



A joint Research Councils Programme co-sponsored by Defra and SEERAD

Data resources for rural sustainability research: realising their combined potential.

Annex A.

Questionnaire survey of RELU research community

Helen M^cKay, Nigel Boatman, Naomi Jones and Robert Stones.

Central Science Laboratory, Sand Hutton, York YO41 1LZ, U.K.

h.mckay@csl.gov.uk

29 April, 2005.



Annex A. Questionnaire survey.

Contents

Executive Summary 3

Aim..... 5

Methods 5

Results..... 5

Response rate..... 5

The sample 5

Data accessed 8

Difficulties with access 14

Data integration 18

Services respondents would like RELU to provide..... 21

Further work 27

Appendix A. The questionnaire. 28

Appendix B. Datasets currently accessed (respondents were asked to give up to ten examples) (Q6). 39

Appendix C. Difficulties in accessing datasets (Q9)...... 48

Appendix D. Datasets respondents would like access to (Q11)...... 53

Appendix E. Datasets respondents anticipated difficulties in obtaining access (Q13). 56

Appendix F. Examples of integrated datasets used by 29 respondents, and for spatial data, the scale of integration (Q15). 60

Appendix G. Examples of datasets that 37 respondents intended to integrate, and for spatial data, the scale of integration (Q17). 63

Executive Summary

1. The survey

A questionnaire survey of the RELU community (not confined to award holders) was undertaken as part of a scoping study on data resources. The aim was to investigate data access, management and informatics requirements. Over 650 individuals were contacted, and 114 completed questionnaires received, giving a response rate of 17%. However, about 60% of non-respondents contacted felt the questionnaire was not relevant to them. Most respondents were from educational institutions (68%) or government (14%). Nearly half indicated they worked on the environment and a third on rural development, the rest mainly on farming and food or economics. The main research interests of respondents suggested that the majority (58%) were environmental scientists rather than social scientists (23%) or economists (19%). Nearly a half had experience of other research programmes, not only funded by the UK Research Councils (48% of respondents, specifying 26 programmes), but other government and non-government programmes (68% of respondents) and overseas programmes (7% of respondents).

2. Access

Three quarters of respondents (74.6%) indicated that they required access to external datasets (i.e. those held outside of their organisation) to carry out their research. The most widely used types of data were land use/land cover, agriculture/horticulture and administrative boundaries. Compared to current use, in the future there was a greater need for datasets on social attitudes, social structure/social exclusion, recreation, energy and waste. Specific datasets most frequently accessed were Digimap (Ordnance Survey) and Countryside Survey. A list of the datasets currently used by respondents is provided (Table 6 and Appendix B).

3. Discovery of datasets

Most respondents discovered datasets from colleagues (68%), and/or the owning organisation website (55%). Data portals were used by 40% of respondents, almost a third used a library (31%) and only 10% used a data catalogue.

4. Difficulties with access

Of the 85 respondents who required access to external datasets, 51% had difficulty (38% of all 114 respondents). Specific datasets and the nature of difficulties encountered are described; the most frequently cited datasets were Defra Agricultural Census (due to confidentiality) and National Soil Research Institute Soil survey (due to cost). The majority of respondents indicated they would like to gain access to additional datasets to undertake their research (67% as opposed to 12% that would not). The datasets most frequently needed were soil, climate and land use/cover. Nearly half of the respondents anticipated difficulties with access to data, specifying in particular soil data (cost) and scheme uptake data (confidentiality).

5. Data integration

A quarter of respondents indicated they currently integrate or use integrated datasets. Examples given by respondents are listed in Appendix F. Nearly a third of the examples given (29%) involved inter-disciplinarity (between environmental and social data). When asked, most respondents did not know if they were likely to encounter difficulties when integrating datasets (62%), the most frequently anticipated difficulty related to spatial issues. The most common data integration tool, used by three quarters of respondents, was a spreadsheet. Mapping/GIS and statistical software were used by half of the respondents. Databases (36%) and graphical software (24%) were less commonly used. CIS was used by 10% of respondents. Most respondents felt there were no processes that they could not carry out because the appropriate tools were not available (49%), less than half said they didn't know and only 7% thought they needed further tools.

6. Services

The most commonly cited frustrations with data access & compatibility were: spatial scale, cost, confidentiality, format and time taken to access and compile data. When asked about other research programmes, most respondents favoured a proactive approach to data management, including provision of information on data ownership and access, and collation of data and metadata produced by the programme. Specific services which respondents would clearly like to be provided by RELU include:

- information on data sources/availability
- help with data access and acquisition
- communication facility for interaction with other award holders
- a collaborative facility for data sharing with other award holders

There was less support for provision of tools for data integration, possibly due to ignorance. Finally, it seemed clear that respondents were in need of advice on legal issues including FOI Act and confidentiality and IP issues including data ownership. When asked about other developments in data management they would like to see implemented in RELU, widening the availability of data in an appropriate form seemed to be the overriding message.

7. Further work

Social scientists and economists seem to be less well represented in the questionnaire survey than environmental scientists. The next phase of the scoping study will include targeted consultation with social scientists (to investigate why many thought the questionnaire was not relevant to them) and data specialists (to investigate issues arising from the questionnaire as well as trends and technological developments in data management), also a workshop on data integration (planned for 19 May).

For further information on the workshop, see the scoping study web site:
<http://reludata.csl.gov.uk/>

Aim

To investigate data access, management and informatics requirements relevant to the RELU programme.

Methods

A list of questions was drawn up and submitted to the RELU Programme Director's office for comment. Based on this, a web-based questionnaire form was designed and Microsoft Word version prepared (see Appendix A). An e-mail requesting participation in the survey was sent to 653 addresses (those who had expressed an interest in the RELU first call) on 7 December, 2004. Two reminders were sent (on 14 and 21 December). On 4 January, 2004, non-respondents were asked why they had not taken part. By 2 February, 2005, 114 completed forms had been received.

Respondents were placed in categories for further analyses. The type of institution from which respondents originate (Q2) was used to allocate respondents to government or education. Main research interest(s) (Q1) were used to allocate respondents to social scientists, environmental scientists or economists. Not all respondents were allocated to categories.

Results

RESPONSE RATE

Of the 653 e-mails sent out, seven were returned undelivered and 114 forms were completed, resulting in a response rate of 17%. However, of 160 non-respondents giving a reason for not replying, 98 indicated they felt the questionnaire was not relevant to them (61%). Twenty-one percent of non-respondents (33) indicated lack of time was the reason for non-response.

THE SAMPLE

The results are based on 114 completed forms. Table 1 shows that 68% of respondents were from universities or colleges, and 14% from government research institutions.

Table 1. The type of institution from which respondents originate (Q2).

Type of institution	N	%
University/college	78	68.4
Government Research Institute	16	14.0
Charity/not-for profit institution	10	8.8
Other	6	5.3
Research Council Institute	4	3.5
Total	114	

When asked about their research interests (Table 2), nearly half of the respondents (43%) indicated they worked on the environment (water/catchments, other natural resources, land use, biodiversity, pollution/waste or policy), 31% on rural development (sustainable development, social issues or public health), 27% on farming and food (markets/economics, food, ecology or policy) and 11% on economics (mostly agricultural). Other interests included analytical and statistical methods, climate change, remote sensing, ecology, marine science and biochemistry.

Annex A. Questionnaire survey.

Table 2. Main research interest(s) of respondents (N = 114) (Q1).

	%		%
Environment	43.0	Water/catchments	14.9
		Other natural resources	14.0
		Land use	13.2
		Biodiversity	6.1
		Pollution/waste	4.4
		Environmental policy	3.5
		Other/ not specified	3.5
Rural development	30.7	Sustainable development and rural economy	13.2
		Social issues	7.9
		Public health	5.3
		Other/ not specified	8.8
Farming and Food	27.2	Markets/agricultural economics	10.5
		Food	6.1
		Agricultural ecology	6.1
		Agricultural policy	2.6
		Other/ not specified	1.8
Economics	11.4	Agricultural economics	7.0
		Other/ not specified	4.4
Other	31.6		

% is the percentage of the 114 respondents who indicated research interest in each category

The type of institution from which respondents originated and their main research interest(s) were used to allocate respondents to categories. Firstly: government (N=16) or education (N=78). Secondly: social scientists (N=26), environmental scientists (N=66) or economists (N=22). Sample sizes for other institutions or disciplines were insufficient for comparative purposes (N<=10). Not all respondents were allocated to categories.

Table 3 lists the research programmes respondents had experience of. Jointly, respondents specified 26 past and present programmes of NERC, ESRC, EPSRC and BBSRC. Many respondents did not name individual research programmes, but instead indicated the funding body: 48% of respondents had experience of RCUK funding and 68% of respondents non-RCUK funding. The most common non-RCUK research programmes specified were funded by EU (37% respondents) and Defra (19% respondents). Some respondents (7%) also had experience of overseas funding.

Annex A. Questionnaire survey.

Table 3. Research programmes respondents had experience of (Q3).

Funding body	%	Programme	N
Research Council - joint	11	JAEP	4
		'Organic Farming Study'	1
		Land Use Programme (NELUP), NERC/ESRC	1
		Culture of Consumption, ESRC/AHRB	1
		Geographic Information Handling, ESRC/NERC	1
		Programme not specified	4
Research Council - NERC	29	GANE	9
		LOCAR	3
		URGENT	2
		LOIS	2
		GeneFlow	1
		Soil Biodiversity	1
		TSEC	1
		Earth Observation (LINK)	1
		EDGE	1
		AFI	1
		Environmental Diagnostics	1
		TIGER	1
		Programme not specified	9
		Research Council - ESRC	17
Science in Society	1		
Programme not specified	13		
Research Council - BBSRC	13	Biological Adaptation to Global Environmental Change (BAGEC), (previously AFRC)	1
		BIRE	1
		Programme not specified	11
Research Council - EPSRC	8	WITE	1
		EQUAL	1
		Sustainable Urban Environments (SUE)	1
		PURE	1
		SU:BRIM	1
		Programme not specified	3
Non - RCUK	77		

% is the percentage of the 114 respondents indicating they had experience of funders in each category
N is the number of respondents out of 114, indicating they had experience of each research programme

Annex A. Questionnaire survey.

DATA ACCESSED

Three quarters of respondents (74.6%) indicated that they required access to external datasets (i.e. those held outside of their organisation) to carry out their research, and 21.9% indicated that they didn't. The rest (4.4%) didn't know (Table 4).

Environmental science and government respondents were more likely than other categories to require access to external data.

Table 4. Number of respondents who require access to external datasets (i.e. those held outside their organisation) to carry out their research (Q4).

	Discipline			Institution		All
	Social scientists	Environmental scientists	Economists	Government	Education	
N	26	66	22	16	78	114
% Yes	69	74	68	88	71	75
% No	27	17	27	6	26	22
% Do not know	4	5	5	6	5	4

N is the number of respondents in each category

Table 5a lists the types of data currently used by respondents. The 'other' datasets listed by four individual respondents were: health, legal services, rural services and woodland. Table 5b lists the types of data that respondents would like to have access to (if they currently didn't use them). For each type of data, respondents were asked to indicate its source (external or internal to their organisation), type (time series, spatial or neither) and whether it was international.

The most widely used types of data were land use/land cover, agriculture/horticulture and administrative boundaries (>45% of respondents replying to this question), followed by soil, wildlife/biodiversity, landscape character, water quality, and weather/climate (32-36%). Datasets which respondents would like access to in the future followed a different pattern. Those categories to which fewer wanted access than were already using included the three most widely used currently (land use/land cover, agriculture/horticulture and administrative boundaries), and also soil, water quality and weather/climate. There were several categories for which a substantially higher number of respondents identified a future need than the number who were already using that category. These included social attitudes, social structure/social exclusion, recreation, energy and waste (Figure 1).

