are formed by the user joining tables by relating columns. Three figures are included in this report (Figures 1, 2 and 3) which demonstrate some of the potential links between the data tables. Although the tables appear separate, all tables can be linked as the standard entity through all the work is a kilometre square.

Information describing the data, in terms of definitions of the fields in the table are being appended to tables in the associated comments fields. A system is under development which will allow interrogation of the data by users without giving edit access, and, where data is altered, records of the changes and date of change is stored so that analyses carried out prior to update can be repeated. The system does carry a number of indexes for most of the tables; to operate in an efficient way the indexes may require tuning for specific queries.

Data held in the ORACLE database can be linked to the cartographic digital coverages held in ARC/Info using the RDBIO link. In ARC/Info, several thematic coverages are stored for each surveyed square; there are over 10,000 digital coverages stored at Merlewood relating to the Countryside Surveys. The data from ORACLE can also be extracted and loaded onto the Countryside Information System (CIS). Computer programs, written in SQL (the Standard Query Language used with ORACLE) can produce means and standard errors ready for installing using the Sample Database program. The CIS is designed to hold far more qualifying information and comments defining the limits to which the data can be used and any specific flaws or problems.

The CIS holds land cover area (for each of the 58 Definition codes) for 1978, 198 and 1990, linear feature measures (both Definition codes and matrix codes) for 1984 and 1990, soils data (at soil series level) for 1990, and summary vegetation data (1978 and 1990). It is not intended to install the complete vegetation dataset onto the CIS, but summarise the data into species groups and vegetation types or present number of species. Only selected species were used in the analysis so as to remove problems caused by aggregates and possible mis-identifications; the same species have been used to produce the data on the CIS.

