e-Science and the Life Cycle of Research

Life cycle models are shaping the way we study digital information processes. These models represent the life course of a larger system, such as the research process, through a series of sequentially related stages or phases in which information is produced or manipulated. In the United Kingdom, for example, the Life Cycle Information for E-Literature (LIFE) and the Digital Curation Centre both have been influenced by this approach.[1] Similarly, data archivists are advancing life cycle models to improve practices in preserving research data.[2] I have applied a life cycle framework with two projects employing standardised metadata to bridge various stages of the research life cycle and two other projects exploring the application of digital repositories in a life cycle context.

This discussion presents ways in which the life cycle approach provides insight into the relationships among the stages and activities in research. An understanding of this perspective may contribute further insight into the function of e-science in the larger picture of scientific research.

One representation of a life cycle model of research is depicted in Figure 1 below.[3] The chevrons in this figure signify discrete stages in research. Each stage consists of processes, many of which generate information. The gaps between chevrons are the transitions that occur in research as products are finished and passed to the next phase. These transition points in the life cycle are the places most vulnerable to information loss. While a product may pass successfully to the next stage, the volume of background information documenting the product can be abandoned in the previous stage. For example, an output from the data collection stage may consist of the completed questionnaires from a survey that are then passed to the data processing stage. The information left behind could include the sources of the questions, the method for selecting the survey's sample, the details about administering the questionnaire and other information about gathering the data. Some of this information may be compiled in a user's guide to accompany the data but often this information is in a format that cannot be easily integrated for use with information in other stages. *The life cycle approach makes us more aware of possible information losses in the gaps between stages.*

The knowledge transfer (KT) stage in Figure 1 deals with the wide variety of communications that flow from research (see Figure 2 for an elaboration of this

---

[1] See the homepages for the LIFE project at http://www.ucl.ac.uk/ls/lifeproject/ and the Digital Curation Centre at http://www.dcc.ac.uk/index

[2] See the presentations made at the 2005 IASSIST Conference in two sessions: C1: The Life Course of Survey Data: Evidence from New Tools and D2: Enriching Metadata: the Lifecycle Perspective. Powerpoint files of the presentations in these sessions are available at http://www.iassistdata.org/conferences/2005/presentations/

[3] Adapted from a model developed by the DDI Structural Reform Working Group. http://www.icpsr.umich.edu/DDI/committee-info/working-groups.html

stage). Only a fraction of the information underlying these communications is available at this stage. Imagine having easy access to the information and data behind each research finding at the KT stage. This would greatly enrich the knowledge transfer experience of future researchers as well as casual readers of research findings. The big-picture view helps determine the value of this additional information in each stage and identifies the relationships among materials to facilitate access throughout the life cycle. This would entail selecting information and preparing metadata in anticipation of the possible linkages of this information across stages. *The life cycle framework helps us visualize or imagine relationships across the stages of research that might otherwise be overlooked.*

The transition points in the life cycle are also crucial junctions for other important activities in research. Many streams of activity flow throughout the life cycle of research, including project administration, grant procurement, data management, knowledge creation, ethical judgments, intellectual property supervision and technology management. Bridging activities across stages requires coordination and a sense of continuity in the overall process. For example, the hand-off of research products between stages often involves a shift of responsibilities from one person to another or from one organization to another. These shifts in responsibility are key to the concept of digital custodianship, which I characterise as the process through which formal agreements are negotiated among partners about the custodial responsibility for digital objects throughout the full life cycle. Similarly, the custodial care of such objects should address issues of intellectual ownership. The Creative Commons[4] is a recent development that provides new ways of declaring the ownership of digital objects that might well be applied in the context of research. *The life-cycle framework can encompass other significant developments in information and knowledge management through the streams of activity occurring across the stages of research.*

In the same way that it takes the proverbial village to raise a child, it takes the commitment of the knowledge community to preserve and provide access to research data. The stakeholders in the knowledge community include granting agencies, universities, researchers, data producers, publishers, research libraries and data archives. James Jacobs and I have argued that research libraries have a significant role to play along with data archives and scientific publishers in providing "a safety net for capturing and retaining the products of scientific research."[5] Collaboration among these stakeholders will be essential in the development of e-science. *Through the life-cycle framework, the roles that partners assume across the various stages of research can be more easily defined.*

---

[4] See http://creativecommons.org/ for more information about the variety of flexible copyright licences available for digital objects.

