



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



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

<http://reludata.csl.gov.uk>

*Data integration*

**Nigel Boatman, Alistair Murray, Helen M<sup>c</sup>Kay,  
Naomi Jones, James Aegerter & Robert Stones**

**Central Science Laboratory**



## Background: aim

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

# Background: **methods**

- Target
  - The RELU research community
  - Wide variety of individuals were contacted
  - Not **confined** to award holders
- Questionnaire design
  - Based on Defra horizon scanning study questionnaire
  - With input from RELU office
- Timing
  - 2 months window for responses from 7 December 2004
  - 2 reminders + non-respondents surveyed
- Response rate
  - 17% (114 completed questionnaires)
- Representativeness
  - 60% non-respondents felt it was not relevant
  - Rest mainly 'lack of time'

# The sample

## Origin of respondents:

University/College	68%
Government Research Institute	14%
Charity/not-for profit institution	9%
Other	5%
Research Council Institute	4%

## Main research interests:

Environment	43%
Rural Development	31%
Farming and Food	27%
Economics	11%
Other	32%

## Categorisation of sample

		N	%
Institute	Education	78	68%
	Government	16	14%
Discipline	Environmental scientists	66	58%
	Social scientists	26	23%
	Economists	22	19%

# Data integration

- 25% currently integrate data or use integrated datasets

	Discipline			Institution		
	Social scientists	Environmental scientists	Economists	Government	Education	All
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

## Data integration: current and future

- In the future, socio-economist respondents intend to use integrated data more

	Social scientists	Environmental scientists	Economists	Government	Education	All
N	26	66	22	16	78	114
% Yes	12 27	33 30	14 32	31 38	21 29	25 30
% No	18 15	42 18	64 18	50 6	62 21	57 18
% Do not know	4 54	20 47	23 50	19 56	18 50	18 52

## Types of data integrated

- A third of the integrated datasets currently used by respondents are from different disciplines

Units	Discipline			Total	%
	EE	ES	SS		
Landscape	32	7	0	39	67
Political	4	9	4	17	29
None	1	1	0	2	3
Total	37	17	4	58	
%	64	29	7		

# Types of data integrated: **future** intentions

- In the future, fewer integrated datasets used by respondents will be from different disciplines

Units		Discipline			Total	%
		EE	ES	SS		
Landscape	Current	32	7	0	30	59
	Future	28	2	0		
Political	Current	4	9	4	11	22
	Future	3	4	4		
None	Current	1	1	0	10	20
	Future	6	0	4		
Total	Current	37	17	4	51	
	Future	37	6	8		
%	Current	64	29	7		
	Future	72	12	16		

# Data integration: a quarter of respondents anticipated difficulties

	Discipline			Institution		
	Social scientists	Environmental scientists	Economists	Government	Education	All
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

## Data integration: the nature of anticipated difficulties (N=25)

- 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 to integrate (2)

# Tools for data management and integration

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

## Further work

- Consultation
  - Social scientists
  - GIS/statisticians
  - Data managers
  - Data providers
  - Computer scientists

# Consultation: some preliminary results

- Methods used for integrating data :
  - Put data into common framework  
e.g OS Mastermap, LCM2000, British National Grid
  - Historical with environmental data:  
put on paper, using maps, scanning, geo-correcting,
  - Produce digital data sets, matched data sets.

# Consultation: some preliminary results

- Issues:
  - Reconciling differences between datasets e.g. scale, units, raster/vector
  - Availability of information which would indicate differences between datasets
  - Lack of information on methods used to collect, interpolate or summarise data
  - Availability of information on quality
  - How to map qualitative data