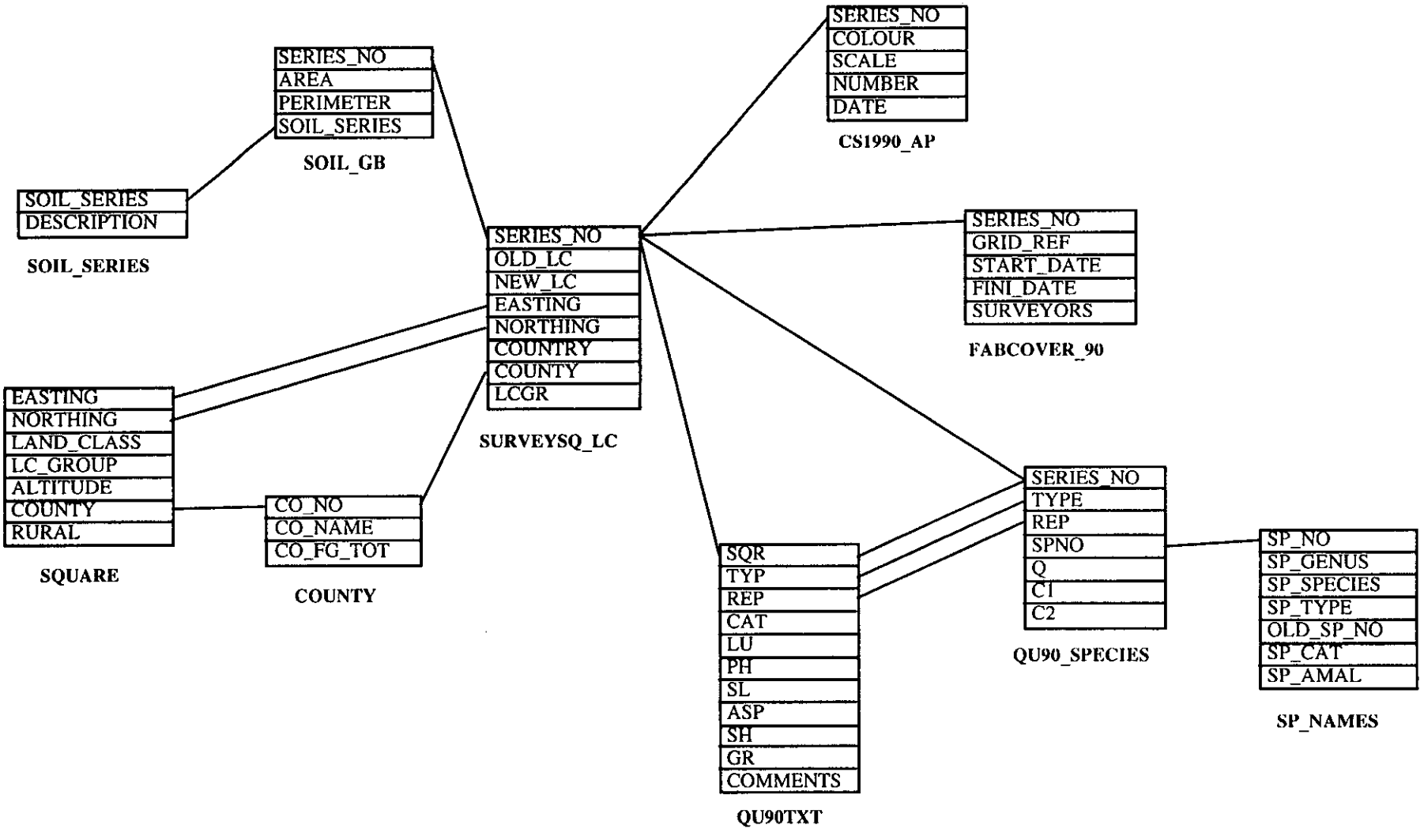


Figure 1 Table relationships for tables holding data on all squares in Great Britain, survey sites in CS1990, air photographs, vegetation from quadrats and soils.

