

4th DRAFT

**The Lowland Catchment Research Programme
Data Management Plan**

Roger V. Moore
C. Isabella Tindall

Centre for Ecology and Hydrology
Institute of Hydrology
Wallingford
Oxon
UK

Tel: +44 (0)1491 838800
Fax: +44 (0)1491 692424
E-mail: rvm@ceh.ac.uk
E-mail: cit@ceh.ac.uk

CONTENTS

1. INTRODUCTION.....	3
BACKGROUND.....	3
PURPOSE OF THE PLAN	3
2. APPROACH.....	3
STRATEGY.....	3
THE DATA COMMITTEE.....	5
<i>Terms of Reference</i>	5
<i>Membership</i>	5
<i>Meetings</i>	5
DATA CENTRE	5
<i>Terms of Reference</i>	5
<i>Role</i>	6
<i>Membership</i>	6
DATA POLICY	6
IDENTIFYING REQUIREMENTS FOR DATA	8
DATA ACQUISITION	8
DATA MANAGEMENT.....	9
ACCESS TO DATA.....	9
LONG TERM SECURITY	9
PUBLICATION	9
EXPLOITATION	10
DATABASE DEVELOPMENT	10
3. DELIVERABLES	10
4. PROGRAMME OF WORK	10
5. RESOURCES	11
STAFF.....	11
CAPITAL ITEMS	11
<i>Computing</i>	11
<i>Data</i>	11
RECURRENT	11
6. DATA CENTRE WORK PLAN AND COST SUMMARY 2000.....	11
APPENDIX I - OPERATIONAL PROCEDURES.....	13
FINDING OUT WHAT DATA EXIST	13
REQUESTING DATA.....	13
SUBMITTING AND RECEIVING DATA FROM THE PI'S	13
RECEIVING DATA FROM THE FIELD PROGRAMME.....	13
APPENDIX II - PROGRAMME OF WORK	14

1. INTRODUCTION

BACKGROUND

The Natural Environment Research Council (NERC) believes its data to be an important national and corporate asset that must be maintained and exploited for the benefit of the country as a whole and science in particular. Therefore, NERC Data Policy requires that all its programmes have a planned approach to the assembly, management, long term security and exploitation of their data and that appropriate resources are allocated to carry out the plan.

PURPOSE OF THE PLAN

This plan sets out the overall approach to be adopted for data management in the Lowland Catchment Research Programme (LOCAR). It does not attempt to define matters of detail, which will be the responsibility of the LOCAR Data Committee.

