



MAKING THE MOST OF OUR DATA

MRC's Data Sharing & Preservation Initiative

Tues 5 December 2006

Peter Dukes
Medical Research Council

Contact: Allan Sudlow
Allan.sudlow@headoffice.mrc.ac.uk

www.mrc.ac.uk/PolicyGuidance/EthicsAndGovernance/DataSharing

MRC Vision



- Proactive sharing and preservation are **normal behaviour**
- Publicly funded research **datasets are fully used** for high quality and ethical research...

Caveats - 2003



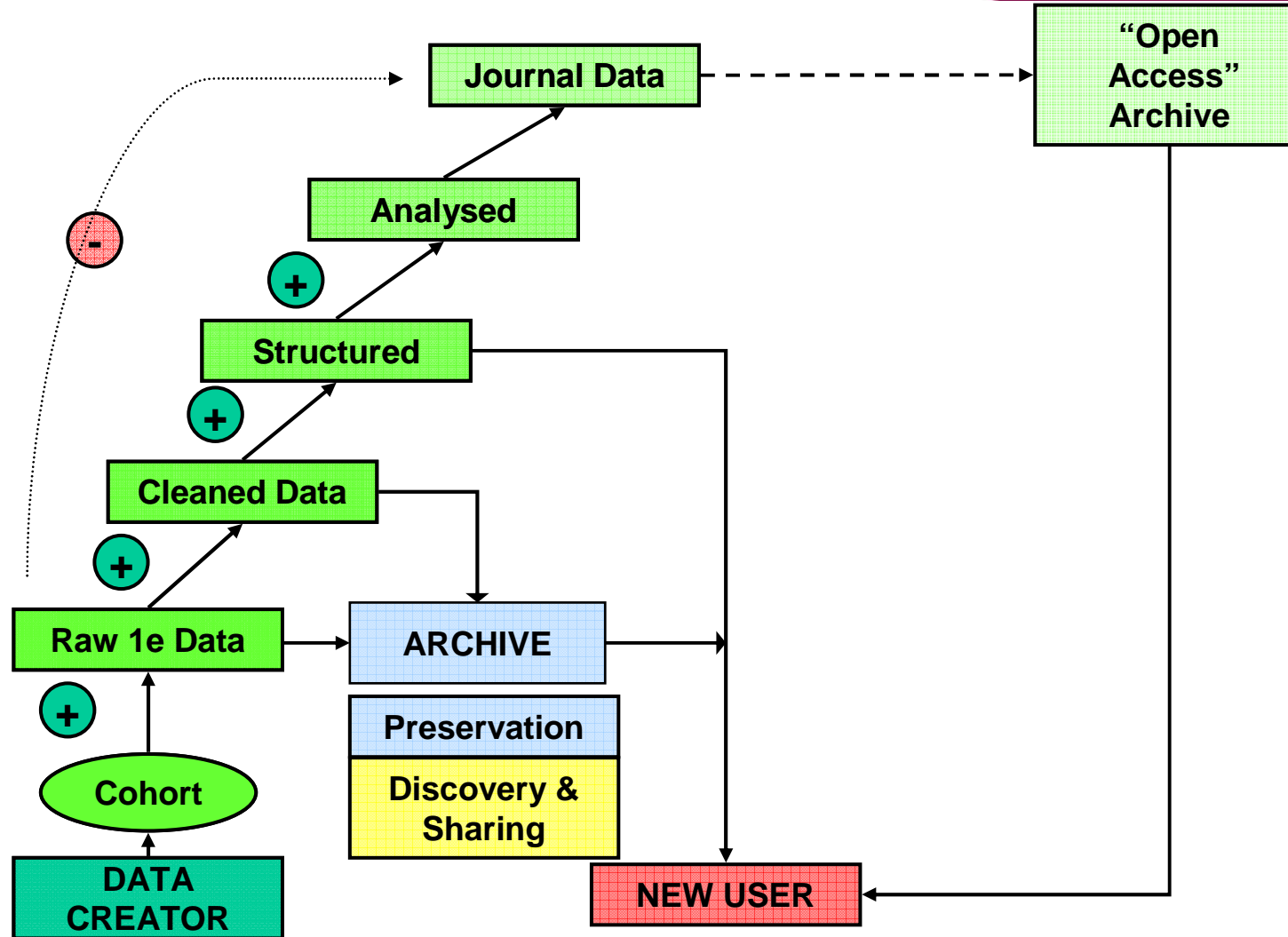
- *I don't know all the answers*
- *No-one knows all the answers*
- *Not all questions have an answer...*
- *And some have more than one*