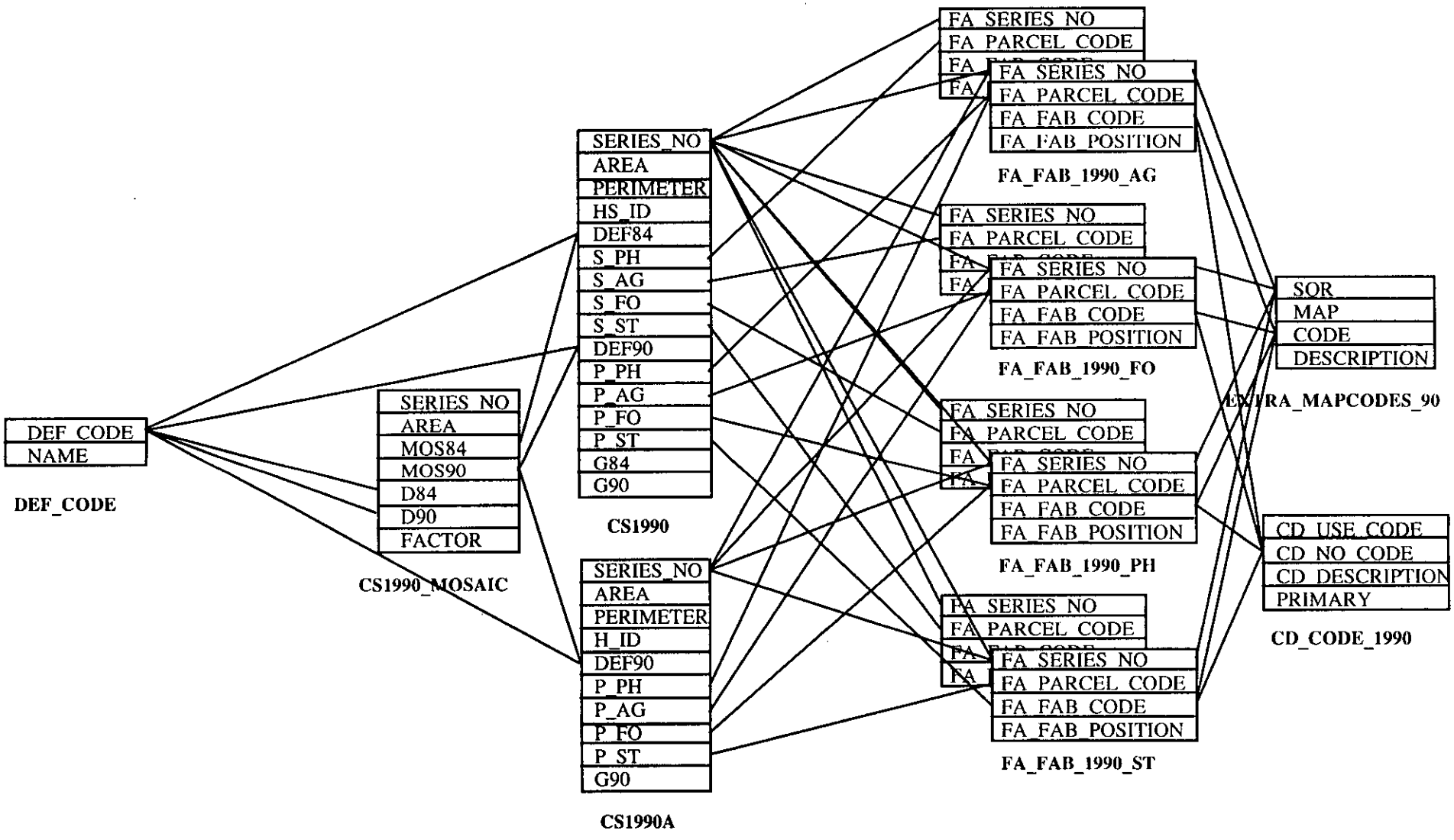


Figure 2 Table relationships showing links between tables holding information realting to sampled land cover from CS1990. Only part of the 1984 data is shown (for simplicity) and none of the 1978. The tables can be linked with those in table one through SERIES_NO for example with table SURVEYSQ_LC.

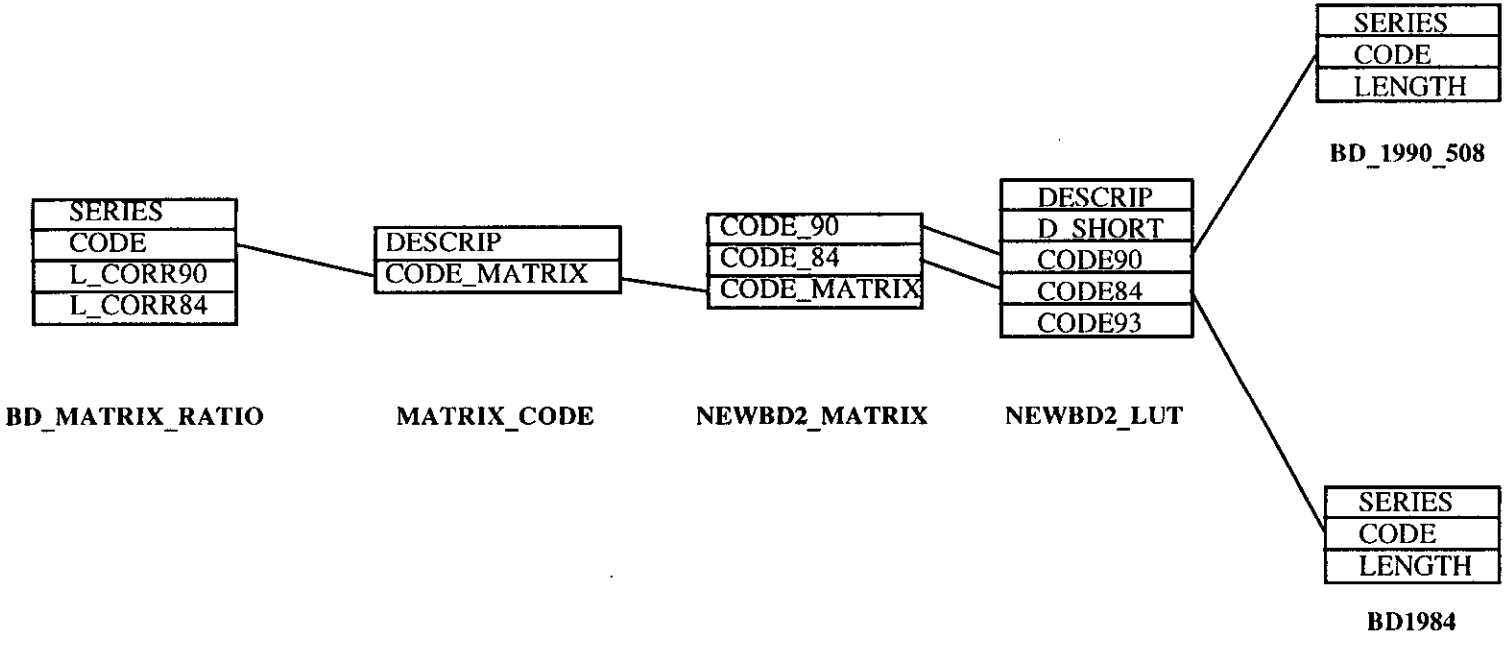


Figure 3 Table relationships for tables holding information about linear features. Not all data tables are shown, and the information can be linked with information in Figures 1 and 2 using SERIES_NO and SERIES.

