

Medical Research Council

Peter Dukes

Strategy & Liaison Manager
Research Management Group

MRC *Scope of MRC's Programme*

Molecules

⇒ **Data !**

- structures
- mechanisms
- associations

**Cellular &
Physiological
Systems**

⇒ **Data Sets**

- descriptions
- analyses
- models
- linkages

**People &
Populations**

Kinds of Data Set

Genomes / Bioinformatics

- Sequencing: high volume
- Annotating function
- Interoperability

Biomolecular Structures

- Real time experiments
- Modelling

Functional Imaging

- Complex information
- Integration across data

Clinical data bases

- 10 to 10^4 people
- Preserve for 20 years
- Meta-analysis - trials
- Personal information

Genetic data bases

- Family data
- Confidentiality
- Linkage to other data

Population data bases

- 10^3 to 10^4 people
- Multiple & repeat measures
- MRC/Wellcome Large Cohort - 5×10^5 people

Project scope

- Population data
- Archiving
- Access

Stage 1

Scoping survey

- Kinds of research design
- Size of studies
- Capture & storage media
- Data formats
- Extent of archiving

Policy questionnaire

- Public data sets register?
- Criteria for prioritising archiving?
- Expertise and capacity to archive
- Consent
- Model of archiving
 - central
 - distributed

MRC *MRC Population-based Data Sets*

Characteristics of MRC Data sets

Variables

- Complex
- Diverse
- Quantitative & qualitative
- Contentious
- Context dependent

Capture / Storage Media

- Paper, video, electronic, cards...

Formats

- Questionnaires, interview forms, diaries, instrument readings, images

Access Issues

Ownership

- Intellectual capital
- Value for the nation's money

Quality

- Independent scrutiny of access requests?
- Controls on 2e research outputs?

Consent & Data Protection

- What did people consent to?
- Anonymisation

Collaboration

- Agreements

What can the GRID offer?

New opportunities for science

Metadata Issues

Extracting meaning from data

Defining

- target population
- interventions
- measures
- instructions to interviewers

Code books, software

Level of confidentiality

Skills and Capacity

Central focus

- IT and computing resource
- Data archiving skills

Shared central and distributed

- Access management

Stage 1 Conclusions

New research proposals

- should *include* archiving

Archiving older data sets

- “All should be preserved”
 - if technically sound
 - if ethically sound
- Needs skills, resources!

Models

- Social sciences
- Bioinformatics
- Biological imaging

Real concerns

- Loss of PI control
- Added bureaucracy

Stage 2 Plans

Several case studies

Examine

- current practices, skills, capacity, cost

Issues

- Primary data preservation
- IT & computing needs
- Metadata “readiness”
- “Ownership”
 - PIs, Funders, Nations
- Technologies and culture to support data access & exchange
- **Scientific opportunity!**