Focus on Population Data: E.g. Children of the 1950s Study



- **Survey of all schoolchildren in Aberdeen in 1962**
 - about 14,000
- **Linked to**
 - birth records
 - interviews with subsample of mothers
 - mortality records of study participants and parents
 - birth records of female participants
 - NHS hospital discharge records
 - follow up questionnaire in adulthood
 - data on schools and neighbourhoods
- **Data held in**
 - punch cards
 - paper schedules
 - magnetic tapes
 - microfilms
- **Data being collected/analysed by people in**
 - Aberdeen, Glasgow, LSHTM, Bath, Stockholm, Bristol

Life Cycle Data Management



What do researchers do now... (Survey data, 2001)



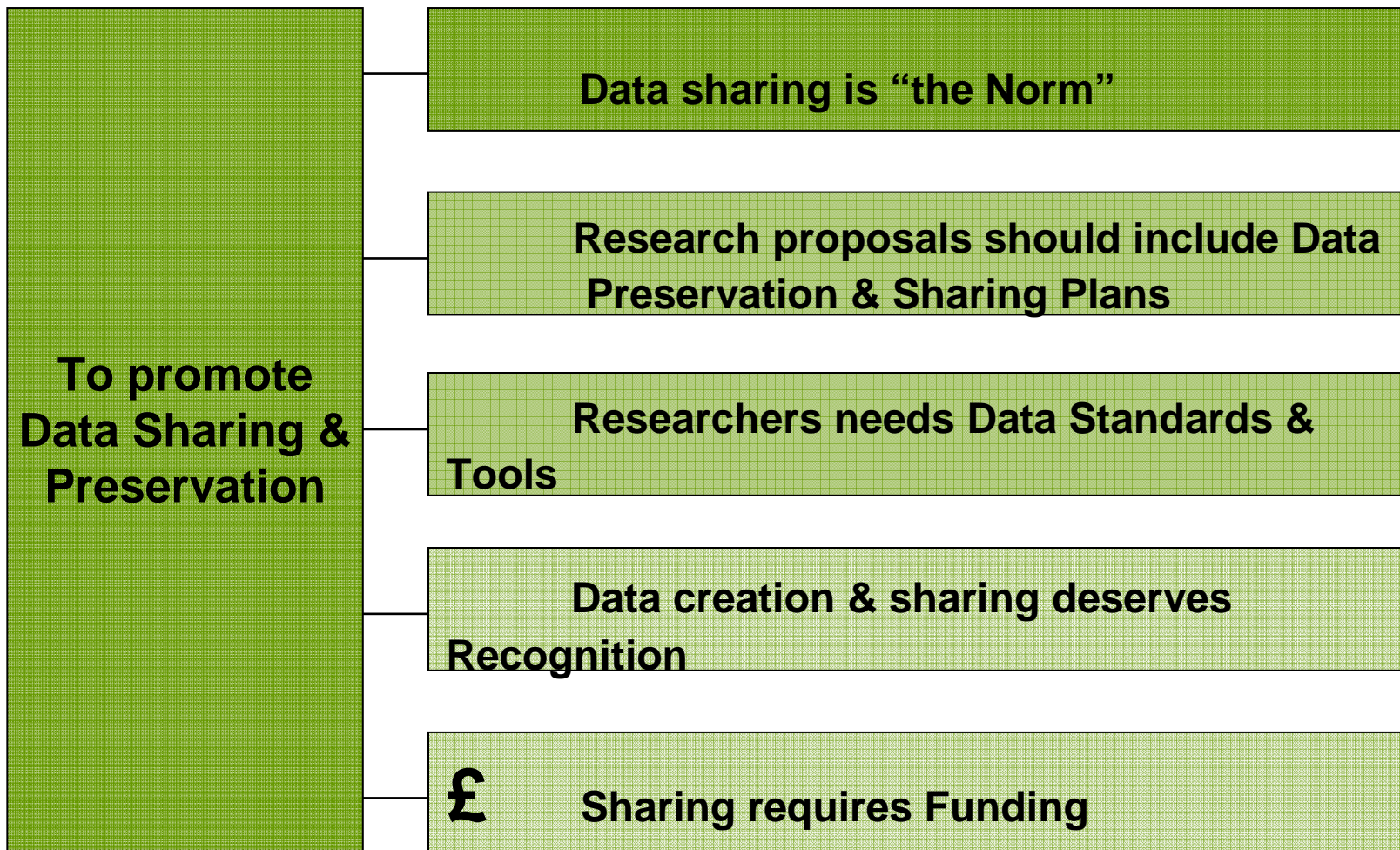
- **Current practice**
 - **Little attention to preservation for sharing**
 - 90% of MRC survey respondents preserve (or plan to) 1e data
 - 12% of respondents archive data to enable use by others
 - 67% of respondents agree archiving = beneficial
 - Good practice in data management for primary research purposes, but not for preservation
 - **Scale of the paper problem**
 - “12 x 4 drawer filing cabinets, 2 cabinets, 8 boxes...”
 - “47,000 questionnaires – 75 metres of shelving...”
 - **Resources**
 - “We cannot guarantee data management beyond the funding period of the trial.”
 - “No additional funding for expanding archives”
 - “Whose responsibility is it?”