2. APPROACH

STRATEGY

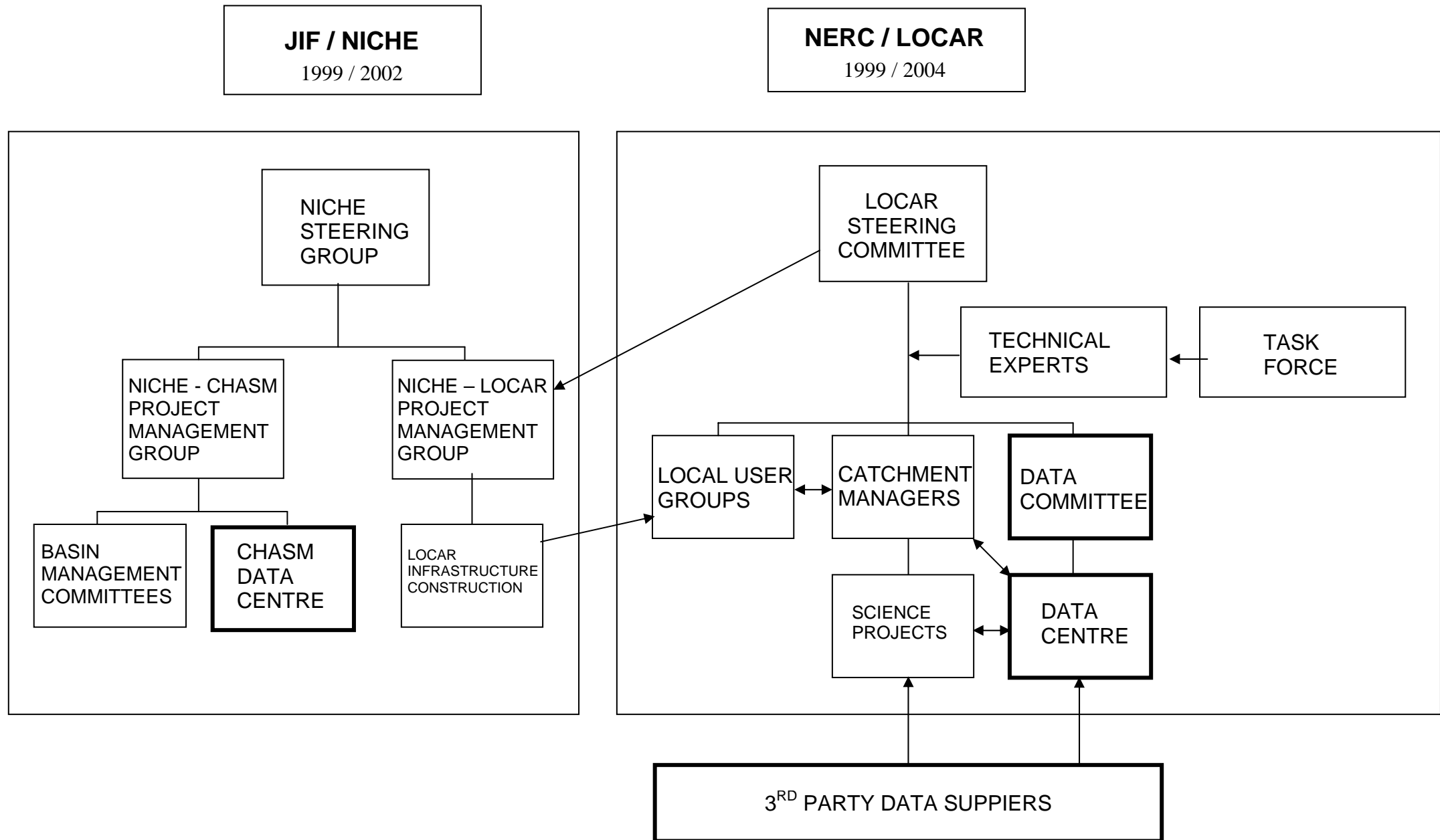
Data management activities within LOCAR will be co-ordinated by a Data Committee, which will take advice from and interact closely with the Steering Committee. They will also liaise with the Task Force, Catchment Managers and the relevant NERC Designated Data Centres. Figure 1 shows these interactions.

There will be one Data Centre for LOCAR data, staffed by IH, IFE and BGS staff. Initially the Data Centre's main task will be servicing users data needs. Later this will change to gathering in LOCAR generated data, securing it and finally organising its publication and exploitation. Within the available resources, the Data Centre will aim to be 'seamless' in terms of acquiring, supplying and archiving datasets.

The Data Centre will be responsible for implementing the NERC Data Policy. Expert and specialist advice from outside the Data Committee will be called upon when necessary.

The Data Committee will have a budget and the delegated authority to spend it.

Figure 1. Organogram for LOCAR Data Management



THE DATA COMMITTEE

Terms of Reference

The terms of reference of the Data Committee are:

- To advise the Steering Committee on data policy for LOCAR
- To identify the requirements for data management within LOCAR
- To co-ordinate the activities of the LOCAR Data Centre who will provide data management facilities for LOCAR
- To set standards
- To publish appropriate LOCAR data at the conclusion of the Programme
- To advise the Steering Committee on how data generated by LOCAR can be exploited
- To develop appropriate techniques and tools necessary for the management of LOCAR data
- To ensure the long term security of data generated and acquired for LOCAR
- To manage its budget
- To report to the Steering Committee

Membership

The membership of the Data Committee will comprise:

- A chairman
- The Data Centre manager
- Representatives of:
 - The LOCAR Science community
 - Key data providers
 - Centre/Survey Designated Data Centres
- A technical secretary

Meetings

The Data Committee will meet as necessary, probably three or four times a year.