Annex A. Questionnaire survey.

Table 5a. The types of data currently used by respondents (Q5).

Data Type	Source		Type			Internat.	N	%
	External	Internal	Spatial	Time Series	Neither			
Land use/Land cover	33	21	36	21	4	15	51	60
Agriculture/horticulture	34	14	32	27	2	14	44	52
Administrative boundaries	35	11	33	11	1	12	41	48
Soil	23	10	20	7	2	9	31	36
Wildlife/biodiversity	20	17	18	12	1	4	30	35
Landscape character	20	8	16	5	1	4	28	33
Water quality	23	12	19	17	2	5	28	33
Weather/climate	23	10	21	20	2	9	27	32
Contamination/pollution	14	10	19	14	1	8	25	29
Topography	19	7	21	3	1	5	25	29
Economic activity/diversity	21	7	16	13	4	6	24	28
Demography/settlement pattern	19	7	21	11	2	4	22	26
Business/institutions/ governance	15	5	6	8	5	8	19	22
Food and drink	16	8	11	8	3	7	18	21
Water e.g. hydrology, waterways	16	6	15	7	1	1	18	21
Social attitudes	10	6	6	5	2	4	16	19
Social structure/social exclusion	12	5	11	8	1	4	14	16
Transport	11	2	7	3	2	3	13	15
Recreation	7	4	9	6	0	1	12	14
Air quality	10	5	10	7	1	3	11	13
Historic/archaeology	10	2	6	3	0	0	11	13
Energy	6	1	4	4	1	1	8	9
Fisheries	6	2	6	5	0	1	8	9
Waste	8	1	6	3	0	2	8	9
Disease - plant and animal	4	2	4	4	0	4	6	7
Coastal	5	2	3	1	0	0	5	6
Radioactive substances	2	1	1	0	0	1	2	2
Other	4	4	4	4	0	0	4	5

‘Neither’ means neither spatial or time series.

N is the number of respondents out of 114, currently using data of each type (in any category)

% is the percentage of respondents currently using data (N=85) of any type

Annex A. Questionnaire survey.

Table 5b. The types of data respondents would like to have access to (Q 5).

Data Type	Source		Type			Internat.	N	%
	External	Internal	Spatial	Time Series	Neither			
Land use/Land cover	26	10	24	25	1	12	36	55
Agriculture/horticulture	26	9	22	19	2	10	34	52
Wildlife/biodiversity	25	8	25	22	3	12	33	50
Landscape character	26	9	25	22	0	10	32	48
Economic activity/diversity	22	8	24	20	1	12	29	44
Contamination/pollution	24	5	17	19	1	12	28	42
Social attitudes	21	8	21	19	3	10	28	42
Soil	18	4	16	10	1	7	26	39
Water quality	20	8	17	16	0	8	25	38
Weather/climate	19	9	15	16	0	10	25	38
Recreation	21	4	21	14	1	4	24	36
Social structure/social exclusion	22	7	22	19	2	9	23	35
Water e.g. hydrology, waterways	17	9	16	12	0	7	23	35
Demography/settlement pattern	20	4	17	17	0	7	22	33
Business/institutions/ governance	16	6	13	14	2	7	21	32
Administrative boundaries	12	4	10	8	0	9	19	29
Topography	14	5	13	7	0	5	18	27
Waste	13	6	13	14	2	6	18	27
Disease - plant and animal	16	4	13	11	3	8	17	26
Transport	11	4	12	9	1	4	16	24
Energy	13	7	11	11	1	6	14	21
Air quality	12	3	10	10	0	3	13	20
Food and drink	6	3	9	7	1	5	10	15
Historic/archaeology	7	1	8	5	0	2	10	15
Coastal	7	1	7	2	1	4	9	14
Fisheries	8	2	5	4	1	3	9	14
Radioactive substances	4	1	2	2	0	1	5	8

'Neither' means neither spatial or time series.

N is the number of respondents out of 114 who would like to have access to data of each type (in any category)

% is the percentage of respondents who would like to have access to data (N=66) of any type

Annex A. Questionnaire survey.

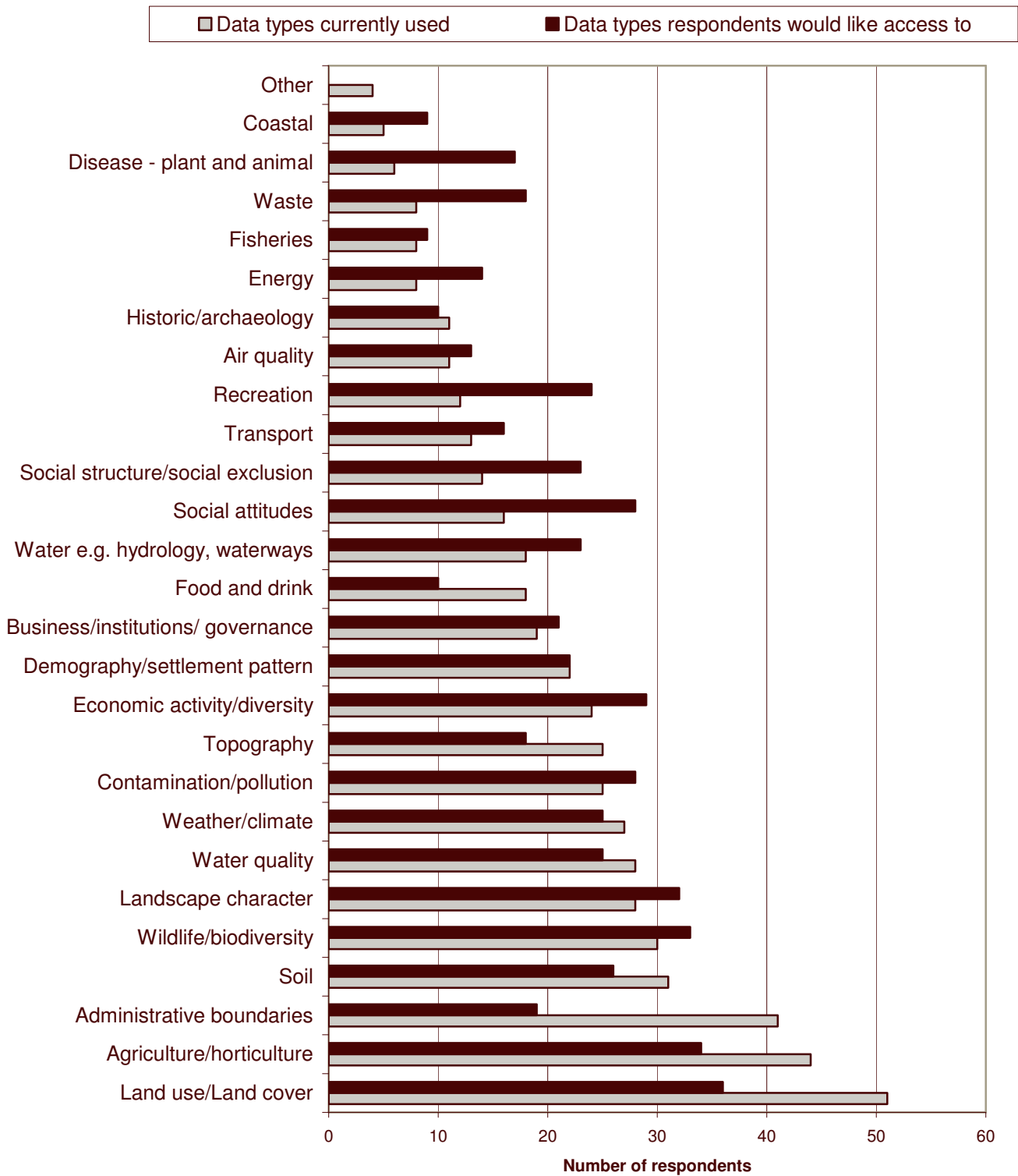


Figure 1. Comparison of data types currently used with those to which respondents would like access in the future (Q5).

Annex A. Questionnaire survey.

Respondents were asked to give up to ten examples of datasets they currently access, their source and access arrangement (Q6). Where the description (given by respondents) was not clear, an attempt was made to identify the datasets that were cited by more than two respondents, and these are shown in Table 6. This process required a certain level of knowledge of each dataset, so some may not have been correctly classified from the information given by respondents. The complete list is shown in Appendix B.

Digimap (Ordnance Survey via EDINA) was the most frequently accessed, by 21 respondents (21%), closely followed by Countryside Survey (CEH) with 20 respondents (20%). Population and agricultural census, river flow and meteorological data were also widely used. Apart from the population census, the majority of datasets identified were land use or environmental.

Digimap was accessed by respondents using a licence, and Countryside Survey by all four types of arrangement specified on the questionnaire form (free access, self-owned, licence, access rights). The datasets on Administrative Boundaries (OS via EDINA) and soil (NRSI, Cranfield University) were the only frequently used datasets that seemed to be accessed solely by licence.

Annex A. Questionnaire survey.

Table 6. Datasets most frequently accessed (Q6).

Dataset	Source	Access Arrangement	N	%
Maps: Digimap (Raster, Land-Line, Land-Form)	OS, via EDINA	licence	21	20.6
Countryside Survey	CEH	free access, self-owned, licence, access rights	20	19.6
Census of population	ONS	free access	15	14.7
Agricultural Census	Defra, SEERAD	free access, licence, access rights	15	14.7
River flow, quality, discharge	Environment Agency	free access, access rights	11	10.8
Meteorological data	Met Office	free access, licence	10	9.8
Common Bird Census/Breeding Bird Survey/nest records	BTO	free access	5	4.9
Designated sites, habitat inventories	English Nature, SNH, CCW	free access	5	4.9
Farm Business Survey	Defra	free access, access rights	5	4.9
Soil	NSRI, Macaulay Institute	licence, self-owned, Do not know	5	4.9
Climate	BADC	free access	4	3.9
MAGIC datasets	MAGIC website	free access	4	3.9
Administrative Boundaries	OS via EDINA	licence	3	2.9
Rural (agricultural) statistics	Defra	free access	3	2.9

N is the number of respondents

Key:	BADC	British Atmospheric Data Centre	EN	English Nature
	BTO	British Trust for Ornithology	NSRI	National Soil Resources Institute
	CCW	Countryside Council for Wales	ONS	Office for National Statistics
	CEH	Centre for Ecology and Hydrology	OS	Ordnance Survey
	Defra	Department for Environment, Food and Rural Affairs	SEERAD	Scottish Executive Environment and Rural Affairs Department
	EA	Environment Agency	SNH	Scottish Natural Heritage

Annex A. Questionnaire survey.

When asked how they discovered the datasets they used in their research (Q7), most respondents indicated they used colleagues (68% of 105 replying to this question) and/or the owning organisation website (55%). Data portals were used by 40% of respondents, almost a third used a library (31%) and only 10% used a data catalogue (Table 7). Six respondents did not reply to this question and three gave inappropriate answers, indicating they generated their own data. Four respondents used other methods: search engine/web, and experience/ research knowledge were specified. There appeared to be little difference between categories of respondents in how they discovered datasets.

Table 7. How respondents discovered the datasets they used in their research (Q7).