Where is MRC today?



- **We have a corporate commitment**
 - Policy
 - Funding
- **We are building an implementation “tool kit”**
 - Basic digital preservation guidance
 - Guidance on preparing & evaluating DSP plans
 - A route map through consent & confidentiality
 - Principles & models of governance – access to data
- **Work to be done**
 - Increasing MRC data asset *visibility*
 - Developing & implementing *access* solutions
 - Implementing available *preservation* solutions
 - Joining solutions into a *system*
 - *Costing & prioritisation* for data enhancement
- **Next steps**
 - **1st: define the *business model***
 - through a *Transformational Partnership*
 - **2nd: Create a *Service***
 - responsive, joined-up, sustainable, value adding

5 Policy Principles



Data sharing is “the Norm”



“The idea is OK, but for many investigators restricted access to their data set is their intellectual capital”

“Not everyone can be a scavenger feeding of the work of other.”

- **Focusing initiative on rich shareable studies**
 - **Currently**
 - National Survey of Health & Development
 - Avon Longitudinal Study of Parents & Children
 - **Next step**
 - 4-6 substantial population based studies
 - **Identify**
 - Benefits & beneficiaries
 - Cost drivers, issues of scale
 - Lifecycle data management models

Research proposals should include Data Preservation & Sharing Plans



- **Data sharing & preservation plans**
 - **All funding proposals must contain a short DSP plan**
 - **Guidance & exemplars are available**
 - Applicants
 - Peer reviewers
 - **Early days**

Researchers needs Data Standards & Tools



Three web-based guides to curation

Guide 1

- Development & Assessment of researchers' DSP plans

Guide 2

- Data & Metadata Management for Population-based Medical Research

Guide 3

- Strategies & Solutions for Data Preservation for Sharing

Search:

Welcome to the Data and Tissues Tool Kit

On this site you will find practical help with implementation of the Data Protection Act (1998), Section 60 of the Health & Social Care Act (2001), the common law of confidentiality, the Human Tissue Act (2004) and Human Tissue Authority codes of practice, and all associated guidance.

The site has been developed primarily for health researchers and research managers working with personal data and human tissue samples, in the academic sector, but will also be of use to other health professionals.

To help you to navigate the regulatory environment, much of the information is organised in Route Maps. Your feedback on the content and presentation will help us improve the site.

The site has been developed by the UK Medical Research Council for use in all academic research studies. Further Info...



Access Toolkit Bibliography

Find links to relevant documents on the web, plus specially commissioned pieces to help you in the approvals process.

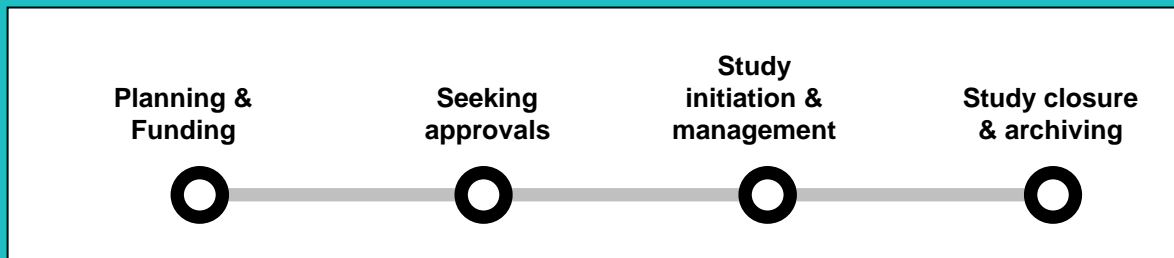


Provide feedback on Toolkit

This is the first draft of the Toolkit. We need feedback on which areas are useful and which are not – and where you may need more help.