Appropriate members of the committee will attend LOCAR meetings and workshops to maintain good co-ordination and communication, to receive feed back on issues relating to data and to ensure that PI's are informed on all matters relating to data.

DATA CENTRE

Terms of Reference

The terms of reference of the Data Centre are:

- To acquire major data sets from 3rd parties and the field programme
- To provide data to scientists
- To receive the data and model output of the science programme
- To secure the data at the end of the programme
- To set standards

- To publish appropriate LOCAR data at the conclusion of the Programme
- To develop appropriate techniques and tools necessary for the management of LOCAR data

Role

The Data Centre will provide the data management infrastructure for the programme.

Figure 2 shows diagrammatically the data flow from third parties, NERC and the field programme to the Data Centre and the scientists. Appendix I outlines the main procedures by which the data will move between these groups. A key function of the Data Centre will be the continuing development and implementation of these procedures over the life of the Programme.

Membership

The Data Centre will be staffed by:

- | | |
|--------------------|---------------------------|
| ▪ Roger Moore | CEH Wallingford (Manager) |
| ▪ Michael Gravelle | CEH Wareham |
| ▪ Andrew McKenzie | BGS Wallingford |
| ▪ Isabella Tindall | CEH Wallingford |

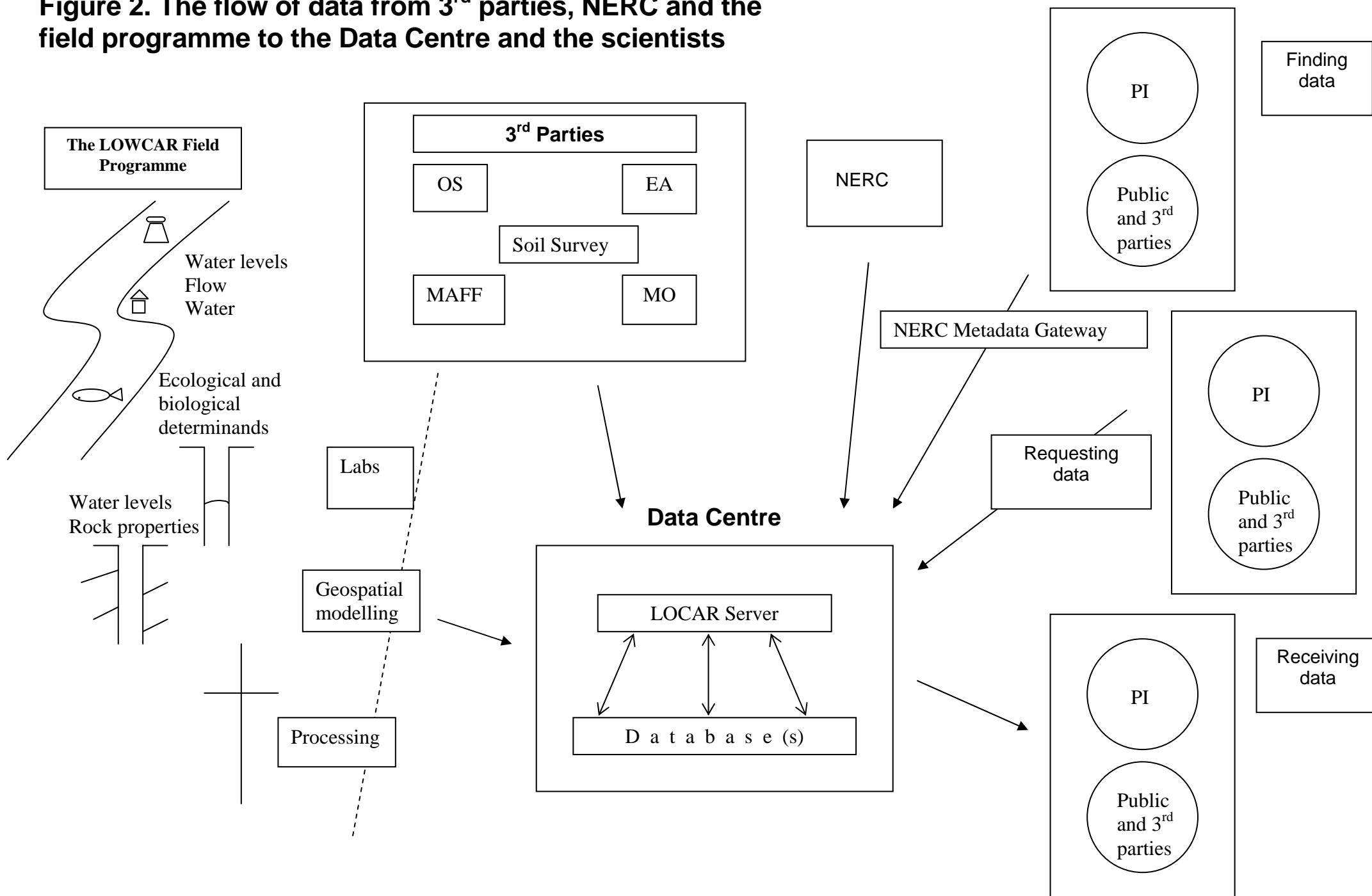
DATA POLICY

NERC policy on data, copyright and intellectual property rights are set out in handbooks and guidance notes available from NERC Headquarters.

The essential points of this policy in relation to LOCAR are:

- Data gathered with LOCAR funds belong to NERC
- All data gathered for LOCAR must be deposited with the nominated Data Centre as soon as practicable with agreed supporting documentation
- Data gathered for LOCAR are to be available to all researchers working within the LOCAR programme and the wider NERC community
- Applications to use LOCAR data by people and organisations external to NERC will be referred to the LOCAR chairman.
- Data users must inform the originators of data of their use
- Credit must be given to the originators of data (citation, co-authorship)
- Copyright and IPR must be observed
- After LOCAR, LOCAR data will be available on the same terms as other NERC data; in general this means free or for the cost of delivery to *bona fide* researchers and at what the market will bear for commercial users