	Discipline			Institution		All
	Social scientists	Environmental scientists	Economists	Government	Education	
N	26	66	22	16	78	105
Colleagues %	50	64	55	69	62	68
Internet: owning organisation website %	46	50	45	56	51	55
Internet: data portal %	35	32	41	31	36	40
Library %	38	20	32	25	36	31
Catalogue %	12	8	9	19	6	10
Other %	15	6	14	13	9	6

DIFFICULTIES WITH ACCESS

When asked if they had any difficulty accessing relevant data sets (Q8), 38% indicated that they did, 49% that they didn't and 13% didn't know (Table 8). However, of the 85 respondents who required access to external datasets, 51% had difficulty. Environmental and economist respondents were more likely than social science respondents to have difficulties with access. Government respondents were more likely than Education respondents to have difficulties.

Table 8. Have you had any difficulty accessing relevant datasets? (Q8)

	Discipline			Institution		All
	Social scientists	Environmental scientists	Economists	Government	Education	
N	26	66	22	16	78	114
%Yes	23	41	41	63	31	38
%No	50	44	55	38	56	49
%Do not know	23	11	5	0	13	13

Those indicating they had difficulty were asked for more details: specific datasets, the nature of the difficulty and whether the difficulties were resolved (Q9). Table 9 lists the types of datasets that respondents had most difficulty accessing. Appendix C

Annex A. Questionnaire survey.

contains the full list. Most problems were encountered with Defra Agricultural Census and NSRI soils (nine of the 43 respondents experiencing problems).

The most common difficulty with Agricultural Census data was confidentiality, and this was solved for only one of the four respondents with this difficulty (by amalgamation of data). Difficulties with expertise/data structure were solved for one of the three respondents. In this case, Defra tailored the data to the respondent's requirements. Two respondents had problems with cost (to access the data via EDINA), and one solved it by finding the money eventually.

Difficulties accessing NSRI soils data were predominantly related to cost. This was solved for half the respondents, by paying the licence fee or being given a discount. The problem with expertise/data structure was resolved by the supplier.

Half the difficulties accessing Ordnance Survey datasets were related to cost (N=2). In contrast, water quality data difficulties seemed to be mainly related to obtaining information on its existence. Bird (BTO) (N=2) and landcover data (N=1) difficulties due to cost were not resolved, but a difficulty with weather data cost was, by obtaining an educational discount.

Table 9. The types of datasets that respondents had difficulty accessing (Q9).

Dataset	N	%	Problems
Agricultural Census data (Defra and EDINA)	9	20.9	Confidentiality (4), Expertise/data structure (3) Cost (2)
NSRI Soils and NATMAP	9	20.9	Expertise/data structure (1) Cost (8)
OS Map data	4	9.3	Expertise/data structure (1) Cost (2) Other (1)
Water Quality (EA and SEPA)	3	7.0	Expertise/data structure (1) Other (2)
BTO bird data	2	4.7	Cost (2)
Farm Business Survey	2	4.7	Confidentiality (1), Expertise/data structure (1)
Landcover Map	2	4.7	Cost (2)
LIDAR	2	4.7	Cost (1) Other (1)
Raw qualitative data, corporate interviews	2	4.7	Confidentiality (1), Other (1)
Weather data	2	4.7	Cost (1) Expertise/data structure (1)

% is the percentage of 43 respondents experiencing difficulties with access

Table 10 shows that the most common difficulties with data access were related to cost (44%), confidentiality (21%) and expertise/data structure (17.5%). Ownership difficulties were relatively infrequent (5%).

Table 10. Difficulties with accessing relevant datasets (Q9).

Difficulty	N	%
Cost	35	43.8
Confidentiality	17	21.3
Expertise/data structure	14	17.5
Ownership issues	4	12.5
Other	10	5.0
Total number of problems	80	

When asked if they would like to obtain access to additional datasets for their research (Q10), most respondents indicated that they would (67% of 58 replying to this question). Only seven (12%) indicated that they would not, the rest didn't know/it didn't apply (21%). Social scientist respondents were less likely than environmental scientist or economist respondents to want access to additional datasets. Government respondents were more likely than Education respondents to want access.

Table 11. Would you like to obtain access to additional datasets for your research? (Q10)

	Discipline			Institution		All
	Social scientists	Environmental scientists	Economists	Government	Education	
N	12	34	10	10	34	58
%Yes	42	76	70	90	65	67
%No	17	12	10	10	12	12
%Do not know	42	12	20	0	24	21

Respondents were asked to list up to five examples of datasets they would like to obtain access to (Q11), and 35 respondents named 77 datasets. These are listed in full in Appendix D. The most common datasets were those relating to soil - nine respondents listing soil data and three respondents NRSI datasets (Table 12). Apart from soils and climate, all datasets identified related to some aspect of land cover or land use.

Table 12. Datasets to which respondents would like to gain access for their research (Q11).

Type of data	N
Soil data, source not specified	9
Climate	5
Land use	5
Land cover	4
Land Cover Map (CEH)	4
Agriculture: Defra Ag.Census	4
Schemes	4
Aerial Photos	3
Satellite Imagery	3
Soil data from NSRI	3

Annex A. Questionnaire survey.

Respondents were asked if they anticipated any difficulty in obtaining access to these datasets (Q12). Two thirds (66%) indicated that they did, 6% that they did not, and the rest didn't know (29%). However, 84 respondents actually answered this question, despite not having listed datasets previously. When all replies were considered, nearly half (46%) anticipated difficulty, 19% didn't and 35% didn't know.

They were then asked to give details of the anticipated difficulty (Q13). Table 13 shows that the most common anticipated difficulty was related to cost (43%), followed by confidentiality (26%) then ownership issues (19%). Expertise/data structure was the least common (13%).

Table 13. Difficulties anticipated with obtaining access to datasets (Q13).

Difficulty anticipated	N	%
Cost	37	43.0
Confidentiality	22	25.6
Ownership issues	16	18.6
Expertise/data structure	11	12.8
Total difficulties	86	

Table 14 shows the datasets two or more respondents anticipated difficulties accessing. The complete list is given in Appendix E. The most frequently mentioned dataset (by 10 respondents) was soil (NSRI and Macaulay Institute), and the main difficulty anticipated was cost. Scheme uptake data (Defra, Forestry Commission etc.) were listed by five respondents and the difficulty anticipated was confidentiality.

Table 14. Datasets respondents anticipated difficulties obtaining access to (Q13).

Dataset	N	Difficulty anticipated
Soil	10	cost (6), confidentiality, ownership issues, Other: availability
Schemes: uptake	5	confidentiality
Biodiversity	4	cost, ownership issues, expertise/data structure, Other: existence of accessible dataset
Climate	4	cost (4), expertise/data structure
Land use	4	confidentiality, cost, ownership issues, expertise/data structure
Satellite data	3	cost, ownership issues, expertise/data structure
Water quality	3	confidentiality, cost, ownership issues, Other: availability of data
Aerial photographs	2	cost, Other: relies on a general Forestry Commission licence, Other: limited coverage of areas required
Agriculture: annual spatial information (crop areas)	2	confidentiality, cost
Atmospheric data: air quality parameters (gas fluxes and aerosols)	2	cost, Other: availability
Land cover	2	cost, ownership issues
Land Cover Map (CEH)	2	cost
LIDAR data	2	cost, Other: licensing and extent

Annex A. Questionnaire survey.

DATA INTEGRATION

Table 15 shows that a quarter of respondents indicated they currently integrate or use integrated datasets (57% didn't and 18% didn't know) (Q14). Environmental scientist and government respondents were more likely than other respondents to integrate or use integrated datasets.

Table 15. Do you currently integrate or use integrated datasets? (Q15)

	Discipline			Institution		All
	Social scientists	Environmental scientists	Economists	Government	Education	
N	26	66	22	16	78	114
%Yes	12	33	14	31	21	25
%No	81	42	64	50	62	57
%Do not know	4	20	23	19	18	18

Respondents were then asked to give examples of integrated datasets that they use (the two datasets and the scale of integration for spatial data) (Q15). Appendix F lists the 58 examples given by 43 respondents.

The datasets in the examples were categorised according to discipline ('environmental' or 'socio-economic') and the spatial unit categorised as 'landscape' (e.g. square kilometre, natural boundaries), 'political' (e.g. ward, county etc), or none. Table 16 shows that the majority of examples given by respondents were of integrated environmental datasets and only four of the examples were of integration between socio-economic datasets. Most were integrated using landscape rather than political spatial units. Nearly a third of the examples given (29%) involved inter-disciplinarity (between environmental and social data).

Table 16. Frequency analysis of the integrated datasets currently used by 43 respondents (Q15).

Spatial scale	Discipline			Total	%
	EE	ES	SS		
Geographical	32	7	0	39	67.2
Political	4	9	4	17	29.3
None	1	1	0	2	3.4
Total	37	17	4	58	
%	63.8	29.3	6.9		

Key: E environmental dataset
S socio-economic dataset

Respondents were then asked if they intended to integrate datasets in the future (Q16). Nearly a third (30%) indicated that they would, 18% that they wouldn't and the rest didn't know (Table 17). There appeared to be little difference between categories of respondents.

Table 17. Do you intend to integrate datasets in future? (Q16)

	Discipline			Institution		All
	Social scientists	Environmental scientists	Economists	Government	Education	
N	26	66	22	16	78	114
%Yes	27	30	32	38	29	30
%No	15	18	18	6	21	18
%Do not know	54	47	50	56	50	52

Respondents were asked to give examples of datasets they intended to integrate (Q17). The examples were categorised as before. Appendix G lists the 51 examples given by 37 respondents. Table 18 shows that again, the majority of datasets were environmental. However, a lower percentage of the examples given were interdisciplinary (12%) compared to the examples of current practice (29%).

Table 18. Frequency analysis of the datasets that 37 respondents intended to integrate in the future (Q17).

Spatial scale	Discipline			Total	%
	EE	ES	SS		
Geographical	28	2	0	30	58.8
Political	3	4	4	11	21.6
None	6	0	4	10	19.6
Total	37	6	8	51	
%	72.5	11.8	15.7		

Key: E environmental dataset
S socio-economic dataset

Most respondents did not know if they were likely to encounter difficulties when integrating datasets (62%) (Q18). However, 24% thought they would have difficulties and only 14% were confident that they would not (Table 19). Economist and government respondents were most likely to anticipate difficulties.

Table 19. Do you anticipate any difficulties when integrating datasets? (Q18)

	Discipline			Institution		All
	Social scientists	Environmental scientists	Economists	Government	Education	
N	26	66	22	16	78	114
%Yes	15	24	32	38	22	24
%No	12	17	5	13	15	14
%Do not know	69	55	64	50	63	62

The most commonly anticipated difficulties (Q19) related to spatial issues, including datasets collected at different spatial resolution and data reported in different output units (Table 20). Availability of, or access to appropriate data, confidentiality and cost were also commonly cited issues for data integration.

Table 20. Type of difficulties anticipated with integrating datasets in the future. (N = 25. Two responses were not relevant to the question).

Q19: if 'yes', please indicate what difficulties you expect to encounter	N
Reconciliation of datasets collected at different spatial scales or reported in different unit areas	9
Availability of/access to appropriate data	8
Confidentiality	6
Cost	4
Time for the integration	2
Compatibility	1
Time to generate spatial datasets and enter data	1
Aggregation of data to generate comparable datasets	1
Access to data in a suitable format	1
Dataset-specific problems	1

Table 21 shows the tools used by respondents for data integration (Q20). The most common tool, used by three quarters of respondents was a spreadsheet. Mapping/GIS and statistical software were used by half of the respondents. Databases (36%) and graphical software (24%) were less commonly used. CIS was used by 10% of respondents. Other tools specified included modelling, image analysis software, writing own software and cartography. Spreadsheets were used most by Environmental scientist respondents, mapping/GIS and statistical software was used less by Social scientist respondents and Economist respondents used graphical tools less and didn't use Database software at all. CIS (Countryside Information System) was used least by Environmental scientist respondents.

