4 DDI COMBINED LIFE CYCLE MODEL
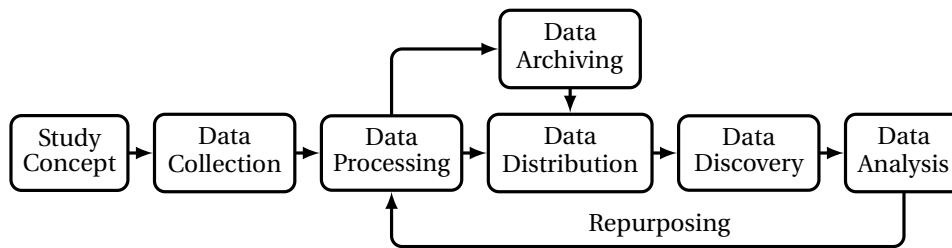

Figure 3: DDI version 3.0 Combined Life Cycle Model.

The Data Documentation Initiative (DDI) version 3.0 Conceptual Model [Str04] contains a Combined Life Cycle Model for research data, particularly social science data. The name signifies the fact that the model was formed by combining the initial draft of the model, constructed from a data application perspective, with elements from another model [GK02]. The model is linear for the most part, with one alternative path and one feedback loop; a graphical representation can be found as Figure 3. It has the following eight elements.

Study Concept. The earliest stage in the model is the point at which a survey is being designed. The model includes in this not only the choice of research question to answer and the methodology for collecting the requisite data, but also plans for the way in which the data will be processed, analysed and used to answer the question, and the form that answer will take. Researchers should also define at this stage the relationships that will exist between the data products of the research.

Data Collection. Examples given of collection methods and sources include surveys, censuses, voting or health records, commerce statistics or Web-based collections. Primary and secondary data sources should be clearly distinguished.

Data Processing. Once the input data has been assembled, it is processed and analysed to produce output data (e.g. a statistic or set thereof) that answers the research question. These output data may be recorded in a machine-readable form or human-orientated form such as a technical report.

Data Archiving. In order to ensure long-term access to the data, they should be passed to an archive rather than merely kept by researchers. The archive not only preserves the data (and metadata) but also adds value to them over time.

Data Distribution. The data are distributed to users either directly or via a library or data archive.

Data Discovery. The data may be publicised through books, journal publications, Web pages or other online services.

Data Analysis. The data may be used by others within the bounds of the original conceptualization; for example, picking out key statistics for a research report.

Repurposing. The data may also be used within a different conceptual framework; examples include sampling or restructuring the data, combining the data with other similar sets, or producing pedagogic materials.