First time on this site? [Visit the Getting Started page](#)

Ready to use a Route Map?



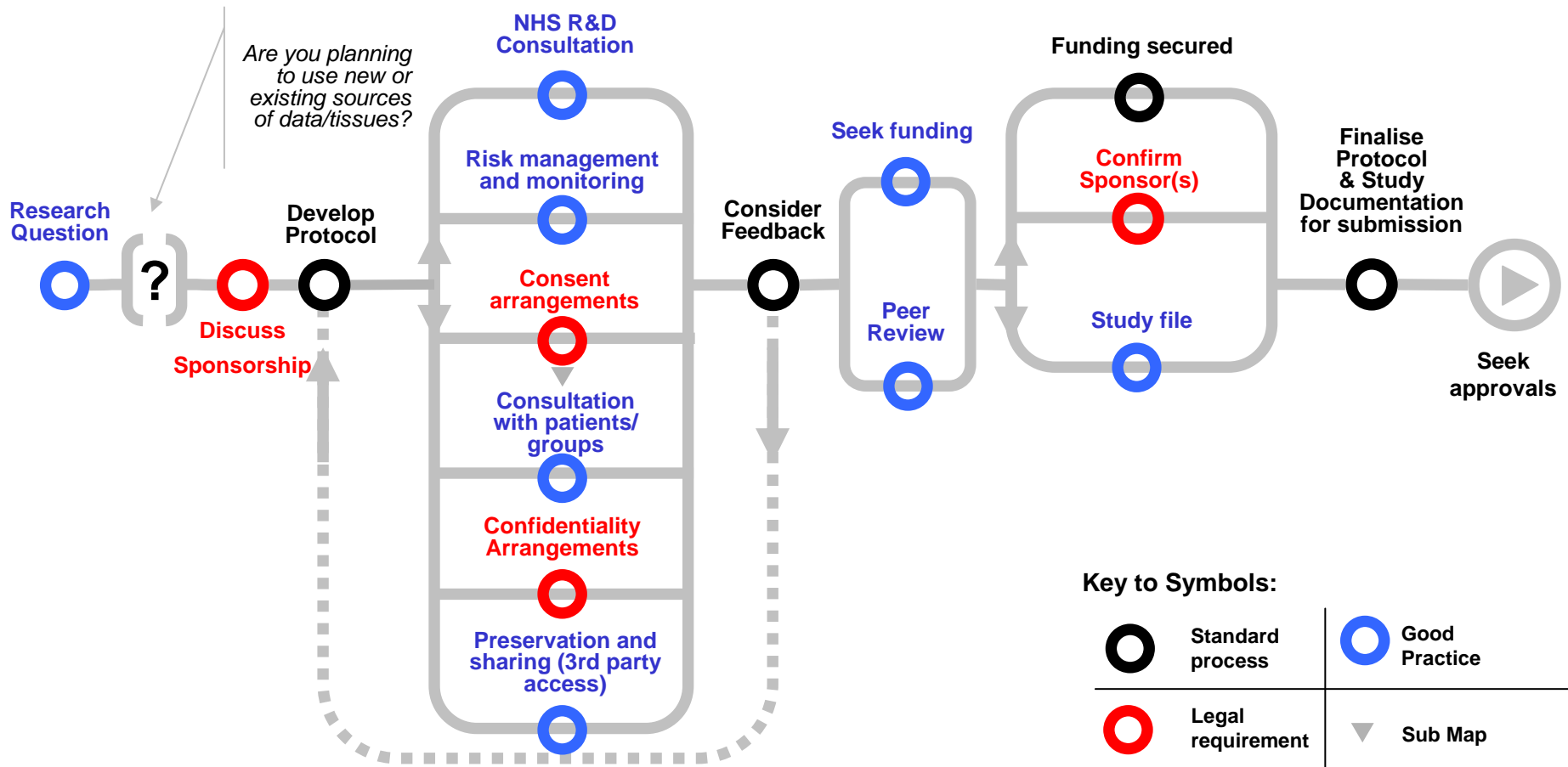
Data and Tissues Tool kit

[Home](#)
[Routemaps](#)
[About this site](#)
[What's New](#)
[FAQs](#)
[Glossary](#)
[Useful Info](#)
[Medical Research](#)

