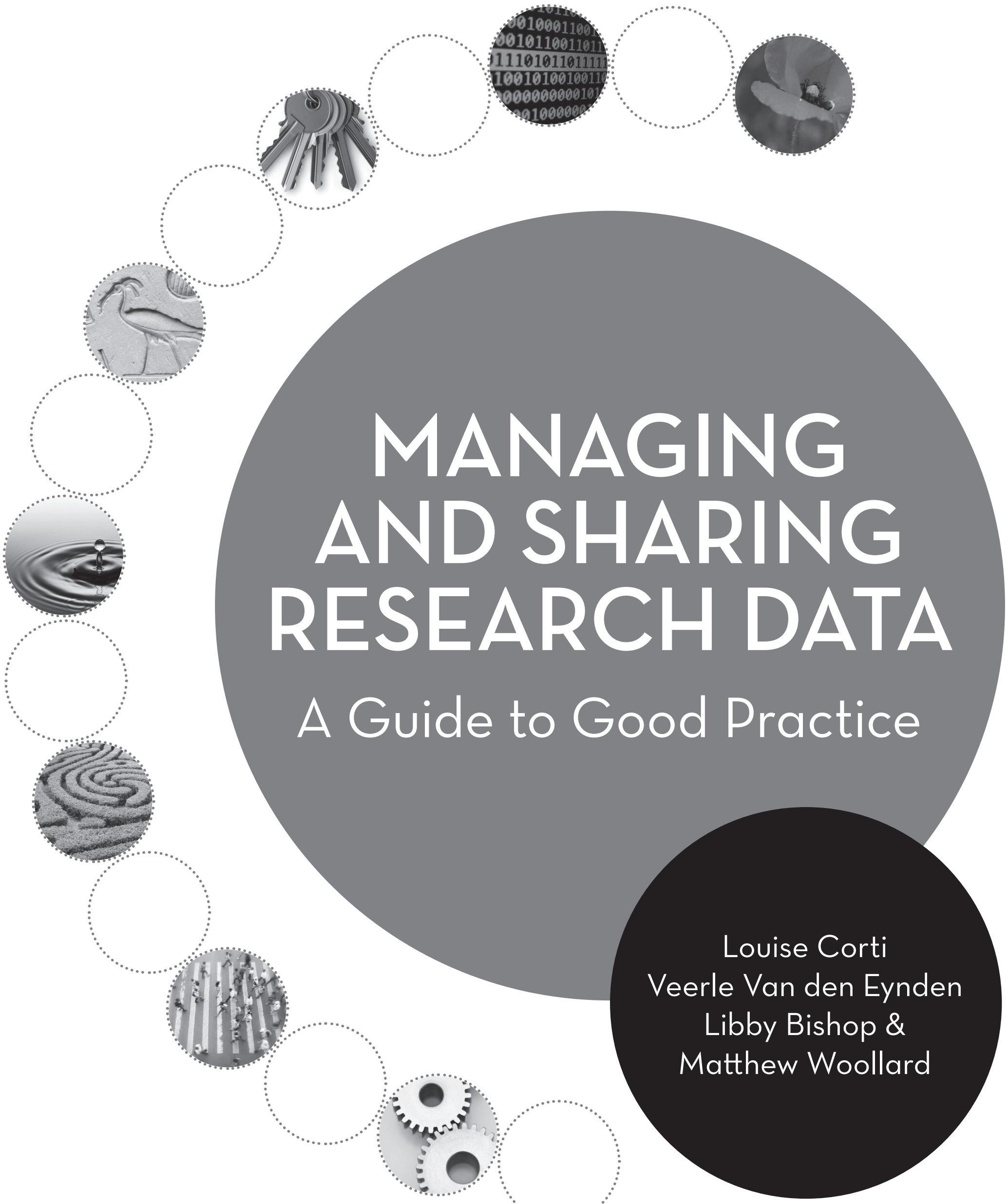

MANAGING AND SHARING RESEARCH DATA

A Guide to Good Practice

Louise Corti
Veerle Van den Eynden
Libby Bishop &
Matthew Woollard

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

TWO
The Research Data Lifecycle

Most data often have a much longer lifespan than the research project that creates them. Researchers may continue to work on data after funding has ceased, follow-up projects may analyse or add to the data, or data may be reused and repurposed by other researchers. If data are well managed during the course of a research project, and if they are properly preserved, curated and made accessible for the longer term, they will be able to be reused in future research.

During the 1990s and early 2000s the data lifecycle was promoted as a concept to support digital preservation and data curation practices. The notion of a data lifecycle is one that has gained popularity as the culture of data sharing becomes part of our everyday research language. The data lifecycle extends the typical research cycle. Table 2.1 sets out an overview of data-related activities typically undertaken in the research data lifecycle.