Table 21. Tools used for data management and integration (Q20).

	Discipline			Institution		
	Social scientists	Environmental scientists	Economists	Government	Education	All
N	26	66	22	16	78	113
% Spreadsheets	73	77	64	69	74	75
% Mapping/GIS	31	53	59	44	50	53
% Statistical software	23	39	36	50	36	50
% Database software	19	32	0	44	21	36
% Graphical	42	61	32	63	47	24
% Other or N/A	12	9	9	19	9	15
% CIS (Countryside Information System)	23	12	14	25	14	10

Only eight respondents (7%) indicated that there were processes that they could not carry out because the appropriate tools were not available (Q21). The other responses were evenly split between 'no' (49%) and 'don't know' (44%). Details of these eight responses indicate a range of difficulties (Table 22) (Q22). At one end of the spectrum is simply that appropriate software is not available to the researcher (e.g. GIS) or a lack of technical understanding of the possibilities for data integration.

More complex issues of data integration included the need to develop land use change tools that currently do not exist, 3D visualisation of landscape data and simulation of landscapes under different scenarios.

Table 22. Individual responses to question on availability of tools for data integration. (N = 8)

Q22: if there are data processes, which you would like to carry out, but are unable to because the tools are not available, what are they?
Unavailable tools
Unfamiliarity with what software might be used.
Difficulty in persuading the department to purchase the relevant statistical packages
GIS; Mapping
GIS
Processing of pdf files to create databases
We will need to develop land use change tools as they do not exist
Better spatial modelling of habitats in regard to management and bird densities.
3D visualisation of landscape data (integration of remotely sensed image data and LiDAR based elevation data) at very high spatial resolution (~1m).
Virtual reality simulation of rural landscapes under different management scenarios.

SERVICES RESPONDENTS WOULD LIKE RELU TO PROVIDE

The most commonly cited frustrations with data access and compatibility were: spatial scale, cost, confidentiality, format and time taken to access and compile data (Table 23) (Q23). Data produced at different spatial scales were a common problem both between and within subject areas and when comparing datasets from different countries.

Several respondents suggested that datasets collected with public money should ideally be freely available. It was pointed out that in the United States, data is freely available on the web. One respondent also commented that organisations that were the holders of major datasets were at a competitive advantage over others.

Several respondents found the process of getting access to data and compiling it in a suitable format to be very time consuming. This can be very frustrating, particularly where it is necessary to access the data simply to determine if it is appropriate. It was also pointed out that, although summary data are often available, it is difficult to establish if more detailed data are available to external bodies. One respondent also commented that, although data is theoretically available, lack of time for individuals to make it available can be a serious barrier to data acquisition.

One respondent raised the issue that there was too much emphasis on data rather than information.

Issues with the quality of datasets ranged from a lack of documentation (e.g. sampling procedures) and incomplete labelling to errors in the data.

Annex A. Questionnaire survey.

Table 23. Respondents' frustrations with data availability etc (Q23). Note that some categories may overlap (e.g. variation in format of data may be linked to time to compile data) (N = 73).

Response	N
Different spatial scales and lack of common spatial units	12
Cost	11
Confidentiality	10
Variation in format/structure	10
Time to access, compile data, decide if appropriate	9
Gaps in/lack of (appropriate) data	7
Lack of availability	7
Knowing what exists	5
QA/inadequate documentation	4
Where and how to access	2
Access restrictions (not specifically because of confidentiality)	2
Complex licensing agreements	2
Lack of ownership of major datasets	2
Incompatibility (different indicators/changes in definitions)	2
Software/web access	2
Lack of statistical knowledge/data management expertise	2
Time from collection to release	1
Information on structure and content	1
Concern about continuous reinvention of data integration	1
Scattered datasets	1
Estimates of errors	1
Emphasis on data not information	1
Nothing	1
N/A	5

Only 17 respondents quoted other research programmes as examples of good/poor practice in data management (Q24). These comments (or a shortened form) are reproduced in Table 24. A further four generated/managed their own data under other research programmes and one simply stated that data management could be improved for wastewater re-use research. Two respondents commented on aspects of specific studies rather than research programmes, but their responses reiterate the need for consistency of methods and formatting and the importance of establishing protocols before data collection begins.

Most positive comments were made about programmes where a proactive approach had been taken to data management, including provision of information on data ownership and access, and collation of data and metadata produced by the programme. Where improvements were suggested, it was apparent that some programmes had no data management policy. Respondents commented that some had made no attempts to co-ordinate communication between award holders to highlight synergies between projects and facilitate data access or exchange. Lack of archiving and licensing arrangements were also criticised.

Annex A. Questionnaire survey.

Table 24. Comments on good and poor aspects of data management under other research programmes (N = 17) (Q24).

Programme	Done well	Could be improved
BAGEC		No data management
Defra programmes		Negotiation of individual licences meant costs were unpredictable and could exceed fixed budget
EC/ESF	Flexibility granted	Support
EU Environmental Monitoring Level II data	Dedicated data management experts meant that data management (protocols, cleaning, storage) was superb	Continuity
European Commission	ACCESS II, IMPEL	Recurrent access is difficult aggravated by licences even if all is publicly funded
GANE	Data Management Plans and Coordination centre	
GANE	Central internet site giving dataset ownership and access rights information. Collation of data and associated metadata produced from studies under the GANE programme	More input into early phases of research with regard to the role of the GANE Data Centre
GANE	Central internet site giving dataset ownership and access rights information. Collation of data and associated metadata produced from studies under the GANE programme	More input into early phases of research with regards the role of the GANE Data Centre
GANE	Strong co-ordination of process	Little early discussion of potential synergies between projects and hence little adaptation of project data collection to optimise synergy
JAEP		Little attempt to co-ordinate communication, data access or exchange between award holders
JAEP		Integration and interdisciplinarity
LOCAR	Licences negotiated centrally	
LOIS	Central database set up	Timeliness of data becoming available.
LOIS	Collation of data from diverse sources in a common format.	Making data readily available to interested parties.
NERC data management fund	Supplied funding for the retrieval of past records and transfer of these on to modern media	
QLIF	The study invested time to develop methods to collect comparable data from different regions	
SEERAD Core	Funding for research data	Data archiving and sharing
TIGER, JAEP, FP5		Data archiving at programme level
United States programmes	Government derived data is free or minimal cost	

Annex A. Questionnaire survey.

Respondents were asked, if they already had or were to have a RELU award, what services they would like to see provided by RELU (under five categories) (Q25). In the first category (*information on the existence and availability of relevant datasets*), there was a general consensus of opinion that it would be very useful if RELU provided greater on-line information, and better metadata. Suggestions for the format of this information ranged from simple lists to web-based portals with search facility and links to sites holding the data. Many respondents simply stated that metadata would be useful. Of those that were more specific about what information they would like, availability/access was most important followed by where to find the datasets. Other suggestions for information on sources and availability included a consultation service and prices.

Table 25. RELU services: information on data sources/availability (N = 51) (Q25).

Category:	N
Information on datasets	
A simple 'yes' response	11
On-line information	
Web based system	11
Catalogue/list	8
Database	5
Simple overview with links to more detail	1
Links to sites holding data	3
Search/filter facility	2
Sample records	1
Metadata information	
Simply 'metadata'	14
Availability/access	10
Description	2
Contact details/where to find data	7
Others	
Cost	1
Consultation service	1
Key datasets made available	1

A total of 41 respondents felt that RELU could *help with data access and acquisition* (Table 26). The most common suggestion was for central provision or negotiation of key datasets. Several respondents thought that these datasets should be provided free, but it was also recognised that central negotiation would be more efficient than many individuals undertaking the same process. Five respondents reiterated that information on data sources/availability would be helpful and three that an internet portal through which datasets could be accessed would be useful. One suggested that this could be linked to an online license agreement. One respondent felt that the UK focus of the programme meant that RELU would not be able to help with data access.

Table 26. RELU services: help with data access/acquisition? (N = 42) (Q25).

Category:	N
Help with data access/acquisition	
A simple 'yes' response	11
Provide or negotiate access/licence centrally for key datasets (preferably free)	12
Information on data sources/availability	5
RELU helpdesk should fulfil requirements	3
Technical help (formatting, data extraction)	3
Internet portal	3
On line licence	1
Budget allowance for data acquisition (already possible)	1
Would like help in order to be able to compete with universities and government on an equal financial footing	1
Designated contact person within project	1
Help with issues of confidentiality	1
Workshops and seminars	1
Common data format	1
Unlikely given restricted UK focus of programme	1

Respondents were much less clear that RELU could help provide *tools for data integration* with only 21 positive responses out of 27 replying (Table 27), although three thought that software could be provided and two thought information about what was available would be helpful. Two respondents commented on how such tools should function – one felt that they should run on commonly used software (e.g. excel, access) whereas the other stressed that tools must be compatible with open source applications and not be restricted to Microsoft products.

Table 27. RELU services: tools for data integration (N = 27) (Q25).

Category:	N
Provision of tools for data integration	
A simple 'yes' or 'possibly' response	8
Software/GIS tools	3
Information on what is available and where	2
If such tools are developed then must be able to run on a range of platforms	2
Common tasks for common problems	1
Basic guide	1
Access to advice from someone who has done something similar	1
Dissemination of integrated datasets likely to be used by > 1 project	1
Common standards and geo-definition for subsequent data deposition	1
Integration of spatially explicit data with area classification (e.g. natural areas)	1
Consider need for/availability of training	1
Not sure this is possible for extensive high resolution spatial data	1
No (already have what required/develop in-house)	5

Respondents were enthusiastic about the possibility of RELU providing a *communication facility for interaction with other award holders*. It was felt that this would be particularly useful for sharing solutions to problems, looking for synergies/links between projects and as a route for information exchange in general. A web forum or email listing were the most common suggestions for such a facility (Table 28). Other possibilities included a newsletter, subject specific or general meetings and more information on the website to facilitate communication between projects.

Table 28. RELU services: communication facility for interaction with other award holders (N = 39) (Q25).

Category:	N
Provision of a communication facility	
A simple 'yes' response	13
Web forum/discussion board	10
Jiscmail or similar email list	6
Programme meetings (annual/occasional)	3
Emailed newsletter	2
More info on website (e.g. noticeboard of meetings, data requests)	2
Data/metadata database to facilitate collaboration/data access	2
Independent identification of links/synergies between projects	1
Not useful unless very similar projects	1
No	1

Most respondents felt that a *collaborative facility for data sharing with other award holders* would be valuable and one pointed out that it would be particularly useful for unpublished datasets (Table 29). However, there were concerns that this would be complicated to achieve, would have cost implications and that it would not be appropriate to make data widely available before it has been analysed and papers produced. The most common suggestion for such a facility was a (web based) data repository or central data management system. One respondent suggested that the existing UK Data Archive would be an appropriate mechanism. Others suggested that a metadata database would be sufficient or that data sharing could be facilitated by simply providing information on availability and access.

Table 29. RELU services: collaborative facility for data sharing with other award holders (N = 41) (Q25).

Category:	N
Provision of a collaborative facility for data sharing	
A simple 'yes' response	21
Data repository/central data management system (web based)	8
Facilitation of data sharing – information on availability and access	3
Metadata database	2
Concerns about IP	2
News sheet	1
Licenses negotiated centrally	1
Agreement for data sharing between Research Councils	1
Data management software	1
More appropriate at end of project if collecting data	1

Only four respondents had additional suggestions for services that RELU could provide. Two expressed concern about the conflict between data sharing and IP and one felt that RELU could help resolve confidentiality issues. One respondent felt that the Research Councils should be responsible for data management rather than individual programmes.