Figure 2. The flow of data from 3rd parties, NERC and the field programme to the Data Centre and the scientists



IDENTIFYING REQUIREMENTS FOR DATA

The first task of the Data Centre is to study the Task Forces Report on data. Through this it should gain an overall understanding of the data that will be required for and output by the programme. This knowledge will be used to plan the assembly of existing data and the purchase of external data and also to plan for the reception of LOCAR generated data over the life of the project.

One of the aims of LOCAR is to understand how catchments work as an integrated whole. A consequence of this is that it is likely that the various sub-projects of LOCAR will need to pass data between each other. A useful overall impression of the data inputs, transfers and outputs can therefore be obtained through a meeting where a giant wall chart is prepared that shows for each project, process or model what its inputs and outputs are and their expected sources or destinations.

This exercise is valuable as it highlights:

- Areas where standards are important, e.g.:
 - Names
 - Definitions
 - Units
 - Methods
 - Meta-data
- Areas where timings are important
- Data sets required by many projects for which central funding and purchase are likely to be the most efficient means of acquisition

It is appreciated that this chart will not be definitive and will change over the life of the project as plans change in response to the problems encountered.

Follow up meetings with individual PI's will deal with detailed arrangements and requirements and will cover:

- Data requirements not covered by the above exercise
- Agreeing formats for data exchange
- Agreeing media for data exchange
- Building dictionaries of variables
- Agreeing a timetable
- Preliminary discussion of the arrangements for submitting data at the end of LOCAR

DATA ACQUISITION

Each LOCAR Principal Investigator (PI) will be assigned to a component of the Data Centre (IH, BGS or IFE) through which he will make all requests for and submissions for data.

The Data Centre will have resources for purchasing minor items of data and will be able to do so on its own authority. Major purchases will be authorised by the Data Committee who will liase directly and immediately if speed is of the essence.