Table 2.1 Typical activities undertaken in the research data lifecycle

Activity	Key features
Discovery and planning	Designing research
	Planning data management
	Planning consent for sharing
	Planning data collection, processing protocols and templates
	Finding and discovering existing data sources
Data collection	Collecting data – recording, observation, measurement, experimentation and simulation
	Capturing and creating metadata
	Acquiring existing third party data
Data processing and analysis	Entering data, digitizing, transcribing and translating
	Checking, validating, cleaning and anonymizing data where necessary
	Deriving data
	Describing and documenting data
	Analysing data

(Continued)

Table 2.1 (Continued)

Activity	Key features
	Interpreting data
	Producing research outputs
	Authoring publications
	Citing data sources
	Managing and storing data
Publishing and sharing	Establishing copyright of data
	Creating discovery metadata and user documentation
	Publishing or sharing data
	Distributing data
	Controlling access to data
	Promoting data
Long-term management	Migrating data to best format
	Migrating data to suitable medium
	Backing up and storing data
	Gathering and producing metadata and documentation
	Preserving and curating data
Reusing data	Conducting secondary analysis
	Undertaking follow-up research
	Conducting research reviews
	Scrutinizing findings
	Using data for teaching and learning

The Data Documentation Initiative (**DDI**) is a popular documentation standard for social science data, and was one of the first initiatives to conceive the idea of a data lifecycle (DDI Alliance, 2013). The idea integrates research processes and activities with concepts of data curation, data preservation, data publishing and data sharing. Figure 2.1 shows this initial DDI lifecycle, which uses similar but not identical terminology to those concepts in Table 2.1. We will learn more about the DDI and its importance for describing social science data in Chapter 4 on documenting data.

Also in the domain of the social sciences, Humphrey set out a lifecycle model of research as it may be applied to the concept of 'e-science' (Humphrey, 2006). Shown in Figure 2.2, the chevrons in this diagram are discrete stages in research, each representing processes that generate information. The knowledge transfer (KT) stage in Humphrey's model deals with the wide variety of communications that flow from research, such as research outcomes, publications and so on.

