2.3.4.5 Data Lifecycles

Return to Data Taxonomy

- Data Management Platform (DMP)

The **Data Lifecycle** covers the stages a particular piece of data transitions through from its initial generation or capture to its eventual archival and/or deletion at the end of its useful life. The Data Management (DM) is the coordination and administratoin of all data associated with a project, program or effort. DataManagement includes the meta data as well as the inidivaul pieces of data as it progresses through the Data Lifecycle. Data management follows a Data Strategy which also includes the required business rules, especially those captured in any governing Legal Documents such as the Charter, By-Laws, and policy and Procedures. Figure 1 reflects the major stages in the Data Lifecycle:

1. **Create** - The data is created, captured, copied from other sources
2. **Store** - The data is stored into a Datastore (i.e., Database, Document, Files, etc)
3. **Use** - The Data is actually used, accesed and referenced by an on-going business process
4. **Propagate** - The Data is copied to other data structures or to other nodes within the system. For example, servers, clients, tiers, and alternative formats (i.e., RDBMS to XML or JSON)
5. **Share** - The data is made public and can be shared with other internal or external business processes using any number of mechanisms such as

   HTTP(S), File Transfer Protocol (FTP), SMTP, SMS, IPFS, DDS,]]dido:public:ra:xapend:xapend.a_glossary:b:blockchain]], Distributed Ledger Technology (DLT), etc.

6. **Archive** - The data is preserved as part of the system legacy. In the past, the amount of data archived was limited primarily due to the cost of storage, but with the advent of inexpensive offline storage, most data is now preserved.
7. **Destroy** - The data is considered as having no value, become a laibility or is required to be destroyed by law (i.e., The Right to Be Forgotten). In traditional data systems, the data is usually overwritten with newer, more germain data. For example, old room temperatures are replaces with new ones if there is no requirement to archive the old temperature.

   Note: Figure 1 also includes a **Plan** Stage, however, since the data does not exist during planning, it is not considered as an actual Data Sage, but this in no way means it is not an important stage for data. It is during this stage the requirements for the data are identfied using system architecture and engineering and modeling. Planning usually includes Use-Cases, Prototypes, and the various Data Models (i.e., Conceptual, Logical and Physical), legal conciderations and the pragmatics of things like the quanity and quality of data. Sometimes during this stage, data that is “planned” never comes to fruition.

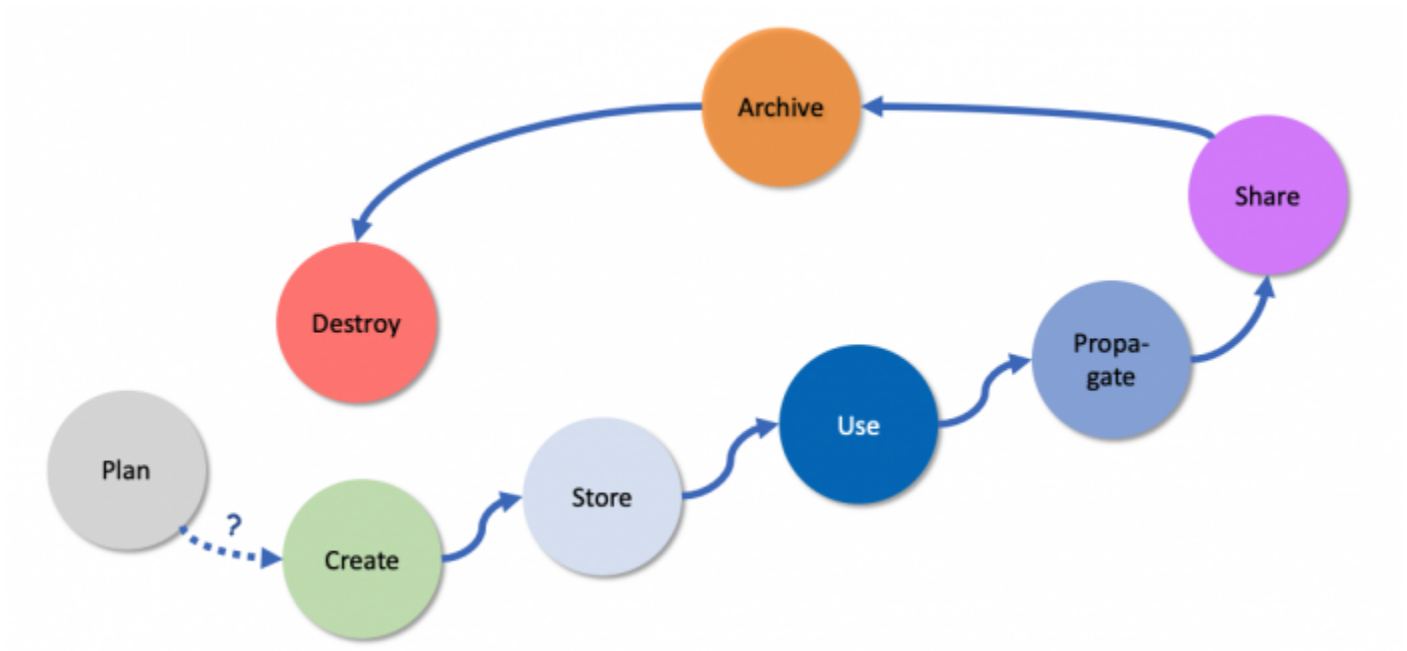

Figure 1: Traditional Data Lifecycle Stages

Although specifics vary, data management experts often identify six or more stages in the data life cycle. Here's one example:

