DATA MANAGEMENT PLAN

PROJECT DETAILS

Project Reference	
Project Title	Assessing the potential risk of, and possible responses to, zoonotic diseases on the development of recreational use of British forests and wild-lands (Full title) Assessing and communicating animal disease risks for countryside users (Short title)
Principal Investigator	
Address	
Telephone Number	
Email Address	
Start date agreed with ESRC	1st September 2007
Completion date	31st August 2010

DATA MANAGEMENT

This section of the Plan covers overarching issues concerning data management. Its concerns all qualitative and quantitative data generated by the project. Before completing the section please read the RELU Data Management Policy, a copy of which can be found at http://www.relu.ac.uk).

Please refer to the accompanying Guidance Notes in order to be able complete the form properly.

Award holders will be required to provide full metadata* together with a description of the datasets which their project generates. The technical arrangements for data management and archiving (including decisions concerning final archiving destination for project datasets; formats for supply of data; licence agreements; IPR etc.) will need to be subsequently agreed with the RELU Data Support Service (DSS-RELU@essex.ac.uk). Award holders will be required to meet/liaise with DSS at appropriate intervals to report progress, resolve problems and exchange information.

* Metadata is the information necessary to interpret, understand and use a given dataset without reference to the original data collector, for example the dataset's title, summary, principal investigator and so on. Example metadata entries are provided in the guidance notes.

1. REQUIREMENTS FOR ACCESS TO EXISTING DATASETS

Please list and describe any existing datasets which will need to be acquired for the research to be carried out (third party data sources). Please also identify any specific issues relating to access to these data and how you will overcome any difficulties.

Dataset name	Description of dataset	Data owner / source	Access issues
Land cover, land use, and vegetation type			
Land Cover Scotland 1988 (LCS88)	Remote sensed dataset derived from aerial photography taken in 1988; provides broad habitat definitions at 1:25000 scale. GIS vector (polygon) dataset at 0.1m resolution.	Macaulay Institute	Forest Research have licence agreement
Land Cover Map 2000 (LCM2000)	Satellite derived remote-sensed datasets providing broad habitat definitions. GIS vector (polygon) dataset at 0.1m resolution.	Centre for Ecology & Hydrology	Forest Research have licence agreement
National Inventory of Woodland and Trees	Derived from LCS88 dataset plus updated to 1995 from Forestry Commission sources; provides info on broadleaved/conifer woodland > 2ha and small woods and trees (0.1-2ha)	Forestry Commission	Full access as part of Forestry Commission
Forester and sub-compartment dataset	National forest estate woodland data; provides detailed information on woodland composition	Forestry Commission	Full access as part of Forestry Commission

Woodland expansion areas	Potential areas for new woodland creation	Forestry Commission	
Vegetation management	Identifies strategies for managing vegetation density and proximity to footpaths	Various – appropriate to be determined	
Land use type	Can indicate management prescriptions	e.g. Macaulay Institute	
NBN Gateway	Species and habitat data, including National Vegetation Classification, Phase 1 Habitat Classification	National Biodiversity Network	Full access allowed through www.searchnbn.net
Topography, geography and climate			
Ordnance Survey Pan-Government product portfolio	Products include: 1) for large scale mapping - OS MasterMap; Land-Line; 1:10000 Scale Raster; 2) for small scale mapping – 1:50000 Scale Colour Raster; 1:50000 Scale Gazetteer; 1:250000 Scale Colour Raster; Strategi; Meridian 2	Ordnance Survey	All partners should have access www.ordnancesurvey.co.uk
Climatic data	A description of precipitation, temperature, frost, etc. which may influence the distribution of carrier species such as ticks.	e.g. Tyndall Centre for climate change research	
Climate change predictions	High / low scenarios	e.g. Tyndall Centre for climate change research	
Ticks, hosts and disease			
Incidence of reported zoonotic disease	Actual cases of Lyme disease and unconfirmed cases; if possible to be used as patient sample	Health Protection Agency	Potential confidentiality issues
Tick abundance	Numbers of ticks sampled under controlled procedures on selected sites	Oxford University	
Animal host abundance	Number of hosts to determine hotspots and potential trends	e.g. Deer Commission for Scotland	
Visitors, other social data			
Forest visitor surveys	Provides visitor profile information and numbers	Forestry Commission	FR should have access to raw data, plus public reports from FC website
Forest visitor numbers	Number of visits to non-FC woodlands	Various - appropriate to be determined	
Human population numbers	Projections for increases in the number and type of people who might visit countryside areas	National statistics office, Govt actuary's dept.	
Holiday / recreational data	Projections for trends in the number and type of people who might visit countryside areas currently and for future scenarios	Various - appropriate to be determined	
Forestry Commission staff / forest workers	Numbers of staff involved in outside activities	Forestry Commission	

2. DATASETS TO BE PRODUCED

Please list both quantitative or qualitative data that will be collected or generated by the project, together with a brief summary of each dataset (description and methodology), its format and how it will be managed and stored. Please add extra rows where necessary.