In general, the major data suppliers prefer programmes such as LOCAR to provide a single point of contact through which all requests for data are routed. The Data Centre will act as the contact point for the LOCAR Programme.

The Data Committee and Data Centre will not be responsible for buying small data sets that are highly specific to a single project and are unlikely to be of wide interest. These are the responsibility of the project.

DATA MANAGEMENT

The day to day task of the Data Centre will be to provide the services and facilities required for negotiating the purchase and assembly of data, storing them in an orderly manner and making them available to users in a convenient form.

Given the volumes of data involved, it will not be reasonable or feasible for the Data Centre to quality control all the data they receive. However, where validation systems are in place, they will be used and errors detected will be reported back to the suppliers. Errors detected by users should be notified to the Data Centre and referred back to the supplier for correction.

The Data Centre will be responsible for the security of its data, for ensuring that users are aware of licence conditions attaching to any data they use and it will be responsible for maintaining suitable meta-data describing their holdings.

ACCESS TO DATA

Subject to the limits imposed by resources and commercial or confidentiality agreements, all NERC data are available to LOCAR researchers.

Access will be negotiated to all relevant data held by LOCAR collaborators such as the Environment Agency.

Non-NERC data purchased for or made available to LOCAR are likely to have conditions attached to their use. It is the responsibility of PI's to find out from the supplier of each data set what conditions apply to the data set and to ensure that the members of his/her project observe those conditions.

LONG TERM SECURITY

PI's will be responsible for ensuring that all LOCAR data in their care is held securely and is adequately backed up until it is passed to the Data Centre with agreed supporting documentation.

At the conclusion of LOCAR, PI's will deposit all data collected with the LOCAR Data Centre together with adequate documentation. If appropriate, the data will be absorbed into national collections, otherwise they will be catalogued and archived.

PUBLICATION

Over the early part of LOCAR the Data Committee will, on receiving a request from the Steering Committee, prepare a plan for the publication of the LOCAR data. The technologies available for publishing data are changing rapidly, however, the underlying tasks remain the same. The plan will cover:

- the objective of the publication
- the audience
- the budget and pricing policy
- the content and style
- the most appropriate medium
- the data and documentation
- the software
- the documentation
- artwork
- packaging
- publicity
- reviewing the material and testing the system
- delivery of the material to the publisher
- distribution

EXPLOITATION

The Data Committee will be responsible for advising the Steering Committee on how data generated by LOCAR can be exploited.

DATABASE DEVELOPMENT

The Data Committee will identify and commission the development of any techniques and tools necessary in order to streamline data collection or dissemination.

3. DELIVERABLES

The major deliverables are:

- The Data Management Plan (This document when completed by the Data Committee – it will, of course, be modified as the programme progresses)
- A data management service
- Standards (as required) covering:
 - Data definitions
 - Analytical methods
 - Quality control
 - Meta-data
 - Interchange formats
 - etc
- The LOCAR data deposited securely in the NERC Designated Data Centre
- The LOCAR data published on suitable media (e.g. CD-ROM, the WWW, etc.)
- If feasible, an exploitation plan, prepared and executed

4. PROGRAMME OF WORK

Appendix II sets out an initial programme of work for the LOCAR Data Committee and its Data Centre. This programme will need to be updated by the Data Committee as the LOCAR programme progresses.

5. RESOURCES

STAFF

In Section 6, staffing has been estimated for the year 2000. Outline figures have been provided for the rest of the programme in Appendix II.

CAPITAL ITEMS

Computing

It is anticipated that the Data Centre components will use their existing systems to service the needs of LOCAR. However, the principle has been established in earlier programmes that each programme contributes to the maintenance and replacement of the systems and equipment that it uses. Without yet knowing the scale of the LOCAR operation it is difficult to identify the equipment requirements of LOCAR. The following list is based on past programmes:

- Additional disk space for databases and file store 50-100Gbytes
- Upgrading of central processing power
- Replacement/upgrading of work stations
- Replacement/upgrading of PC's
- Software

Data

Funds will be required for the purchase of third party data sets. The provisional estimate of the costs of these data are indicated in Appendix II.