Finally, respondents were asked whether there were any other developments in data management that they would like to see implemented in RELU and similar programmes (Table 30) (Q26). Several respondents felt that datasets should be more

Annex A. Questionnaire survey.

readily available (or centralised) and should be at the lowest possible or no cost (particularly where both data and research is publicly-funded). However, there was some concern about data ownership and the issues raised by the Freedom of Information Act. The scale of reporting and format of data were felt to be important, both for ensuring data were available at an appropriate scale to answer questions, and to make processing more efficient. Advice on data management and statistical analysis would be helpful to some respondents, and one suggested setting up a list of participants who would be willing to share expertise within the programme.

Table 30. Other issues in data management that could be implemented in RELU (N = 16) (Q26)

Category:	N
Other developments respondents would like to see implemented in RELU	
Improved data availability (and cost)	3
Data reported at smallest scale/different units	3
Advice on data management/statistical analysis	2
Metadata – advice/central list	2
Standardisation of data types and formats	1
GRID computing to encourage data sharing and integration	1
Details of projects funded under the research programme (aims, plan)	1
Archiving of historical data	1
Feedback of recommendations on data use/collection to organisations funding collection	1
Standardised system of quality control and implementation	1
Concern about data ownership	1

Further work

Results from the questionnaire analysis indicate social scientists and economists are less well represented than environmental scientists. In the next phase of the scoping study we hope to target socio-economists to investigate why many thought the questionnaire was not relevant to them and had trouble completing it. In addition, we intend to consult specialists to investigate issues arising from the questionnaire (such as data policy in other countries, particularly the US) as well as to investigate trends and technological developments in data management.

Although respondents were less enthusiastic about data integration tools than other services that could be provided within a research programme such as RELU, responses revealed a degree of ignorance of the complexities and issues involved. We believe the ability to integrate data is a prerequisite for interdisciplinary research. For these reasons it was decided to hold a workshop on data integration. The workshop will include presentations on projects involving integration of data from socio-economic and natural science disciplines which will discuss the practical issues involved and approaches to resolving common difficulties. Information will also be provided on the RELU Data Support Service and feedback given from the questionnaire survey.

Appendix A. The questionnaire.

RELU Questionnaire

Please use this word document only if you are unable to complete the web based version.

Please return to Helen McKay, email: h.mckay@csl.gov.uk
or post to: Central Science Laboratory, Sand Hutton, York, YO41 1LZ

This questionnaire is being distributed to all contacts on the RELU e-mail list, as part of a RELU scoping study looking at issues involved in data management and the development of strategies for the future. This study is complementary to the activities of the RELU Data Management subgroup, and does not duplicate their activities. Whether or not you are in receipt of an award (or hope to be!), please answer the questions which are relevant with respect to research you currently conduct, and hope to carry out in future, within the Rural Economy and Land Use field. We thank you in advance for your help, which will enable us to develop proposals for improving data services both in the RELU programme and for similar programmes in the future.

Name:	
Address:	
Postcode:	
Organisation:	
Department	
e-mail:	
Telephone:	

Annex A. Questionnaire survey.

1. What is/are your main research interest(s)?

--

2. Please tick one of the following categories:

Government Research Institute	<input type="checkbox"/>
University/college	<input type="checkbox"/>
Research Council Institute	<input type="checkbox"/>
Charity/not-for profit institution	<input type="checkbox"/>
Other (please specify):	<input type="checkbox"/>

3. What other research programmes do you have experience of?

--

4. Do you require access to external datasets (i.e. those held outside your organisation) to carry out your research?

Please tick

1	Yes	<input type="checkbox"/>
2	No	<input type="checkbox"/>
3	Don't know	<input type="checkbox"/>

Annex A. Questionnaire survey.

5. The following question lists types of data.

Please select all data types that are relevant to you, and specify type if "Other"

For each, please could you:

i) consider whether the data are external or internal to your organisation, and indicate whether you currently use this type of data (option a), or if not, would like to have access to it (option b).

ii) consider whether the data are time series and/or spatial, and indicate a or b as before, and

iii) consider whether the data are international, and indicate a or b as before.

		External		Internal		Time Series		Spatial		Neither spatial or time series		International	
		(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
1	Administrative boundaries												
2	Contamination/pollution												
3	Radioactive substances												
4	Waste												
5	Demography/settlement pattern												
6	Social structure/social exclusion												
7	Historic/archaeology												
8	Transport												
9	Land use/Land cover												
10	Landscape character												
11	Social attitudes												
12	Economic activity/diversity												
13	Business/institutions/governance												
14	Recreation												
15	Agriculture/horticulture												
16	Food and drink												
17	Fisheries												
18	Energy												
19	Soil												
20	Disease - plant and animal												
21	Wildlife/biodiversity												
22	Topography												
23	Coastal												

Annex A. Questionnaire survey.

24	Water e.g. hydrology, waterways												
25	Water quality												
26	Weather/climate												
27	Air quality												
28	Other (please specify):												

Annex A. Questionnaire survey.

6. Please give examples of datasets you currently access.

Fill in up to 10 examples.

	a) What datasets do you currently access? <i>give name or description</i>	b) Source	c) Under what arrangement do you gain access? <i>1. free access 2. self-owned 3. licence 4. access rights</i>
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

7. How do you usually discover the datasets that you use?

Please tick

1	Catalogue	
2	Internet: owning organisation website	
3	Internet: data portal	
4	Library	
5	Colleagues	
6	Other (please specify):	

8. Have you had any difficulty accessing relevant datasets?

Please tick

1	Yes		(go to question 9)
2	No		(go to question 10)
3	N/A		(go to question 10)

Annex A. Questionnaire survey.

9. If the answer to the previous question is 'yes', please give the following details:

	a) Which datasets did you have difficulty accessing? <i>give name or description</i>	b) What was the difficulty? <i>1. confidentiality 2. cost 3. expertise/data structure 4. ownership issues 5. other (please state)</i>	c) Was the difficulty resolved? <i>1. yes 2. no 3. partially</i>	d) How was the difficulty resolved? <i>describe</i>
1				
2				
3				
4				
5				

Annex A. Questionnaire survey.

10. Would you like to obtain access to additional datasets for your research?

Please tick

1	Yes	<input type="checkbox"/>	(go to question 11)
2	No	<input type="checkbox"/>	(go to question 14)
3	N/A	<input type="checkbox"/>	(go to question 14)

11. If so, please list up to 5 examples:

1	
2	
3	
4	
5	

12. Do you anticipate any difficulty in obtaining access to any of these datasets?

Please tick

1	Yes	<input type="checkbox"/>	(go to question 13)
2	No	<input type="checkbox"/>	(go to question 14)
3	Don't know	<input type="checkbox"/>	(go to question 13 if you wish to add comments)
4	N/A	<input type="checkbox"/>	(go to question 14)

13. If the answer to the previous question is 'yes', please give the following details:

	Dataset	Difficulty anticipated: 1. <i>confidentiality</i> 2. <i>cost</i> 3. <i>ownership issues</i> 4. <i>expertise/data structure</i> 5. <i>other (please specify)</i>
1		
2		
3		
4		
5		

Annex A. Questionnaire survey.

14. Do you currently integrate or use integrated datasets?

Please tick

1	Yes	<input type="checkbox"/>	(go to question 15)
2	No	<input type="checkbox"/>	(go to question 16)
3	Don't Know	<input type="checkbox"/>	(go to question 16)

15. If the answer to the previous question is 'yes', please give examples of integrated datasets which you use. For spatial data, please give the scale at which the integration takes place (e.g. km², parish, census output area, etc.).

Dataset 1	Dataset 2	Spatial scale

16. Do you intend to integrate datasets in future?

Please tick

1	Yes	<input type="checkbox"/>	(go to question 17)
2	No	<input type="checkbox"/>	(go to question 20)
3	Don't know	<input type="checkbox"/>	(go to question 17 if you wish to add comments)

17. If the answer to the previous question is 'yes', please give examples of datasets which you intend to integrate. For spatial data, please give the scale at which the integration takes place (e.g. km², parish, census output area, etc.), if known.

Dataset 1	Dataset 2	Spatial scale (if known)

Annex A. Questionnaire survey.

18. Do you anticipate any difficulties in this process?

Please tick

1	Yes	<input type="checkbox"/>	(go to question 19)
2	No	<input type="checkbox"/>	(go to question 20)
3	Don't know	<input type="checkbox"/>	(go to question 19 if you wish to add comments)

19. If the answer to the previous question is 'yes', please indicate what difficulties you expect to encounter:

20. What tools do you use for data management and integration?

Please tick all that are relevant and specify type(s) if "Other"

1	Spreadsheets	<input type="checkbox"/>
2	Statistical software	<input type="checkbox"/>
3	Database software	<input type="checkbox"/>
4	Graphical	<input type="checkbox"/>
5	Mapping/GIS	<input type="checkbox"/>
6	CIS (Countryside Information System)	<input type="checkbox"/>
7	Other (please specify):	<input type="checkbox"/>

21. Are there data processes which you would like to carry out but are unable to because the necessary tools are not available?

Please tick

1	Yes	<input type="checkbox"/>	(go to question 22)
2	No	<input type="checkbox"/>	(go to question 23)
3	Don't Know	<input type="checkbox"/>	(go to question 23)

Annex A. Questionnaire survey.

22. If the answer to the previous question is 'yes', what are they?

--

23. What do you find most frustrating in terms of data availability/access/compatibility?

--

24. If you have experience of other research programmes, what aspects of data management were, in your opinion, managed well or less well

Research programme	Done well	Could be improved

Annex A. Questionnaire survey.

25. If you already have or were to have a RELU award, what data services would you like to see provided by RELU?

	Type of service
1. Information on data sources/availability	
2. Help with data access/acquisition	
3. Tools for data integration	
4. Communication facility for interaction with other award holders	
5. Collaborative facility for data sharing with other award holders	
6. Other (please specify).	

26. Are there any other developments in data management that you would like to see implemented in RELU and similar programmes?

Thank you for your help in completing this questionnaire.

Appendix B. Datasets currently accessed (respondents were asked to give up to ten examples) (Q6).

N=number of respondents

	Dataset	Source	Access Arrangement	N
1	Access	Local Authority	free access	1
2	Access to Services	Countryside Agency	free access	1
3	Administrative Boundaries	OS via EDINA	licence	3
4	Aerial photographs	Environment Agency	free access	1
		Various	license	1
5	Agricultural Census	Defra	free access, licence, access rights	14
		SEERAD	free access, licence	4
		AgraEurope	access rights	1
		not given	free access	1
6	Agricultural Statistics	FAO	free access	1
7	Agri-environment Scheme uptake	Defra (RDS)	access rights	1
8	Air Quality	EA	free access	1
		Air Quality Archive, AEA Technology	free access	1
9	AMADEUS	via university library	licence	1
10	Ancient woodland	EN, SNH etc	free access, licence	2
11	Archaeological Site Management Record	Gwynedd Archaeological Trust	licence	1
12	Areas of organic land	Defra Statistics	free access	1
13	Beilstein Chemical Reactions	http://www.mimas.ac.uk/crossfire/	access rights	1
14	Benefits Data	Department of Works and Pensions	free access	1
15	Biodiversity	Local Record Centre	free access	1
		Biological Records Centre	free access	1
		CEH	free access, self-owned	1
		National Biodiversity Network	free access	2
		County Records Centres	licence	1
		OXON council	free access	1
		None Given	self-owned	2

Annex A. Questionnaire survey.