Dataset name	Summary (description /methods)	Format	Dataset management and storage
Lyme Patient Interviews	To collect information on the perceptions, attitudes, emotions and behavioural responses of Lyme disease patients	MS Word document, analysed in NVivo	All versions of digital data (text and audio recording to be stored on University networked drive
Lyme Patient Questionnaires	To collect information on the perceptions, attitudes, emotions and behavioural responses of Lyme disease patients	SPSS and analysed in NVivo	Scanned using 'Eyes and Hands' software. All versions of digital data (text and audio recording to be stored on University networked drive
Focus group	To collect views on scenarios, public perceptions, and behavioural responses. Will be used to guide the survey design (see below)	MS Word document, analysed in NVivo	Surrey data: All versions of digital data (text and audio recording to be stored on University networked drive Forest research centre data: data stored on network drive and automated remote back-up
Questionnaire survey with woodland workers and visitors	Questionnaires given to Forestry Commission (FC) rangers, managers and other organisations. Additional 400 questionnaires to be handed out at FC visitor centres		Questionnaires to be scanned and stored in networked drives
Tick abundance	Fine–scale estimates from study sites in this project	Excel & GIS	To be kept on backed-up computers at University of Oxford and the project's central data management server.
Climatic sub-sets	Sub-sets of existing climatic data – may be manipulated to produce new datasets	Excel & GIS	To be kept on backed-up computers at University of Oxford and the project's central data management server.
Habitat quality	Determined from landcover and management data	GIS	
Projective and prospective Lyme disease risk	Derived from scenario modelling	GIS, database	
Dynamic risk maps	Biological models of tick abundance driven by climate surfaces and host abundance within a GIS	GIS	To be kept on backed-up computers at University of Oxford and the project's central data management server.
Prototype short-term early warning system	Datasets representing spatially explicit climate, weather, host abundance and seasonally variable human activities	GIS	To be kept on backed-up computers at University of Oxford and the project's central data management server.

3. QUALITY ISSUES

Please briefly describe the procedures for quality assurance that will be carried out on the datasets (Quality issues to be addressed could for example include: documenting the calibration of instruments, the collection of duplicate samples, data entry methods, data entry validation techniques, methods of transcription).

Quality Assurance protocols should be derived for all data collection and processing stages to ensure data are collected as consistently as possible and filenames are standardised.

At Forest Research, the procedures laid down by the FR Quality Assurance Scheme will be followed; this accords with the Defra Joint Code of Practice.

At Surrey the procedures laid down by the University of Surrey code of practice (http://www.surrey.ac.uk/SHS/ethics/documents/good-practice.htm) will be followed.

At Oxford, tick abundance data and predictive risk maps will come from the project's own activities, and will be appropriately documented with a metadata scheme.

4. DATA BACK-UP PROCEDURES

Please describe the data back-up procedures that you will adopt to ensure the data and metadata are securely stored. For example: 'Recognising the susceptibility of hard disks to failure, collected digital data will be transferred on a weekly basis to IOMEGA Zip disks, which will be stored in the University fire safe.' Methods of version control should also be stated.

Data must be stored either on each institution's back up server or on a separate data storage device that is kept in a secure and fireproof location, separate from the main data point.

In practice, this will be achieved –

At Forest Research, by researchers storing copies within fire proof safes, and through use of the automated network back-up.

At Surrey all data is fully backed up onto Ultrium LTO2 tapes - incremental backups are taken Mon-Thurs, full server backups are taken over Fri/Sat/Sun. Tapes are securely stored in a separate building. In addition all hard drives attached to the server are in RAID 5 arrays.

At Oxford, by researchers using the University regular back-up system (Tivoli Storage Manager) and storing copies on external hard disks and computers housed separately from main data point.

