Exploring Data and Information Lifecycles

Prepared for MOIMS-DAI working group by

Terry Longstreth

Document Purpose

This document is intended to stimulate discussion and inform the decisions of the working group related to the eventual production of a process document to be used in satisfying funders' and producers' common interests in data and information management as it relates to eventual transfer, of the managed data, to an archive.

Information and Data Lifecycles

As a general rule, lifecycle models are used to illustrate progression from state to state or from use to use. In contrast to biological lifecycles, where state transitions as well as the existence of the lifecycles themselves are usually discovered empirically, the stages in lifecycles used to illustrate progression in engineering or business environments are human inventions, prepared by the enterprise, and used to illustrate the organization's perceptions of process or data flow. Data (or information flow) is the basis for an information lifecycle, but for the purposes of this document, we have adopted a very high level generic view of the lifecycle of information from its initial recognition within the organization (e.g. engineering , research or procurement project; business enterprise, government entity, research institution, intellectual property purveyor).

The information content of an enterprise's or project's data can only be fully and accurately managed by understanding information use in all lifecycle stages pertinent to a particular class or instance of information. Lifecycle management ensures the information content of an organization's data can be fully exploited by tracking the data's pedigree through the life cycle.

The information lifecycle adopted here to illustrate the stages needed for digital data management planning is shown in Figure 1. Data is put to differing uses at different times or for differing purposes, and the information lifecycle is intended to illustrate that.

> The lifecycle is composed of 5 activities pertinent to most enterprise environments (including business, operational or mission and probably other, environments). The information content of an enterprise or project's data can only be fully and accurately exploited by tracking the data's pedigree through the life cycle.
>
> The figure also shows a sixth activity, representing handover of the enterprise produced information either because of a statutory requirement or perceived declining value of the information to the enterprise via a formal turnover procedure (for example, PAIS) to a separately organized or governed long-term repository .

The Lifecycle Stages are not necessarily separated in time or by function. While we expect information to have a conceptual grounding before it is recorded and shared, in scientific domains particularly, the conceptual basis for an item of information is often one of the hoped-for products of a scientific dialectic. Concepts don't necessarily need labels, but when they become associated with things, the concepts require labels for them to be described, discussed, documented or discarded. For example, the concept Pluto:(the ninth planet) arose in the 1930s and was discarded (or relabeled) 60 years later.