Appendix B . continued

	Dataset	Source	Access Arrangement	N
16	Botanical data	cranfield university	self-owned	1
17	British Household Panel Survey	UK Data Archive	do not know	1
18	Business Activity x Contacts	Private Databases	free access	1
19	CACI Paycheck Data on Household Income	CACI	licence	1
20	Catalogues of entomopathogenic microorganisms	on line	free access	1
21	Catchment boundaries	CEH	self-owned,licence	1
22	Census of population	ONS	free access	15
23	Census of population boundaries	UKBORDERS	access rights	1
24	Chemistry Databanks	Daresbury Laboratories	access rights	1
25	Climate: UK	EU ATEAM	access rights	1
		Environment Agency	free access	1
		BADC	free access	4
		Self-generated	self-owned	1
26	Climate: European Climate data (50 km grid)	JRC/DG Research MARS data base	access rights	1
27	Climate change scenarios	Tyndall Centre	do not know	1
28		UKCIP02	licence	2
29	Common Bird Census/breeding bird survey/nest records	BTO	free access	5
30	Common land	Defra	free access	1
31	Community Environmental Valuations	Questionnaires	self-owned	1
32	Companies House records	Dun and Bradstreet	licence	1
33	Corine landcover	CEH	self-owned,licence	1
34	Corporate and NGO interviews	Personal negotiation	free access	1
35	Council Tax and Housing Benefit Information	Local authorities	access rights	1
36	Countryside Information System	CEH	free access	1
37	Countryside Survey	CEH	free access, self-owned, licence, access rights	20
38	County wide reports	County Council & inward investment agency	free access	1

Annex A. Questionnaire survey.

Appendix B . continued

	Dataset	Source	Access Arrangement	N
39	Crop production & yields	Ministry of Agriculture, Government of Botswana & FAO	free access,access rights	1
40	Crop trials	SAC, various	access rights	1
41	CSO Neighbourhood Statistics	CSO	free access	1
42	DEM (90m)	EROS data centre	free access	1
43	Designated areas	Countryside Agency	free access	1
44	Designated sites, habitat inventories	English Nature, SNH, CCW	free access	5
45	DeTR acid deposition maps	Through library	licence	1
46	Digital Terrain Model	CEH	self-owned	1
47	Digitized river network England and Wales	CEH	self-owned	1
48	Economic Accounts for Agriculture	EUROSTAT	free access	1
49	Economic Regeneration	Scottish Neighbourhood Statistics	free access	1
50	Economic Time Series	CSO/treasury	free access	1
51	Employment	NOMIS	do not know	1
52	Energy use	IEA	do not know	1
53	ESA	CEH	free access	1
54	EU Level II environmental monitoring data	EU and Forest Research	free access	1
55	European Bioinformatics Institute	http://www.ebi.ac.uk/	free access	1
56	Expasy databanks including genomic information	http://us.expasy.org/	free access	1
57	Expenditure and Food Survey	ESDS	free access	1
58	FAAM aircraft data	BADC	access rights	1
59	Family expenditure survey	uk data archive	free access	1
60	Farm Accounting Data Network	EC	free access, access rights	2
61	Farm Business Survey	Defra	free access, access rights	5
		ESRC Data Archive	access rights	2
62	Farm Enterprise Costs/Income	Data booklets	licence	1
63	Farm Practices Survey	Defra	free access	1

Annex A. Questionnaire survey.

Appendix B . continued

	Dataset	Source	Access Arrangement	N
64	Farmer attitudes	None Given	self-owned	2
65	Fertilizer use	Defra	free access	1
66	Fisheries	FRS	do not know	1
		FAO	free access	1
67	Food Chain Data	IGD and AgraRurope	free access	1
68	Food Safety Incidents	FSA/USDA/FDA/ Central Labs	access rights	1
69	Food Survey	Defra	free access	1
70	Forest Enterprise Sub-Compartment Database	Forestry Commission	free access,licence	1
71	FSA IID DATA	FSA	free access	1
72	Funding announcements	GDN	free access	1
73	Gamebag Census, Partridge Count Scheme database	GCT	self-owned	2
74	Gamekeeper survey	GCT, NGO, CA, SGA, MGA	self-owned	1
75	Genomics - eg BLAST	NCBI	free access	1
76	GLC2000	ISPRA	free access	1
77	Global Assessment of Soil Degradation (GLASOD)	FAO	free access	1
78	Global Entrepreneurship Monitor (various data bases)	London Business School	free access	1
79	Global meterological	UAE	access rights	1
80	Global soils	ISRIC	free access	1
81	Greenhouse gas emmisson factors and emmissions	various, ipcc, internal	self-owned	1
82	Gross/Net farm income breakdowns	SAC	licence	1
83	Habitat Maps	Devon and Cornwall Wildlife Trusts	free access	1
84	Health Data	Public Health Institute for Scotland	free access	1
85	Historic Features (SAMs, Listed Buildings, etc.)	English Heritage	free access	1
86	Historical field data	BADC	access rights	1
87	HOST	CEH	self-owned	1
88	House Price Data	Land Registry	free access	1
89	Household Expenditure Survey	CSO	free access	1

Annex A. Questionnaire survey.

Appendix B . continued

	Dataset	Source	Access Arrangement	N
90	Household survey data	Own research	self-owned	1
91	HPA CSSS	HPA	free access	1
92	Human Census - spatial and attribute	ESRC and JISC	free access	1
93	Hydrology data	environment agency	free access	1
94	Hydrometric data from CEH	Internet	free access	1
95	IAHS household statistics	EUROSTAT	access rights	1
96	IGER grass experiments	IGER	access rights	1
97	IMD2004	internet	free access	1
98	Import/Export/Trade	AgraEurope	free access	1
99	Income Data	Inland Revenue	access rights	2
100	Index of multiple deprivation	Welsh Assembly Government	free access	1
101	Information on desertification in the Mediterranean	EU-funded MEDACTION project	free access,self-owned	1
102	Infoterra	UK Satellite Image Data Service, via MIMAS/JISC	licence	1
103	Internal project monitoring	Self generated	self-owned	1
104	International and national economic data	various, OECD,USDA, HMT etc.	free access,access rights	1
105	International Passenger Survey	not given	access rights	1
106	JNCC Macrophyte Database	JNCC	licence	1
107	Land cover	Macaulay Institute	self-owned	1
108	Land use	EU ATEAM	access rights	1
		North York Moors National Park	licence	1
		None Given	self-owned	2
109	Land Use and Soils (1 km)	European soils/ land use data base	access rights	1
110	Landesvermessungsamt data from German Laender	Landesermessungaemter	licence	1
111	Landmap	UK Satellite Image Data Service, via MIMAS/JISC	licence	1
112	LANDMAP 25m DEM	www.landmap.ac.uk (no cost access up to Sept 2004)	access rights	1

Annex A. Questionnaire survey.

Appendix B. continued

	Dataset	Source	Access Arrangement	N
113	Landowner Contact Details	NFU, Yellow Pages, BT, Maps, Local People	free access, access rights	1
114	Leader observatory reports	EC	free access	1
115	Leisure Day Visits Survey	not given	access rights	1
116	Library Information	IC libraries	free access	1
117	Local Government fact sheets	Local Authority websites	free access	1
118	Local Knowledge	internet	licence	1
119	LOIS Data centre	NERC	free access	1
120	MAGIC	website	free access	4
121	Mammal Survey	Mammal Society/University of Bristol	access rights	1
122	Maps: Bartholomew 1:200k UK	Bartholomew (CHEST)	licence	1
123	Maps: Digimap data (Raster, Land-Line, Land-Form)	Ordnance Survey, via EDINA	licence	21
124	Meteorological data	Met Office	free access, licence	10
125	Migation Data	Health Authorities	access rights	1
126	Mintel reports	University licence	licence	1
127	Mortality and morbidity data	ONS, ISD	free access	1
128	Nation Diet and Nutritional Survey	Internet	free access	1
129	National Inventory of Woodland and Trees	Forest Research	self-owned	2
130	NCBI	http://www.ncbi.nlm.nih.gov/	free access	1
131	Nextmap 5m Digital Elevation Model (DEM)	NERC/NEODC	access rights	1
132	NOx , Ozone, SO4	EMEP	free access	1
133	Nucleotide sequence information	GenBank	free access	1
134	Nutrient Data Laboratory	USDA	free access	1
135	Oceanographic Data	BODC	free access	1
136	OS GPS Base station network	OS	access rights	1
137	PCT	Relevant PCT	access rights	1
138	Pdb - enzyme structural databank	http://www.rcsb.org/pdb/	free access	1
139	Pesticide Usage Survey	CSL	self-owned, licence	2

Annex A. Questionnaire survey.

Appendix B . continued

	Dataset	Source	Access Arrangement	N
140	Planning Permissions	Local Planning Authorities	access rights	1
141	Population dynamical time series and associated data	self-owned	self-owned	1
142	Postcode descriptors	Commercial	free access	1
143	Postcode directory: all fields	ONS Geography via MIMAS	licence	1
144	Prices, production, trade, input use, for 18 countries and commodities	various, oecd, fao, internal	self-owned	1
145	Property restitution data for Germany and Poland	Bundesamt fuer Restitution	do not know	1
146	Public Perceptions of Pesticides	MORI	free access	1
147	Qualitative socio-economic	focus groups	self-owned	1
148	Rainfall: Botswana	Meteorological Services, Government of Botswana	free access	1
149	RDA Observatories	website	free access	1
150	Regio	Eurostat	licence	1
151	Reindeer populations and parasite datasets	CEH	free access	1
152	Remote sensing: airborne	NERC Airborne Remote Sensing Facility (ARSF)	free access	1
153	Remote sensing: satellite	Various	access rights	1
154	River flow, quality, discharge	Environment Agency	free access, access rights	11
155	River Network	CEH/SEPA	licence	1
156	River water quality	SEPA	free access, licence	2
157	Rothamsted Park Grass Experiment	Rothamsted	access rights	1
158	Rural (agricultural) stats	Defra	free access	3
159	Scottish deer culls	Forest Enterprise Scotland	access rights	1
160	Scottish Executive on line surveys	Scottish Executive	free access	1
161	Scottish Native Woodland Inventory	SNH	licence	1
162	Sewage Treatment Works Data	EA/Water Companies	free access	1
163	Social Attitudes	ONS	free access	1

Annex A. Questionnaire survey.