RECURRENT

Funds will be required for:

- Travel and subsistence
- Consumables (CD-ROM's, Disks, tapes, etc)
- Consultancy (e.g. hire of programmers, etc)
- Printing, publishing, packaging
- Postage
- Insurance
- Minor software

6. DATA CENTRE WORK PLAN AND COST SUMMARY 2000

It should be noted that the costs for years 2001 – 2005 are estimates, which will be reviewed based on proposals from the LOCAR Steering Committee and following consideration of the Task Force Report.

Item	2000 ¹	2001	2002	2003	2004	2005	
	Man days / cost						
Project planning and management	49						
Data management	120						
Data acquisition from 3 rd parties	51						
Receive data from the field programme							
Data provision to LOCAR projects	15						
Secure LOCAR data	0						
Publication	0						
Develop data management tools	90						
Steering Committee duties	3						
Data Committee	16						
Total staff time (days)	344						
Staff costs ²	£77,400						
T&S	£1000						
Computing							
Data	£20,000.00						
Total	£98,400	£111,076	£64,549	£64,249	£71,563	£11,658	£421,495

¹ The years in the plan are nominal and will be adjusted when formal permission to start is received from the Steering Committee

² Staff costs are based on Salary + 46% overhead + Superannuation.

APPENDIX I - OPERATIONAL PROCEDURES

This Appendix contains initial outline proposals for the operational procedures adopted by the LOCAR Data Centre. These will be developed and implemented over the life of the Programme.

FINDING OUT WHAT DATA EXIST

The primary method by which PI's will find out what data exists will be by accessing the NERC Metadata Gateway. The web address for this is:

<http://snowball.ag.rl.ac.uk>

REQUESTING DATA

PI's will request data in writing via email, fax or letter. Requests will be sent to their nominated contact. If and when the mechanism is set up, PI's may also be able to directly request and obtain data via the web.

Most data will be on open access but a few sets will have access restrictions attached to their use.

If a data set is requested that is not already in the possession of the Data Centre, provided the data are of general interest to LOCAR and subject to funds, then the Data Centre will acquire the data.

SUBMITTING AND RECEIVING DATA FROM THE PI'S

PI's will prepare a Data Management Plan which will state what, how and when data will be submitted. The Data Centre will produce guidelines for these plans.

LOCAR projects will be signed off and final payments authorised when the Data Centre has signed to say that all agreed data sets have been satisfactorily submitted.

RECEIVING DATA FROM THE FIELD PROGRAMME

The field managers will be responsible for supplying validated data to the Data Centre in an agreed format.

Further details for this procedure will be prepared when the Task Force Report on data and the appointment of Catchment managers are complete.

APPENDIX II - PROGRAMME OF WORK

It should be noted that the costs for years 2001 – 2005 are estimates, which will be reviewed following consideration of the Task Force Report.

Task	Duration (Man days)	Resource	Grade	Rate	Man days or Cost					
					2000	2001	2002	2003	2004	2005
Project planning and management										
Review draft plan	20				4	4	4	4	4	
Prepare work programme	6				2	1	1	1	1	
Prepare QA procedures	5				5					
Prepare metadata standards:	10				10					
Prepare data standards	15				15					
Workshop ³	8				8					
PI liaison ⁴	30					20	10			
Disseminate	15				5	5	5			
Data management										
Set up										
Plan and acquire hardware and software Set up hardware	20				10	10				
Set up direct user access	10				10					
Set up database										
Set up database	40				40					
Set up metadata catalogue (inc Att. Dicts)	15				15					
Set up Data Centre WWW pages	13				13					

³ Wall paper session to identify inputs and outputs of each project, the transfers between projects, the need for standards and a timetable of events.

⁴ Assume 25 PI's and 2 days work per PI + T&S.