Planning & Funding
Search:

[Map View](#)
[List View](#)
[Previous Station](#)
[Up level](#)
[Next Station](#)

You are here : [Planning & Funding](#)



Researchers needs Data Standards & Tools



- **Data publishing**
 - **Nesstar implementation**
 - National Survey for Health & Development
 - Avon Longitudinal Study of Parents & Children
- **Metadata standards**
 - **Key issue to be addressed**
- **Governance of access**
 - **Transparent & equitable access**
 - **Fair reward and recognition**
 - **Independent oversight**

Researchers needs Data Standards & Tools



Cohort Support Project: Help NSHD & ALSPAC to...

1. Assess current practice

2. Determine Requirements to

**Maintain value through time/custodianship changes
Enable discovery & wider access for informed re-use**

3. Recommend options for enhanced preservation for sharing

Bronze —————> Silver —————> Gold

4. Assess

**Benefits derived – procedural/technical & scientific
Resource (cost & time) implications for implementation**

5. Inform cohort data management plans

Data creation & sharing deserves Recognition



- **Institutional recognition**
 - Research assessment exercises
 - MRC reviews of its Units & Institutes
- **PI recognition**
 - Database citation
 - Peer review – e.g. of database use
- **Ownership & responsibility**
 - **Who “owns” data?**
 - *Our trial involves 33 countries and 4 funding agencies...*
 - Data creator
 - Funder
 - Employer
 - **“Who owns the secondary data?”**

£ Sharing & preservation require Funding

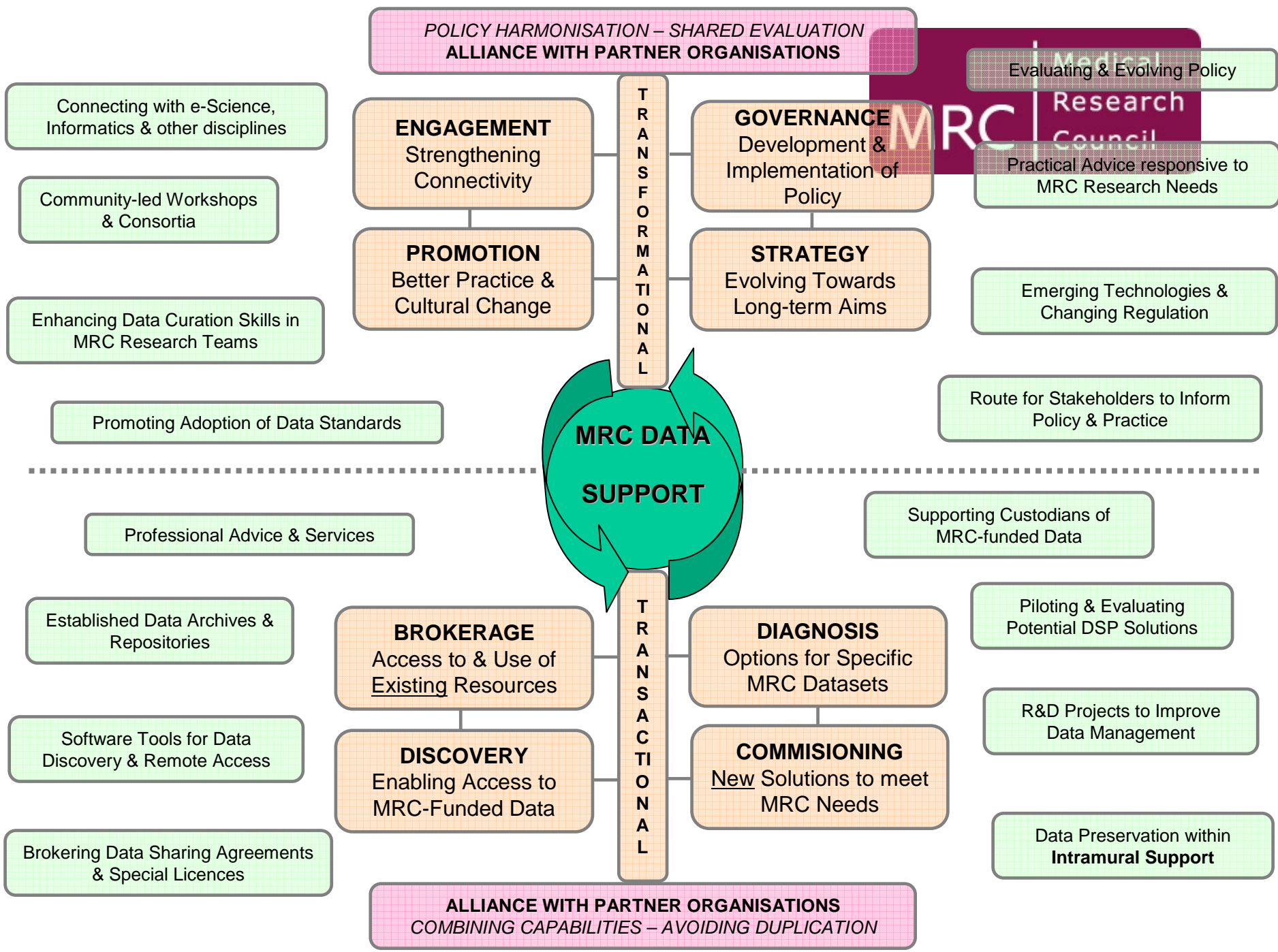
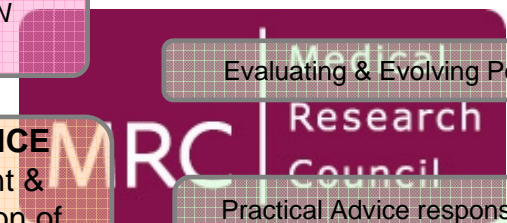