Imagining a data or information lifecycle

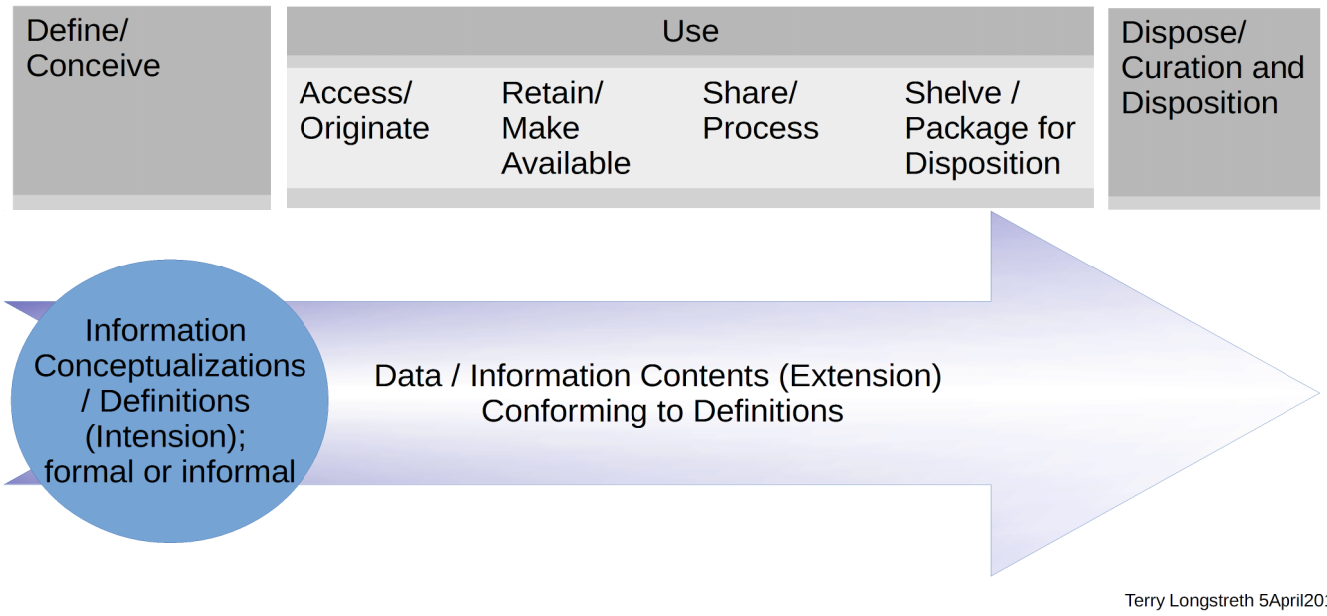

Figure 1: Life Cycle Stages for data or information, labeled by usage

Lifecycle Stages or Phases

The three major subdivisions of the Information Lifecycle are summarized here. For a different perspective on these phases, see Table 1, which follows.

Define

The birth of recorded information occurs when its existence is postulated or conceived in the (Define/ Conceive) stage and the relevant concepts are recognized. In formal settings, such as an engineering environment, this stage is characterized by the creation of descriptive definitions (metadata) and initialization data (constraints and initial dynamic data content). In informal settings, such as production of a document or recording minutes or notes, the Conceive stage may be less crisply defined and definitional information (such as file names in directories, and format or media descriptors supplied by a word processor or other software) may occur and be captured simultaneously with the primary data production of the Use stage.

Use

In the main (Use) stage, enterprise data is directly visible to operators and enterprise software. The Use stage has been subdivided into four activities, to show that use has different meanings depending upon user perspective or reasons for interaction. I emphasize

that these stages do not represent any particular time or precedence ordering. Only the information itself governs the uses for which it may be relevant and the particular circumstances under which it may be appropriate for handling under any of the four ***Activities.***

Disposition

The Disposition stage reflects the process by which live, operational and dynamically manipulated data and attendant ancillary information transitions out of immediate or operational relevance. While it's possible for actual destruction to occur at this point, in most circumstances actual destruction (erasure, removal, expungement) of data would occur as an aspect of the Use phase, but records (logs, journals, audit trails) of the destruction *actions* could conceivably be transferred beyond the operational environment and retained in an archive.

Lifecycle Functions

Table 1. Data Lifecycle phases of Figure 1, illustrates these three major phases in the evolution of data within an enterprise.

Enterprise Data Lifecycle management ensures the information content of enterprise data can be fully exploited by tracking the data's pedigree through the life cycle. The life cycle is composed of 6 activities. The birth of data occurs in the development (Define) phase, with the creation of descriptions (metadata) and initialization data (constraints and initial dynamic data content.). In the main (Use) phase, enterprise data is directly visible to operators and enterprise software. The Use phase has been subdivided into four topics, to show that use has different meanings depending upon user perspective or reasons for interaction.

Enterprise data has intrinsic value beyond that of its initial or enterprise projected utility and that value can be exploited when the data can be saved in a permanent, identifiable, recoverable form. While the table acknowledges the potential for information destruction, the Disposition phase represents this preparation for the potentially and essentially immortal quality of enterprise data.

Verb and Object or modifiers	Subjects, Agent or Agents
Define:	→ Design/Development Staff
•Abstract and concrete programmable structures	
•Associated initial value sets	
Use: The next four activities represent different but usually complementary functions. They occur simultaneously and asynchronously, and comprise components of dynamic Interactions, and Intentional or Intended Information Use	
Access:	→ Computer Programs and Individuals
• Programmed or ad hoc	
• Single, common list of standard access protocols across user community	
Retain:	→ DB, Communication and File Infrastructures

Verb and Object or modifiers	Subjects, Agent or Agents
• Dynamic state	
• Interrelated storage and message services	
Share:	→ Computer Programs and Individuals
• Concurrent or serial review and manipulation	
• Guarantee information consistency among sharing applications	
Shelve:	→ Storage Systems
• Final content representation fixed	
• Permanent invariant label applied	
• Enterprise content declared read-only	
• Ancillary and base (payload, operational) information packaged for Disposition	
Disposition:	→ Archive or Destruction Facility
• Disposition protocols defined and agreed	
• Information disposition protocols invoked and acknowledged and recorded	

Table 1. Data Lifecycle phases of Figure 1