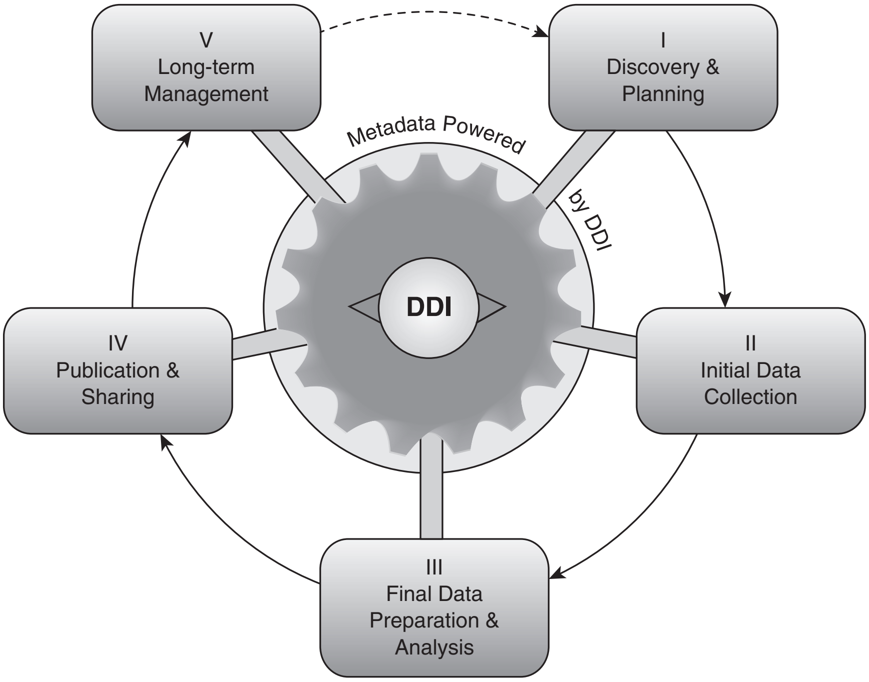

Figure 2.1 Data Documentation Initiative (DDI) lifecycle

Source: DDI, 2013

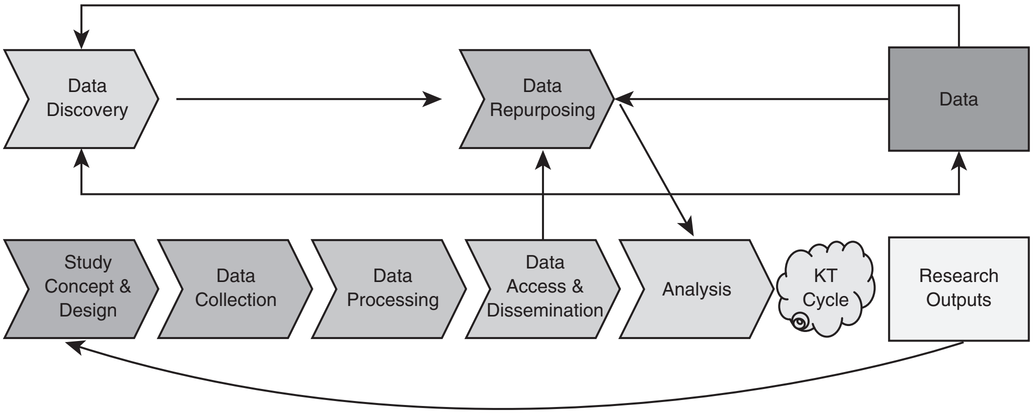

Figure 2.2 The lifecycle model of research knowledge creation

Source: Humphrey, 2006

The important point to note here are the gaps between chevrons: the transitions that occur in research as products are finished and pass to the next phase. Humphrey indicates that these transition points in the lifecycle are the places in a project's cycle that are most vulnerable to information loss, for example detailed sampling procedures for the design of a survey. These transition points are the most significant areas in negotiating the data management plan for a project, and in particular, clarifying who is responsible for the digital objects created at each stage.

Taking Humphrey's concern, critical data management intervention points can be grafted onto the research cycle before the project even begins. For example, formally signing off consent forms so that data sharing is not precluded can be a requirement prior to going into the field to collect interview data. Or spreadsheet templates and pre-set rules for data entry can be set up prior to data being captured.

As a researcher, it is very helpful to consider what aspects of data management might apply to each stage of your research before the project even starts. It is equally helpful to visualize the research cycle and the data lifecycle of a particular research project and its resulting research data to help identify the points at which actions or interventions should be taken.

We will cover the practicalities of such lifecycle planning in Chapter 3, with Exercise 3.1 illustrating how a sketched data lifecycle can be used to identify intervention points for data management procedures.

Since then other organizations have adapted the DDI lifecycle model to create their own version. Examples include the generic Data Curation Centre (DCC) Lifecycle Model (DCC, 2012) and the USA Geological Survey Data Lifecycle Diagram (USGS, 2013). This latter diagram (see Figure 2.3) clearly indicates the linear steps of planning, acquiring, preserving and publishing/sharing data, together with the ongoing processes throughout the lifecycle of describing data, managing their quality and backing up and securing them.

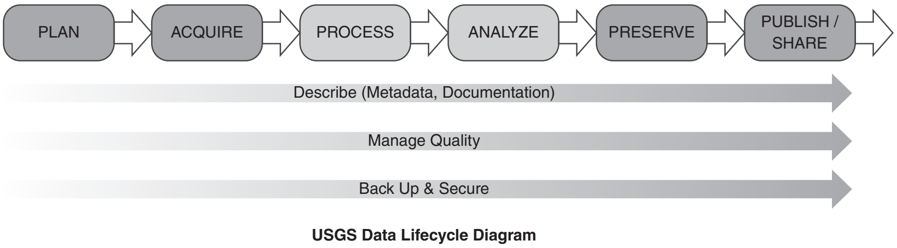

Figure 2.3 The USA Geological Survey data lifecycle model

Source: USGS, 2013

The **OAIS** reference model is a more formalized conceptual framework describing the environment, functional components, and information objects within a system responsible for the long-term preservation of digital materials. It provides a lifecycle model for data archives and is widely recognized in scientific, data management and archival communities (CCSDS, 2012).

In the case study below we give an example of a real-life research project, which shared its resulting data successfully, and its key data management intervention points.

CASE STUDY

Data Lifecycle of the 'Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria' Project (Mort, 2006)

The 2001 foot and mouth disease outbreak had an enormous effect on the economic, social and political life of rural areas in the UK. This research project, which was funded by the Department of Health, produced evidence about the human health and social consequences of the epidemic.