- **Who should pay for DSP?**
 - *“As a small group we cannot guarantee data management beyond the period of the trial...”*
- **MRC funds the DSP costs on grants**
 - Costs are usually implicit
 - Explicit costs have yet to be challenged
- **Pilot study - NSHD & ALSPAC**
 - But difficult to identify DSP-specific costs
- **Retrospective DSP costs >> Prospective DSP**
 - What could MRC afford?

The DSP Vision



- Funders promote **collaboration** and **avoid duplication**
- Proactive sharing and preservation are **normal behaviour**
- Researchers can access & implement **tools and standards**
- **Incentives & rewards** are in place and barriers removed
- **People trust researchers** using the their data and value the widest responsible use of those data
- Users and creators **trust the quality** of each others' work
- Patients, data creators, managers, and secondary researchers are all **recognised**
- Publicly funded research **datasets are fully used** for high quality and ethical research



**POLICY HARMONISATION – SHARED EVALUATION
ALLIANCE WITH PARTNER ORGANISATIONS**

ENGAGEMENT
Strengthening
Connectivity

PROMOTION
Better Practice &
Cultural Change

GOVERNANCE
Development &
Implementation of
Policy

STRATEGY
Evolving Towards
Long-term Aims

**MRC DATA
SUPPORT**

T
R
A
N
S
F
O
R
M
A
T
I
O
N
A
L

T
R
A
N
S
A
C
T
I
O
N
A
L

**ALLIANCE WITH PARTNER ORGANISATIONS
COMBINING CAPABILITIES – AVOIDING DUPLICATION**

BROKERAGE
Access to & Use of
Existing Resources

DISCOVERY
Enabling Access to
MRC-Funded Data

DIAGNOSIS
Options for Specific
MRC Datasets

COMMISSIONING
New Solutions to meet
MRC Needs

Connecting with e-Science,
Informatics & other disciplines

Community-led Workshops
& Consortia

Enhancing Data Curation Skills in
MRC Research Teams

Promoting Adoption of Data Standards

Professional Advice & Services

Established Data Archives &
Repositories

Software Tools for Data
Discovery & Remote Access

Brokering Data Sharing Agreements
& Special Licences

Route for Stakeholders to Inform
Policy & Practice

Supporting Custodians of
MRC-funded Data

Piloting & Evaluating
Potential DSP Solutions

R&D Projects to Improve
Data Management

Data Preservation within
Intramural Support

Evaluating & Evolving Policy

Practical Advice responsive to
MRC Research Needs

Emerging Technologies &
Changing Regulation



Thank you