Task	Duration (Man days)	Resource	Grade	Rate	Man days or Cost					
					2000	2001	2002	2003	2004	2005
Plan and document procedures for transferring data from the field programme to the Data Centre	10				10					
Quality control	10				2	2	2	2	2	
Populate meta data catalogues ⁵	12				5	3	2	2		
User support	75				15	15	15	15	15	
Data acquisition from 3rd parties ⁶										
Negotiate licence agreement and cost	14				10	1	1	1	1	
Obtain funds	1				1					
Receive data, reformat and load into database	75				40	20	5	5	5	
Receive data from the field programme ^{7 8}										
Receive data, reformat, quality control and load data	720					144	144	144	144	
Data provision to LOCAR projects ⁹										
Receive data requests, negotiate and obtain data and disseminate to users	50				15	15	10	5	5	
Secure LOCAR data ¹⁰										
Agree details with PI's ¹¹ , remind PI's, receive data, convert formats, quality control, load data and secure	150					10	20	30	50	40

⁵ Assume 50 data sets.

⁶ Assume 4 organisations, probably, EA, MO, OS and MAFF.

⁷ Assume 10 major field data sets requiring 4 days / month work by each of BGS, IH and IFE.

⁸ Assume the 4 Data Managers will visit the 3 catchments once per year and will incur 12 days T&S per year

⁹ Assume 25 data sets at 2 days work per data set.

¹⁰ Assume 6 days per PI.

¹¹ Assume 1 visit to each PI incurring 25 days T&S.

Task	Duration (Man days)	Resource	Grade	Rate	Man days or Cost					
					2000	2001	2002	2003	2004	2005
Publication ¹²										
Identify user requirement										
Design software										
Write software										
Test										
Assemble data										
Load data										
Check data										
Document data										
Peer review										
Prepare help system										
Final testing										
Publish										
Distribute										
Develop data management tools										
Develop meta-data system	20				20					
Develop tools to receive data from the field programme and 3 rd parties	20				20					
Improve direct user access	60				40	5	5	5	5	
Develop dictionaries	50				10	40				
Steering Committee duties										
Attend 2 meetings per year	12				2	2	2	2	2	2

¹² The content and medium of the publication are not known at this point and much will change in relation to publishing in the next 5 years. We assume the effort will be equivalent to producing a BODC data CD, i.e 1 man-year plus £10k pressing and printing costs. If the equivalent of a LOIS Overview CD were to be contemplated then the cost would be of the order 3 man-years plus £15k for artwork, pressing and printing. At present no costs have been assigned to this task.

Task	Duration (Man days)	Resource	Grade	Rate	Man days or Cost					
					2000	2001	2002	2003	2004	2005
Annual report	5				1	1	1	1	1	1
Data Committee ¹³	80				16	16	16	16	16	16
4 meetings per year										
Prepare, papers, etc										
Attend meeting										
Minutes										
Attend 4 workshops per year										
Total staff time	1427				344	314	243	233	251	42
Staff costs ¹⁴	£346,395				£77,400	£73,476	£59,049	£58,949	£66,013	£11,508
Computing ¹⁵⁻¹⁶						£15,000				
T & S					£1000	£600	£500	£300	£550	£150
Data					£20,000	£22,000	£5,000	£5,000	£5,000	
Annual totals	£421,495				£98,400	£111,076	£64,549	£64,249	£71,563	£11,658

¹³ Assume RVM, CIT, AMcK & MG.

¹⁴ To allow for employing programmers, possibly from an external company, and project supervision at Band 4, assume an average grade of SSO with 46% overheads and hence an average daily rates of the order £225, £234, £243, £253, £263 and £274 in each of the six years of the Programme with inflation at 4%.

¹⁵ Computing requirements will be evaluated in the year 2000. For budget planning purposes an indicative figure of £15K is included in 2001.

¹⁶ Provision should be made for additional Oracle/SQL licences if multiple simultaneous access to the database by more than 10 users is anticipated. (CEH Wallingford currently has 33 Oracle/SQL licences and a typical maximum simultaneous usage of between 16 and 26. The cost of a licence is unknown at present but is likely to be of the order £500 - £1000)