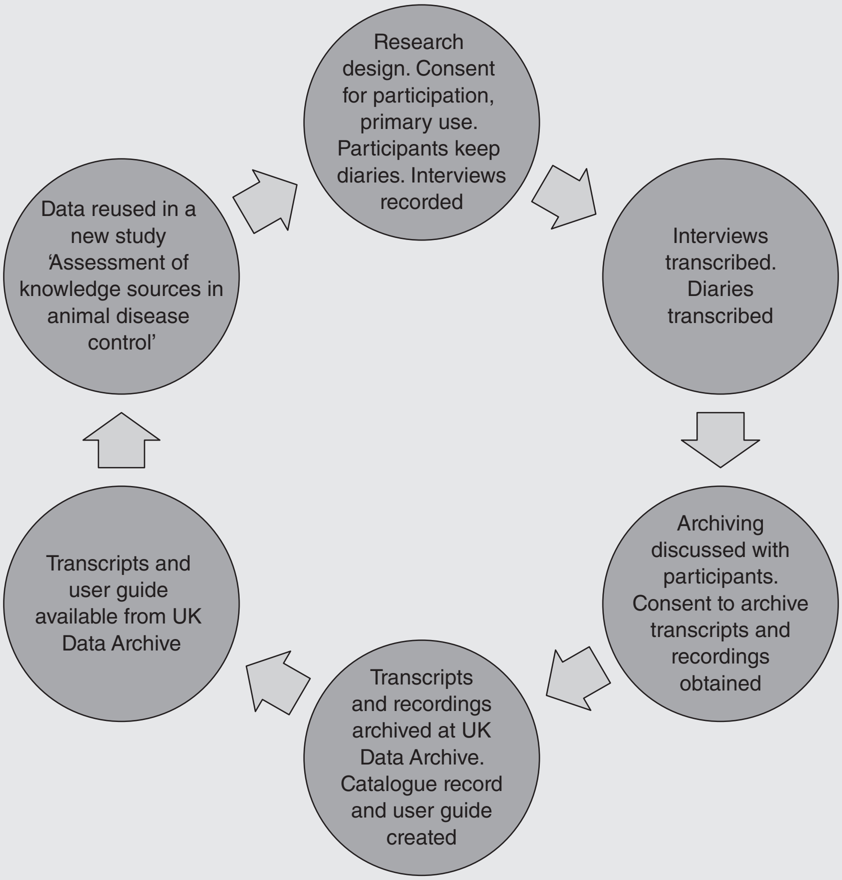
Research design. Consent for participation, primary use. Participants keep diaries. Interviews recorded
Interviews transcribed. Diaries transcribed
Archiving discussed with participants. Consent to archive transcripts and recordings obtained
Transcripts and recordings archived at UK Data Archive. Catalogue record and user guide created
Transcripts and user guide available from UK Data Archive
Data reused in a new study 'Assessment of knowledge sources in animal disease control'

Figure 2.4 Data lifecycle from the project 'Health and social consequences of the foot and mouth disease epidemic in North Cumbria, 2001–2003'

Source: Mort, 2006

(Continued)

(Continued)

The study recruited a standing panel of 54 local people from the worst affected area (North Cumbria). This panel wrote weekly diaries over a period of 18 months describing how their lives had been affected by the crisis and the process of recovery they observed around them. The panel was recruited to reflect a broad range of occupations including farmers and their families, workers in related agricultural occupations, those in small businesses including tourism, hotel trades and rural business, health professionals, veterinary practitioners, voluntary organizations and residents living near disposal sites.

The panel members produced 3,200 weekly diary entries of great intensity and diversity over an 18-month period. The data were supplemented by in-depth interviews with each respondent, focus group discussions, and 16 other interviews with stakeholders.

The research team gained consent from participants for primary participation in the project, but did not get consent for sharing or archiving their data. When the project was finished they wanted to archive their data for future reuse, so had to gain retrospective consent. As we will see later on, it is far more efficient to gain consent for data sharing early on in the fieldwork process. They consulted with the panel of participants about retrospective consent procedures and also sought expert advice from copyright law specialists to help draft terms of agreement that would give respondents a series of options about how their diaries, copies or portions of diaries, and/or their audio material would be archived. The research team gave respondents the choice to decide how their material was used. These data are now available to other researchers from the UK Data Service. Its data lifecycle is shown in Figure 2.4.

The significance of taking an approach which takes into account the whole of both the research lifecycle and the data lifecycle is that it allows all participants to understand their roles and responsibilities in context. Many of the pressure points in ensuring that data are available for reuse fall, temporally, immediately after significant tasks and changes in responsibility. Being clear that the end of an activity denotes a change in responsibility and not the end of a piece of work is vital for the long-term access to data. We discuss planning, roles and responsibilities next in Chapter 3.

References

CCSDS (2012) *Reference Model for an Open Archival Information System* (OAIS), The Consultative Committee for Space Data System Principles. Available at: http://public.ccsds.org/publications/archive/650x0m2.pdf.

DCC (2012) *The DCC Curation Lifecycle Model*, Digital Curation Centre. Available at: http://www.dcc.ac.uk/resources/curation-lifecycle-model.

DDI Alliance (2013) *Data Documentation Initiative*. Available at: http://www.ddialliance.org.

Humphrey, C. (2006) *e-Science and the Life Cycle of Research*, University of Alberta. Available at: http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc.

Mort, M. (2006*) Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001–2003* [computer file]. Colchester, Essex: UK Data Archive [distributor], November 2006. SN: 5407. Available at: http://dx.doi.org/10.5255/UKDA-SN-5407-1.

USGS (2013) *Data lifecycle Diagram*, USA Geological Survey. Available at: http://www.usgs.gov/datamanagement/why-dm/lifecycleoverview.php.