Appendix B . continued

	Dataset	Source	Access Arrangement	N
164	Social Housing Data	Local Authorities and Housing Associations	access rights	1
165	Social indicators	NZ statistics	free access	1
166	Soil	NSRI, Cranfield University	licence	4
		self generated	self-owned	2
		Environment Agency	access rights	1
		Macaulay Institute	do not know, self-owned	3
		Through library	licence	1
		EU JRC ESB Bilateral collaboration	licence	1
167	Spectroscopy databanks	http://www.aist.go.jp/RIODB/SDBS/menu-e.html	free access	1
168	State of Countryside Reports	Countryside Agency	free access	2
169	Sussex Study Database	GCT	self-owned	1
170	Tesco clubcard	dunnhumby	access rights	1
171	The Composition of Foods	McCance and Widdowson (book)	free access	1
172	Tide Height	Hydrographic Office	free access	1
173	Time series data (economic/consumer/price trends)	ONS	free access	1
174	Topographic data	Environment Agency	licence	1
175	Tourism: businesses	Local Tourist Boards	free access	1
176	Tourism: UK Tourism Survey	not given	access rights	1
177	UK Livestock compound feed composition by animal type	Defra	free access	1
178	UK precipitation - chemical quality	Acid Waters Monitoring Network	free access	1
179	UKMO surface station	BADC	access rights	1
180	Urban Extent	OS	free access, licence	1
181	Waste process locations	Environment Agency	free access	1
182	Water and sanitation information	International Water Association	access rights	1
183	Water quality	CEH	free access	2
		Samples	access rights, self-owned	2

Appendix B . continued

	Dataset	Source	Access Arrangement	N
184	Water resources	Waternet	access rights	1
185	Weather	ARCMET	licence	1
186	Weather	Rothamsted	free access	1
187	Welsh National Phase I survey	CCW	licence	2
188	Wetland Inventories	English Nature	access rights	1
		Self-generated	self-owned	1
189	Wheat growing areas of UK	HGCA	free access	1
190	Various	UNEP GEO Data Portal	free access	1
		CIESIN	free access	1

Appendix C. Difficulties in accessing datasets (Q9).

	Dataset	Difficulty	Solved?	Solution
1	Agricultural Census data (parish)	confidentiality	No	Up to 1988, visit the PRO!
2	Agricultural census data spatially explicit data not available below parish level	confidentiality	No	N/A
3	Agricultural data	confidentiality	No	N/A
4	Agricultural data	cost	No	N/A
5	Agricultural data	expertise/data structure	No	N/A
6	Annual Agricultural Census	expertise/data structure	Yes	Data tailored by MAFF to catchment
7	Bird 10km Atlas data	cost	No	N/A
8	BTO bird data	cost	No	N/A
9	CACI Paycheck data	cost	Yes	Pay fee
10	Census	expertise/data structure	No	N/A
11	Climate change data	expertise/data structure	partially	Communication
12	Comparable maps for England, Wales, and Scotland	Other - not sure they all exist	No	I just took what they had and tried to make it work
13	Corporate interview data	confidentiality	No	N/A
14	Council tax and housing benefit	confidentiality	partially	Depends on organisation
15	Crop trials data	Other - Combination of 2,3 & 4	No	N/A
16	CSSS	expertise/data structure	Yes	Contact
17	Data on pesticide contamination levels, transport infrastruxcture, income levels etc disaggregated to local administrative units.	expertise/data structure	No	N/A
18	Data on pesticide contamination levels, transport infrastruxcture, income levels etc disaggregated to local administrative units.	ownership issues	No	N/A
19	Defra Annual Agricultural Census	confidentiality	partially	Amalgamation
20	Digitised River Network	cost	No	N/A
21	Digitised soil and landcover	cost	partially	Inter-institute negotiation
22	EDINA Agcensus	cost	Yes	Money found (eventually)

Annex A. Questionnaire survey.

Appendix C. continued

	Dataset	Difficulty	Solved?	Solution
23	ERDP Scheme uptake	confidentiality	partially	Data obtained for certain schemes
24	FADN in different countries	ownership issues	partially	Formal request, co-operation with local researchers.
25	Farm Accounts Survey Data - spatially explicit data not available	confidentiality	No	N/A
26	Farm Business Survey	confidentiality	partially	No access to spatial information
27	Farm Business Survey	expertise/data structure	No	Only used published tables
28	Fertiliser Survey	ownership issues	No	N/A
29	Get Mapping National Aerial Photographs	cost	No	N/A
30	Health board sub sets	cost	partially	Negotiated access, alternative material
31	Hydrologically Corrected DEM	cost	No	N/A
32	IACS	confidentiality	No	N/A
33	IEA energy use	cost	partially	Wrote for specific data and obtained printed information
34	IGD, AGRAEUROPE	cost	No	Beyond budget
35	IID	expertise/data structure	Yes	Contact
36	Imagery	cost	No	N/A
37	ISD	cost	partially	Negotiated access, waived payments
38	Land Use and crops (DEFRA)	expertise/data structure	No	N/A
39	Landcover Map at more refined level	cost	No	N/A
40	Landowner details	confidentiality	partially	Careful sifting of various public repositories, plus some help from organisations able to release data to me
41	LCM 2000	cost	No	Too costly
42	LIDAR	cost	No	N/A

Annex A. Questionnaire survey.

Appendix C. continued

	Dataset	Difficulty	Solved?	Solution
43	LIDAR data collected by the Environment Agency	Other - unsure of who to approach	No	N/A
44	Livestock census data	confidentiality	Yes	I obtained a research permit in my own name, rather than using that of my PhD supervisor
45	Local Knowledge	cost	Yes	Negotiation with senior management about the usefulness of the data.
46	MAFF / Defra farm census	confidentiality	No	Analysis put on hold
47	Mastermap for study sites	cost	No	N/A
48	Meteorological data	Other - missing values	No	Filled in with data from nearby station
49	Meteorological Data	cost	No	N/A
50	MLURI Soils data for Scotland	cost	partially	Paid high cost for licence
51	NASA WorldWind data	Other - the server cannot cope with demand	partially	Better servers?
52	NATMAP	cost	Yes	Data provided for extraction cost by NSRI for NERC PhD student
53	NSRI Soils data for England and Wales	cost	partially	Paid high cost for licence
54	ONS	cost	No	N/A
55	Ordnance Survey Master Map data	expertise/data structure	partially	Massive data set, currently served to PCs but limited in application.
56	OS digimap	cost	Yes	UWB agreed to centrally fund agreement
57	Raw qualitative data	Other - having the time to chase it down	No	I have a job to do without doing any research too.....so I have never got round to this
58	Restitution data	confidentiality	Yes	Negotiation

Annex A. Questionnaire survey.

Appendix C. continued

	Dataset	Difficulty	Solved?	Solution
59	River discharge (EA)	expertise/data structure	partially	Continued communications and questions to resolve problems of data format and identity
60	Satellite imagery for heather cover	cost	No	N/A
61	Satellite Images of the UK	cost	No	N/A
62	Sewage Treatment Works Data	Other - data set not coordinated	Yes	Liaison with EA who telephoned individual water suppliers for information. limited to only 50 major works
63	Social housing	confidentiality	partially	Depends on organisation
64	Social information	Other - knowing where to get the right information	partially	Gave up in some cases, used data of lesser value in others
65	Soil (NSRI)	cost	No	N/A
66	Soil Association Maps for UK	cost	partially	N/A
67	Soil data	expertise/data structure	Yes	Resolved by supplier
68	Soil maps of UK	cost	Yes	Discount
69	Soil Survey	cost	partially	Research funder agreed to pay licence costs
70	Soil survey	cost	No	N/A
71	Soil Survey Data	cost	No	N/A
72	Standards protocols	confidentiality	partially	Personal contacts
73	UK Postcode Centroids	ownership issues	No	N/A
74	UK soils datasets	cost	No	NA
75	Various	Other - knowing they existed	partially	Luck, searching and asking people we know
76	Water Quality	Other - finding it/contacting the EA	Yes	Contact finally established

Annex A. Questionnaire survey.

Appendix C. continued

	Dataset	Difficulty	Solved?	Solution
77	Water quality (EA)	expertise/data structure	Yes	Continued communications and questions to resolve problems of data format and identity
78	Water Quality data	Other- not easily attained as data is spread out over SEPA	No	N/A
79	Weather (long-term average values)	expertise/data structure	partially	Mapped data obtained for certain variables
80	Weather data	cost	Yes	Negotiated educational discount and reduced the resolution of data
81	Woodland Grant Scheme applications individual applicant data not available	confidentiality	partially	Limited access, but unsatisfactory in that only limited analysis could be undertaken.

Annex A. Questionnaire survey.

Appendix D. Datasets respondents would like access to (Q11).

	Dataset
1	Aerial Photos (National)
2	Access to restitution data has been ad hoc, depending on cooperation at local level. It would be a great help to have blanket access to case files.
3	Additional ONS, ISD and NHS data
4	Aerial Photos
5	Aerial Photos: Getmapping Millennium Aerial photographic coverage
6	Agriculture: annual spatial information on agriculture (crop areas)
7	Agriculture: Defra Census
8	Agriculture: DEFRA land holding info - census
9	Atlas data for birds, bees, Orthoptera and Syrphidae
10	Botswana: time series bush cover data
11	BTO bird data
12	Climate
13	Climate: easier to access climate change
14	Climate: long-term climatic data, spatially referenced
15	Climate: Rainfall, temperature, potential evapo-transpiration (daily)
16	Countryside Survey (CEH)
17	Crop Trials
18	CROW Act Access Land (Countryside Agency)
19	Digitised River Network (CEH)
20	Economic data
21	FADN in New Member States of the EU
22	Farm Business Survey
23	Farm Business Survey with spatially relevant information
24	Farm-based work
25	Fertiliser application
26	High resolution data both spatially and temporally referenced

Annex A. Questionnaire survey.

Appendix D. continued.

	Dataset
27	Historic data
28	Historical meteorological data (site specific)
29	Hydrologically corrected DEM (CEH)
30	Land and property ownership
31	Land Cover
32	Land Cover 2000
33	Land Cover data
34	Land Cover data
35	Land Cover Map (CEH)
36	Land Cover Map at more refined level
37	Land Use
38	Land Use
39	Land Use
40	Land Use and crops
41	Land Use data at high resolution
42	Landcover (UK) CEH
43	LCM 2000
44	LIDAR
45	LIDAR data
46	Mastermap for study sites
47	Pesticides: national coverage, application /use data
48	Pollution: more detailed data
49	Pollution: various information required to estimate levels of diffuse pollution and or carbon sequestration
50	Rainfall chemistry - AWMN, ECN, NETCEN
51	Remote sensing data (NIR-data)
52	River discharge
53	Satellite Imagery

Annex A. Questionnaire survey.

Appendix D. continued.

	Dataset
54	Satellite imagery for heather cover
55	Satellite Images of the UK
56	Schemes: Agri-Environment scheme uptake and related data
57	Schemes: Rural Stewardship Scheme Application Data
58	Scottish Forestry Grant Farmland Premium Scheme Application Data
59	Scottish Forestry Grant Scheme Application Data
60	Social
61	Social attitudes
62	Social data
63	Soil
64	Soil data from NRSI
65	Soil quality
66	Soil survey digitised data
67	Soil: pH, organic matter content, base cations, major anions, cation exchange capacity, Fe, Mn, Al, soil series, texture, soil water storage capacity, hydraulic conductivity, soil depth, soil water available to plants
68	Soils: 100 m grid for soils data
69	Soils: Macaulay Institute Scottish Soils
70	Soils: NATMAP and SOILSERIES from NSRI
71	Soils: physical characteristics of
72	Soils: site specific data
73	Soils: site specific measurements of leaching and erosion for model calibration
74	Soils: site specific measurements of soil water balance for model calibration
75	Soils: UK datasets
76	Species distribution data (Biological Records Centre)
77	Water Catchment Boundaries

Appendix E. Datasets respondents anticipated difficulties in obtaining access (Q13).