5. DATA AVAILABILITY

The Research Councils require all RELU data to be made available for long-term, post-project management within the Research Councils' data centres so they can be made available for secondary research. Do you envisage any difficulties in making any data available, for example, for access constraints or licence conditions of third party datasets used? If so, how might these difficulties be overcome?

Derived data should be made available to RELU; original datasets, such as Ordnance Survey, will be subject to usual licensing issues that need to be respected in considering issues of data availability outwith the project.

Surrey - all survey data will comply with ESRC requirements. Qualitative data will be available once appropriate changes have been made to honour assurances of confidentiality and anonymity

6. DATA COPYRIGHT/ IPR OWNERSHIP
Please state who owns the copyright/IPR of the datasets that you have collected.

The terms of copyright for data collected by partners within this project is defined in the project collaboration agreement.

7. DATA MANAGEMENT RESPONSIBILITIES
Please identify the first point of contact for data management issues, including metadata and quality issues. If different people are responsible for different datasets, please specify below.

1. REQUIREMENTS FOR ACCESS TO EXISTING DATASETS

If you intend to use datasets from third parties (for example land cover, soil or census data) then please record these here. For each dataset specify its name, Please a brief description, the data owner or source any access issues you perceive in acquiring it, such as cost or license restrictions. RELU is currently collating a list of key third party datasets that researchers have requested, and may try to negotiate easier access for the Programme. However, this is likely to be a slow process and it **should not** be assumed that data will be available at reduced cost or restriction-free within the period of the programme.

2. DATASETS TO BE PRODUCED

Please provide details of the datasets you expect to produce arising from the project. For each proposed dataset, please use the Summary column to describe briefly how the data will be collected (description and methods) the format in which it will be collected and an indication of how it will be managed and stored. If standard methods or procedures are to be used in the production of datasets, please either provide references for these or else a brief description of the methodology and how it will be stored for submission with the datasets at a later date. Please note that statistical and model outputs with descriptions should also be treated as datasets.

3. QUALITY ISSUES

There are various guiding documents on project management and quality issues that can be consulted for best practice. For example, the DEFRA document http://www.defra.gov.uk/science/documents/QACoP_V8.pdf and the requirements of the international standard BS ISO 9001:2000 that might need to be considered for some kinds of data. The following list are issues that may be of specific importance to data production and management

- o Project planning and research design, including for example:
- o Setting realistic milestones and re-evaluating them at intervals
- o Collection of duplicate sample data
- o Comparison with known standards
- Documentation of procedures and methods, including:
  - o Version control of amended documents
- Maintenance and calibration of equipment
- The maintenance of electronic and paper records of outputs, including:
  - o Labelled primary data and accompanying metadata
  - o Labelled processed information (and its connection to primary data)
  - o The explanation of non-standard acronyms
  - o The inclusion of measurement units as required
  - o Validation of the transfer from paper records to electronic format
  - o Version control of amended data

Please indicate briefly your strategy for managing the quality issues you see as most important to your project. See **DSS Guidelines to Data Management** for specific examples.

4. DATA BACK-UP PROCEDURES

Please specify your data back up-procedures. All data should be backed-up to prevent its loss, for example through hard disk failure, virus infection or theft. Where IT support is limited, making regular copies of your data on CD and keeping them at a separate location may be considered a useful back-up procedure. Version control procedures should be covered to ensure that working and final versions of data are kept track of.

5. DATA AVAILABILITY

Please list any access constraints or license conditions relating to the datasets that you have collected, with reference to the Data Management Policy on the RELU website at http://www.relu.ac.uk/about/Data%20Management%20Plan.pdf. It is anticipated that where possible, all data will be shareable within the RELU or wider research community. Issues such as informed consent and confidentiality should be addressed upfront to ensure that further usage is not precluded. Where use of third party data (such as OS or census data) have been used to derive new data products, you must state any restrictions of further access and data sharing.

6. DATA COPYRIGHT/ IPR OWNERSHIP

Please state who owns the copyright and IPR of the datasets that you have collected and of any data products you have created utilising any third party data sources. Joint copyright can be held by different funding bodies/sponsors.

3.7 DATA MANAGEMENT RESPONSIBILITIES

Please indicate who will be the first point of contact for data issues. This is the person/s responsible for data management, metadata production, dealing with quality issues and the final delivery of final datasets to the designated Data Centre. If several key people are responsible for these matters, please list their names and data management responsibilities. The persons responsible should be familiar with the **DSS Guidelines on Data Management**.