[5] James Jacobs and Charles Humphrey, Preserving Research Data, **Communications of the ACM**, Vol. 47 (9), pp. 27-29.

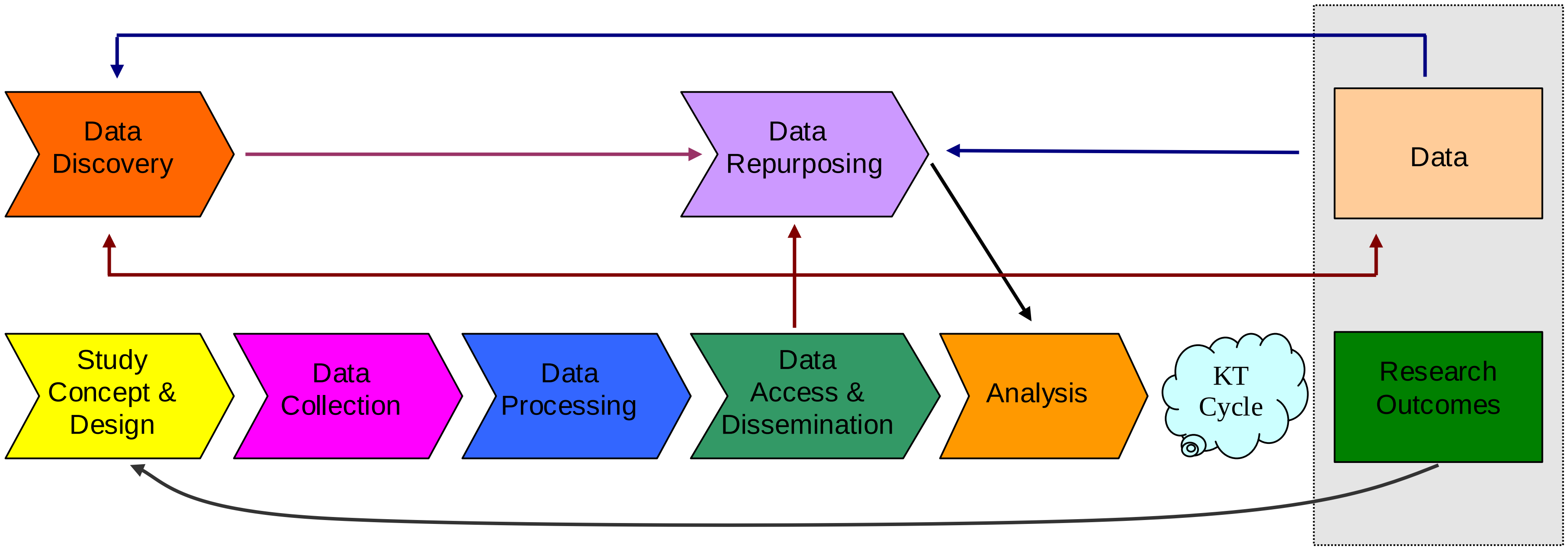
How does e-science fit into the life cycle model of research? I propose that a large part of e-science is located along the top row of Figure 1: the stages of data discovery, data repurposing and data. The activities of e-science involve processes within these stages as well as connections among them. For example, the arrow from Data to Data Repurposing could well represent data and text mining, which are recognised e-science activities. The "mashup" of content on the Web,[6] a new activity associated with e-science, involves relations among Data, Data Discovery and Data Repurposing. In addition, e-science has a fundamental tie to the relationship in Figure 1 represented by the box around Data and Research Outcomes. Both of these entities are important products of research. While research libraries traditionally have not dealt with data and research publications in an equivalent way, the goals of e-science require data to be managed on equal footing with research communications.