Generation or capture: In this phase, data comes into an organization, usually through data entry, acquisition from an external source or signal reception, such as transmitted sensor data. Maintenance: In this phase, data is processed prior to its use. The data may be subjected to processes such as integration, scrubbing and extract-transform-load (ETL). Active use: In this phase, data is used to support the organization’s objectives and operations. Publication: In this phase, data isn’t necessarily made available to the broader public but is just sent outside the organization. Publication may or may not be part of the life cycle for a particular unit of data. Archiving: In this phase, data is removed from all active production environments. It is no longer processed, used or published but is stored in case it is needed again in the future. Purging: In this phase, every copy of data is deleted. Typically, this is performed on data that is already archived. Data lifecycle management (DLM) is becoming increasingly important since the explosion of big data and the ongoing development of the Internet of Things (IoT). Enormous volumes of data are being generated by an ever-increasing number of devices all over the world. Proper oversight of data throughout its life cycle is essential to optimize its usefulness and minimize the potential for errors. Finally, archiving or deleting data at the end of its useful life ensures that it does not consume more resources than necessary.

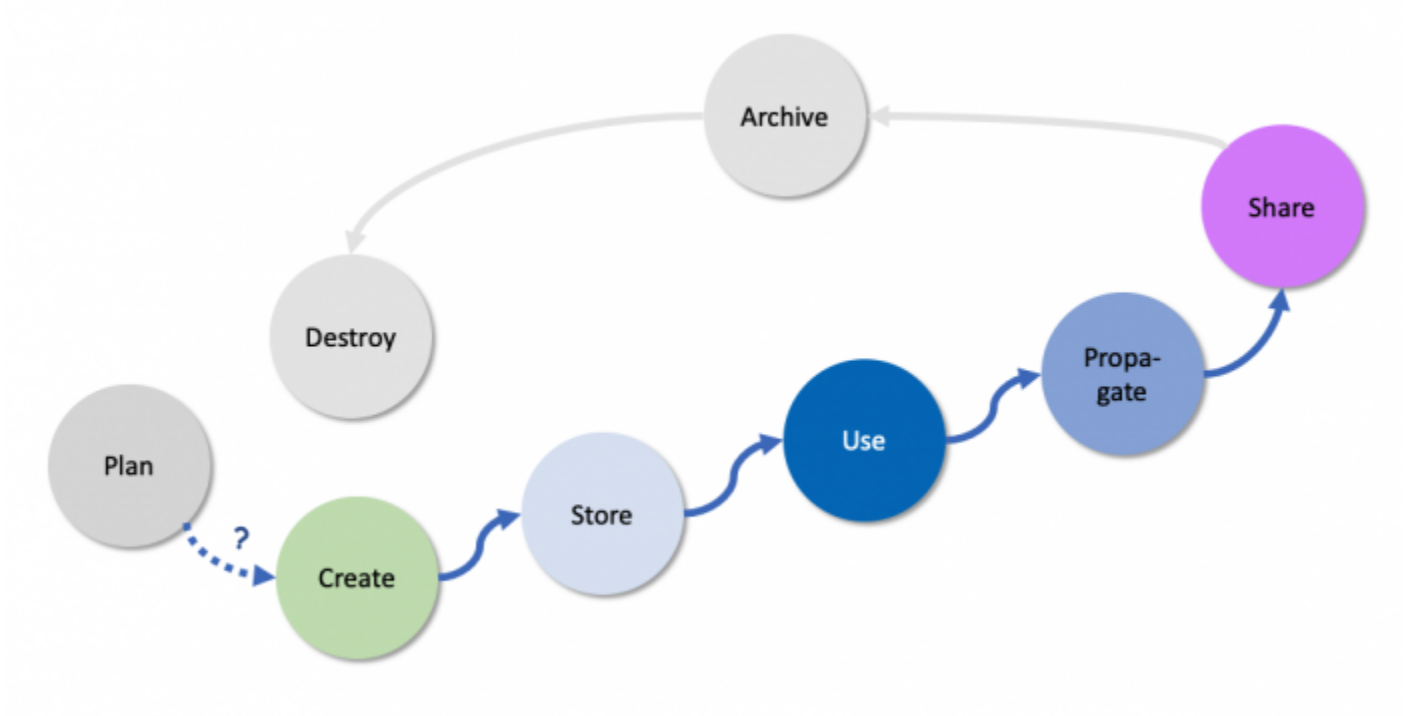

Figure 2: Immutable Data Lifecycle.