Data availability

Data from field survey of Countryside Survey 1990 is available in a number of formats. Apart from the reports and documents there are three major avenues to access:

Countryside Information System (CIS)

The version of CIS (5.3) which is currently commercially available contains summary information describing a number of features. Four data sets containing vegetation data (main plot, hedge plot, stream plot and verge plot) hold mean numbers of species, species aggregations and frequency of occurrence of species. Land cover is described in tables which classify data into the 58 reporting classes as described in Barr *et al.* (1993). Linear features are also presented in the categories given in the Main report. Error terms are provided within both land cover and linear features data sets allowing national and regional estimates to be derived.

Four categories of data are provided for freshwater invertebrates in streams, namely population indexes, species frequency, stream management and stream substrate. All the data sets are derived from databases described in this report.

CIS uses most of the data described in table SQUARE as its core data. LC_GROUP is held as four region files (ARABLE.RGN, PASTURAL.RGN, MARGINAL.RGN and UPLAND.RGN), RURAL is held internally and used in upscaling survey estimates and COUNTY is held as an addressable field. ALTITUDE is not held for reasons of commercial confidentiality. CIS only holds information for 240222 squares, the difference between it and the total given in SQUARE is due to Rathlin Island off the coast of Northern Ireland.

Supply of data tables

Data can be supplied as ORACLE export tables, ASCII dumps of tables or in different data structures (eg dBase DBF files). There are restrictions for release of information which include confidentiality agreements with land owners. The OS grid reference of sites will not be released (table SURVEYSQ_LC). Look up tables (eg SP_NAMES and DEFINITION_NAMES) would have to be supplied with some data tables.

The most commonly supplied table is a sub-set of SQUARE (usually EASTING, NORTHING and LAND_CLASS), which can be supplied for the whole or part of GB. There is a pricing policy for the land classification which is indicated in the table below.

Number of squares	*HEI Research/ Co-funders*	*Government*	*Commercial*
0 - 500	£200	£700	£950
500 - 10,000	£200	£1,200	£1,700
10,000 - 100,000	£200	£2,200	£3,200
> 100,000	£200	£5,200	£7,700

Charges for other datasets are based on the type of application and quantity of data required. Charges include staff and computing time to sort, extract and format data and media for transfer.

Merlewood is connected to the JANET network so data can be passsed using electronic file transfer (eg FTP, E-mail) and in exceptional circumstances access may be given to the data servers for remote access and query.

Data preparation and presentation

Many users find it valuable to request data in semi-processed condition, for example prepared as CSF files for loading onto CIS. ITE can process and interpret information at a variety of scales and resolutions. The charges for such processing include data costs, staff time (at Government staffing rate at 'H' level) and computing, and are dependant upon the use for which the data is required.

The costs of such presentation are variable, but estimates of costs can be made if a written description of the task is sent to John Watkins, ITE Merlewood. Quotations will not be issued verbally.

Appendix

NERC DATA POLICY: DRAFT OF PUBLIC STATEMENT

1. Introduction

NERC holds many digital and analogue datasets which when taken together comprise a valuable environmental science resource for the UK. The nature of the resource varies widely. For example, there are: datasets held by individual scientists, on which they are actively working as part of their research; major "live" databases which are continuously managed and extended by the validation and addition of incoming data, and which constitute the working tools of active scientific groups; and data deposited as archives after research projects have been completed.

This document publicises Council's policy towards this valuable resource. It sets out the principles underlying NERC's responsible stewardship of its data assets; describes the current arrangements for their management; and provides the framework for their exploitation in pursuit of scientific understanding or the creation of wealth.

2. Underlying Policy Principles

2.1 Ownership of datasets and the right to exploit them

A variety of activities funded by NERC and/or undertaken by NERC staff will lead to the accumulation of data; eg research funded by grants and studentships; institute and community scientific programmes; the activities of NERC Scientific Services; institute commissioned work; the collation of data as a public duty; and direct purchase of data from other suppliers.

Datasets owned by NERC (ie to which NERC owns the intellectual property rights) will in general be under the direct control of Council, but this need not always be the case. Conversely, datasets under the control of NERC may not be owned by NERC. This may be quite clear, for example, when NERC has acquired data under licence from a supplier, or collected the data under a contract which specifically assigns IPR to the customer. But other cases may be much more complex. There have been unfortunate historic cases where ownership of data is unclear, and possibly spread among several different parties, not all of whom can now be identified. Such circumstances may present a legal obstacle to the exploitation of data by any of the parties. A fundamental principle, therefore, is that whenever NERC activities will result in the acquisition of data, or data will come into the custody of NERC, the ownership of the IPR should be established at the outset.