The life cycle framework is also a useful tool for planning services in the library. As the manager of research data services for the University of Alberta Library, I am committed to providing access to data. Without the preservation of research data, however, long-term access is jeopardized. The provision of services in both preserving and accessing data is paramount. How we go about planning and implementing these services can be greatly facilitated by building on a solid foundation. *The life cycle of research is a model that helps identify the landscape of possible data services and the partnerships to build in support of these services.*

To be responsive to e-science, research library staff must evaluate the services we provide for both research outcomes and data. Given the stages of the life cycle associated with e-science, we need to determine the services to be provided by research libraries and the partnerships required to implement and sustain these services.

[6] Declan Butler, Mashups mix data into global service, **Nature**, 439, p. 6-7.

FIGURE 1: THE LIFE CYCLE MODEL OF RESEARCH KNOWLEDGE CREATION

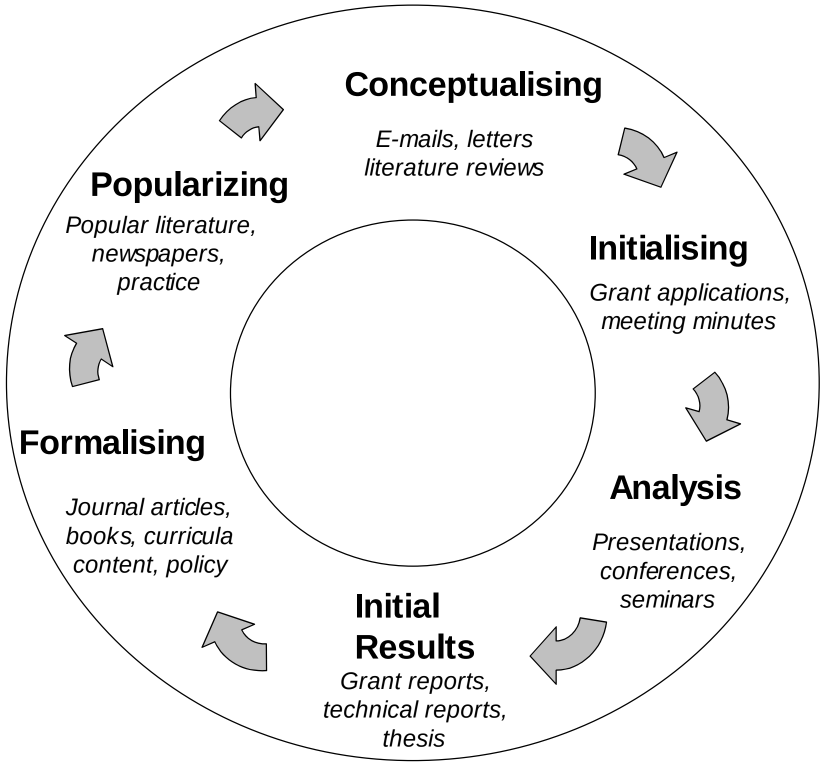


The KT (Knowledge Transfer) Cycle in this model consists of the variety of methods used in communicating research outcomes. This process is presented in a separate model shown on the next page.

The research outcomes and data upon which these outcomes are based collectively document the knowledge for an area of study. This is represented above in the lightly shaded box around Data and Research Outcomes.

Each of the chevrons in the above model represents a stage in the life cycle of knowledge creation from research. The gaps between chevrons symbolize the transitions between stages. These transitions tend to be vulnerable points in the documentation of a project's life cycle. When a stage is completed, its information may not get systematically preserved and instead end up dead-ended (most often on someone's hard drive.) Shifts in the responsibility for the objects of research also tend to occur at these points of transition. For example, the data collection stage passes along completed interviews or questionnaires to the data processing stage; the data processing stage passes one or more clean data files to the data access and dissemination stage. In each transition, someone else usually becomes responsible for the outcomes of the previous stage. These transition points become important areas in negotiating the digital curation plan for a project as partners in the life cycle of research identify who is responsible for the digital objects created at each stage.

FIGURE 2: RESEARCH COMMUNICATIONS IN THE KNOWLEDGE TRANSFER CYCLE



Source: Charles Humphrey and Elizabeth Hamilton (2004). "Is it Working? Assessing the Value of the Canadian Data Liberation Initiative." **Bottom Line**, Vol. 17 (4), pp. 137-146.