N	Dataset	Difficulty
1	Aerial Photographs: Get Mapping	cost, Other, Relies on a general Forestry Commission licence
2	Aerial Photographs: orthophotographic cover	Other, limited coverage in these areas
3	Agriculture (crop rotations and phenology)	Other, Availability
4	Agriculture: annual spatial information (crop areas)	confidentiality, cost
5	Air quality parameters (gas fluxes and aerosols)	Other, Availability
6	Atmospheric data	cost
7	Biodiversity	Other, existence of accessible dataset
8	Biodiversity: atlas data	cost
9	Biodiversity: Biological Surveys	ownership issues
10	Biodiversity: species distribution data (Biological Records Centre)	expertise/data structure
11	Botswana: time series bush cover data	Other, Current remote sensing techniques are not able to distinguish adequately between bush and grass cover on Kalahari sands
12	British Household Panel Survey - easily accessed, but rural identifiers only available up to 1999	Other, availability of rural identifiers
13	BTO bird data	cost
14	Business activity by location over time	Other, will need to be interpolated, I suspect
15	Catchment Boundaries	ownership issues
16	Climate	expertise/data structure
17	Climate	cost
18	Climate data	cost
19	Climate: detailed met data	cost, expertise/data structure
20	Companies House (costly to get large samples)	cost
21	Countryside character	Other, existence of accessible dataset
22	CROW Act Access Land (Countryside Agency)	expertise/data structure
23	Deer population estimates	Other, Already in use by others = IP problems

Annex A. Questionnaire survey.

Appendix E. continued.

N	Dataset	Difficulty
24	Digitised River Network (CEH)	cost
25	Ecological data	ownership issues
26	FADN in New Member States	ownership issues
27	Farm Business Survey with spatial information	confidentiality
28	Food: FSA data on nutrient content	confidentiality
29	Food: manufacturers nutritional information	confidentiality
30	HBS in other countries	confidentiality, cost, expertise/data structure
31	Health surveys	confidentiality, expertise/data structure
32	Health: NHS	confidentiality, cost
33	High resolution data both spatially and temporally referenced	confidentiality, cost, ownership issues
34	Hydrologically corrected DEM (CEH)	cost
35	IEA energy data	cost
36	Land and property ownership	confidentiality, cost, ownership issues
37	Land cover	cost, ownership issues
38	Land Cover Map/Countryside Survey (CEH)	cost
39	Land use	expertise/data structure
40	Land use	confidentiality
41	Land use	cost, ownership issues
42	Land use and crops	ownership issues
43	Land use: holding level data	confidentiality
44	Land Cover Map at more refined level	cost
45	LIDAR	cost
46	LIDAR data	Other, Licensing and extent
47	Nutrition surveys	expertise/data structure
48	ONS-ISD	confidentiality, cost
49	OS Mastermap for study sites	cost
50	Pesticides: national coverage of application /use data	cost

Annex A. Questionnaire survey.

Appendix E. continued.

N	Dataset	Difficulty
51	Pesticide application data	confidentiality
52	Pollution	expertise/data structure, Other, Resolution
53	Population data	ownership issues
54	Restitution data	confidentiality
55	River discharge	ownership issues
56	Satellite data	ownership issues, expertise/data structure
57	Satellite imagery for heather cover	cost
58	Satellite Images of the UK	cost
59	Schemes: Agri Environment Scheme uptake	confidentiality
60	Schemes: ERDP Scheme uptake data (Defra-RDS)	confidentiality
61	Schemes: Rural Stewardship Scheme Application Data	confidentiality
62	Schemes: Scottish Forestry Grant Farmland Premium Scheme Application Data	confidentiality
63	Schemes: Scottish Forestry Grant Scheme Application Data	confidentiality
64	Soil	cost
65	Soil data	ownership issues
66	Soil data from NRSI	cost
67	Soil quality (nutrients)	cost
68	Soil Survey	cost
69	Soil Survey (Cranfield University)	cost
70	Soil: high spatial resolution data (better than current 1km)	Other, Availability
71	Soil: physical characteristics	ownership issues
72	Soils: 100 m grid data	confidentiality, cost
73	Soils: Macaulay Institute Scottish Soils	cost
74	Supermarket Till data	confidentiality, cost
75	Topography	cost, ownership issues, expertise/data structure
76	Tourism	cost, Other, scale and time-series

Annex A. Questionnaire survey.

Appendix E. continued.

N	Dataset	Difficulty
77	Water quality	Other, availability of data
78	Water quality	ownership issues
79	Water quality data	confidentiality, cost

Appendix F. Examples of integrated datasets used by 29 respondents, and for spatial data, the scale of integration (Q15).

	Dataset 1	Dataset 2	Spatial scale
1	Administrative boundaries	Land Cover Map	ward (NUTS Level 5)
2	Agriculture: agricultural census	Land cover	parish
3	Agriculture: agricultural census	Administrative boundaries	parish
4	Agriculture: agricultural census	own survey	catchment
5	Agriculture: agricultural productivity data	GLASOD data	district
6	Agriculture: agricultural productivity data	Qualitative data from farmer interviews	village, Sub-District,
7	Agriculture: crop	Soils	plot
8	Agriculture: crop/Livestock	Gross/Net margins	farm
9	Agriculture: Defra agricultural census	SURPOP population	5km2 or 200 m
10	Agriculture: Parish June Ag. Census	Agriculture: EDINA Agcensus	2x2 km grid
11	Airborne remote sensing data	DEMs	1:2500 - 1:25000
12	Airborne remote sensing data	OS vector data	1:2500 - 1:25000
13	Airborne remote sensing data	LCM 2000 landcover	1:10000 - 1:50000
14	ATEAM Land Use	Climate: UKCIP	10' x 10'
15	Benefits Data	Health Data	SNS Datazone/ Unit Postcode
16	Benefits Data	Household Data	SNS Datazone/Unit Postcode
17	Botanical data	Soil: soil type	hectare
18	BTO bird distribution data	Agricultural data	10-km
19	BTO bird distribution data	LCM 2000	1km or 10-km
20	Building address	Population census	electoral ward
21	CCW Phase I	LCM 2000	Wales (10m resolution)
22	Census	Topography	census output area
23	Climate	Soils	plot/Farm
24	CTD	Trace Metal Geochemistry	km2
25	Enterprise Costs	Land use	approaching 400 km2
26	Environment Data	Demographic data	SNS Datazone

Annex A. Questionnaire survey.

Appendix F. continued.

	Dataset 1	Dataset 2	Spatial scale
27	ESB Soils data	ATEAM Land Use	10' x 10'
28	EXPERIAN 2000	1991 Census EDs	1991 Census EDs
29	Farmer Attitudes	Wildlife Populations	
30	Gamekeeper Survey	SSSI	area overlap - km2
31	Gamekeeper Survey	SPA	area overlap - km2
32	Gamekeeper Survey	Landcover map from CIS at 1 km scale	area overlap - km2
33	Gamekeeper Survey	Common Land in England	area overlap - km2
34	GLASOD data	Remotely sensed vegetation cover & grazing capacity	district
35	GLASOD data	Qualitative data from farmer interviews	village, sub-district, district
36	Land use	Runoff	100 km2
37	Land use	Sediment transfer	100 km2
38	Land use cover	Soil	km2
39	Landscape character / DEM / landcover / access	Landscape character / DEM / landcover / access	approximately 1:50,000
40	Land use	climate/weather	regional
41	LIDAR DEM (1m)	Nextmap (5m)	1:2500 - 1:25000
42	National Inventory of Woodland and Trees	Ancient Woodland Inventory	England (counties and natural areas)
43	Nutrient deposition	Land management info	per ha
44	Parish restitution data	Parish restitution data	parish
45	Partridge Count Scheme Database	Landcover map from CIS at 1 km scale	area overlap - km2
46	Phase 1 habitat cover	Topographic drainage sub-catchment	sub-catchment
47	Rainfall	Qualitative data from farmer interviews	village, sub-district,
48	Rainfall	Agriculture: agricultural productivity data	village, sub-district, district, national
49	Rainfall	River flow	
50	Regio	Agriculture: agricultural census	region
51	River levels	Climate: weather data	hectare

Annex A. Questionnaire survey.

Appendix F. continued.

	Dataset 1	Dataset 2	Spatial scale
52	SAFFIE Inverts	SAFFIE weed	field
53	Soil	Topography	catchment
54	Soils	Land cover	km2
55	Soils	Land cover	100 m2
56	Stream order (reach)	Riparian wetland (buffer zones)	farm, sub-catchment, catchment
57	Terrain (e.g. slope category)	Hydrological (e.g. runoff)	catchment
58	Water Quality (river channels)	Wetland inventory (delineations of functional types)	sub-catchment, catchment

Appendix G. Examples of datasets that 37 respondents intended to integrate, and for spatial data, the scale of integration (Q17).

	Dataset 1	Dataset 2	Spatial scale
1	Agricultural census	Land Cover	hectare
2	Agricultural census	Breeding Bird Survey	km ²
3	Agricultural census data	Neighbourhood statistics	census output area/ward
4	Airborne Remote Sensing data	DEMs	1:2500 - 1:25000
5	Airborne Remote Sensing data	LCM 2000 landcover	1:10000 - 1:50000
6	Airborne Remote Sensing data	OS vector data	1:2500 - 1:25000
7	Aphids	Parasitoids	farm
8	Beneficial insects	Climate: meteorological	km ²
9	Botanical data	Soil type	hectare
10	BTO bird distribution data	Agricultural data	10-km
11	BTO bird distribution data	LCM 2000 (CEH)	1km or 10-km
12	Building address	Population census	electoral ward
13	Business statistics register	Farm Business Survey	
14	CEH	Water quality	50m x 50m (10x10 if possible)
15	Countryside character areas	National inventory of woodlands and trees	parish
16	Countryside Survey	Agriculture: Defra census	ideally, sub 1km
17	Countryside Survey	Breeding bird survey	1km
18	Crime Data	Demographic Data	unit postcode
19	Demographics	Farm labour use	census output area
20	Ecological data	Farm practice	field level
21	EU F6 Analytical results from studies at Reading	EU F6 Analytical results from studies at other partners in Europe (8 partners in all)	
22	Expressed Preference surveys	Till data	
23	Farm Business Survey	CIS	km, parish and census output area
24	Farm Business Survey	Countryside Survey	smallest feasible

Annex A. Questionnaire survey.

Appendix G. continued.

	Dataset 1	Dataset 2	Spatial scale
25	Farm Business Survey	Soil Survey	km2
26	FES	Nutrition/health data	household
27	Fish farm location	LHCC and practice based data (health)	LHCC and PCT areas
28	FSA funded project Analytical results from studies at Reading	FSA funded project Analytical results from studies at other UK partners (5 partners in total)	
29	House price	Population census	full postcode
30	Imagery	Topography	catchment
31	Land Cover (LCM)	Natural Areas/Countryside Character Area	
32	Land use	Climate, nutrient inputs	farm/catchment
33	Land Use (Ag Census)	Natural Areas/Countryside Character Area	
34	Land use (change) data	Hydrological characteristics (esp. flooding)	river (section); possibly catchment
35	Landscape character / DEM / landcover / access	Landscape character / DEM / landcover / access	approximately 1:50,000
36	Lidar DEM (1m)	Nextmap (5m)	1:2500 - 1:25000
37	National Diet and Health	Household Expenditure Survey	
38	National IO table	CIS	census output area
39	Pesticide application data	Pollution data	
40	River levels	Weather data	hectare
41	Rural Services	House Price	full postcode
42	Soil	Land use	field to catchment
43	Soil data	Own survey	catchment
44	Soil quality	Land use	region
45	Soil/Climate	Hydrology	Plot/Farm
46	Soils	Land Cover	100 m2
47	Species Distribution (Biological Records Centre)	Natural Areas/Countryside Character Area	
48	Surface temperature	Vertical pollutant distribution	metres
49	Survey of Personal Incomes	Farm Business Survey	
50	Topography	Runoff	100 km2

Annex A. Questionnaire survey.

Appendix G. continued.

	Dataset 1	Dataset 2	Spatial scale
51	Water quality	Land use	catchment