2.2 Legal and contractual constraints on NERC's data policy

This policy has been framed to be consistent both with legal constraints and formal agreements between Council and other bodies: eg the Environmental Information Regulations, and the framework for the exchange of data for research purposes between the agencies of the Interagency Committee on Global Environmental Change (IACGEC).

2.3 Individual, and corporate, rights to NERC data

Notwithstanding Council's ownership of and rights to data, individual scientists, principle investigator teams and programmes should be permitted a reasonable period of exclusive use during which to work on and publish results from the data collected by such individuals or teams.

2.4 Costs associated with data collection and maintenance

In addition to the obvious costs incurred when data are collected in the first place, the subsequent documentation, validation, storage, maintenance and dissemination of data can also incur substantial costs. Such costs will continue to be incurred for as long as the data are to be kept available. It is NERC's policy to determine these costs; they provide input to management decisions on commercial pricing and on whether specific datasets justify long term stewardship.

2.5 The value of NERC's data

The value of datasets is determined by their end use in enabling science and creating wealth. Their value can be realised by Council: by giving, exchanging or licensing/selling them to scientific researchers; by using them to create wealth within NERC's own institutes; or by licensing/selling them to commercial organisations who will themselves create wealth. Thus data are potentially a tradeable asset.

2.6 Standards

To be readily usable, and hence of value, data should be catalogued and provided wherever possible according to agreed standards of format and quality.

2.7 Justification for the active stewardship of data

Since the ultimate use of data cannot always be anticipated in advance, neither therefore can the ultimate value. But the costs of managing data are such that many datasets will never justify active stewardship. Decisions on which data should be actively preserved need to be taken in the light of their potential future value, and reviewed periodically thereafter. It should be noted that many of NERC's datasets relate to the state of the environment at the time they were collected, and that such historical records may be irreplaceable. Much scientific uncertainty in the environmental sciences can only be reduced by acquisition of time series continuing for as long as possible.

2.8 Purging of data

If a review of particular datasets concludes that Council can no longer afford the costs of stewardship in the light of their likely future value, then NERC will publicise its intention to put at risk or destroy the data before doing so, and make reasonable endeavours to identify some other body or discipline prepared to accept responsibility for them.

2.9 Points of contact

NERC will have well defined points of contact whereby the environmental science community and potential commercial users of Council's data can readily obtain information as to what is available and on what terms.

3. Management responsibility for NERC's data

3.1 Policy development

Advice on data policy issues is provided to Council via its Information Services Advisory Committee (ISAC). Through ISAC NERC will address such issues as copyright, data ownership and intellectual property rights; common standards, protocols and data formats; cataloguing of NERC's data; technical issues such as electronic networking and provision of archive facilities; a policing policy to ensure that data are not misused; quality assurance of data.

3.2 Designated Data Centres

Responsibility for NERC's data, and the implementation of NERC's policy, is divided on discipline related criteria between NERC's Designated Data Centres, which are represented on ISAC. They are as follows

Antarctic Environmental Data Centre: British Antarctic Survey, Cambridge. Responsible for all NERC's data from the Antarctic, regardless of discipline.

British Oceanographic Data Centre: Proudman Oceanographic Laboratory, Bidston Observatory, Birkenhead, Merseyside. Responsible for marine data

National Geosciences Information Service: British Geological Survey, Keyworth, Nottingham Responsible for geosciences data.

National Water Research: Institute of Hydrology, Wallingford, Oxon. Responsible for NERCs hydrological data and for the Government's National River Flow Archive.

Environmental Information Centre: Institute of Terrestrial Ecology, Monks Wood, Huntingdon, Cambridgeshire. Responsible for all other NERC terrestrial and freshwater data.

British Atmospheric Data Centre, ie the former SERC Geophysical Data Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon. Responsible for atmospheric sciences data.

In addition to these discipline-related Data Centres, NERC Scientific Services, Polaris House, North Star, Swindon, Wilts is responsible for NERC's non-discipline-related data holdings, notably the satellite imagery archive at Dundee, and remotely sensed imagery from NERC's airborne campaigns. NERC Scientific Services also manages a top level catalogue of NERC's data holdings; provides a

Natural
Environment
Research
Council