

Data

# 7 phases of a data life cycle

July 14, 2015

Most data management professionals would acknowledge that there is a data life cycle, but it is fair to say that there is no common understanding of what it is. If you Google “Data Life Cycle” you will not find anything that clearly describes it. But, if data management professionals know that there really is a Data Life Cycle, then it is incumbent on us to try to define it.

This is one attempt to describe the Data Life Cycle. It takes the position that a life cycle consists of phases, and each phase has its own characteristics. Einstein, when he was a teenager tried to think what it would be like to ride a beam of light. There is no chance that we can emulate Einstein, but perhaps we can put his idea to use. What would happen if we could ride on a piece of data as it moved through the enterprise? What new experiences would the piece of data have? What phases would it pass through?

## Discover more with Bloomberg newsletters

Subscribe now

### 1. Data Capture

The first experience that an item of data must have is to pass within the firewalls of the enterprise. This is Data Capture, which can be defined as the act of creating data values that do not yet exist and have never existed within the enterprise.

There are three main ways that data can be captured, and these are very important:

- **Data Acquisition:** the ingestion of already existing data that has been produced by an organization outside the enterprise
- **Data Entry:** the creation of new data values for the enterprise by human operators or devices that generate data for the enterprise
- **Signal Reception:** the capture of data created by devices, typically important in control systems, but becoming more important for information systems with the Internet of Things

There may well be other ways, but the three identified above have significant Data Governance challenges. For instance, Data Acquisition often involves contracts that govern how the enterprise is allowed to use the data it obtains in this way.

### 2. Data Maintenance

Once data has been captured it usually encounters Data Maintenance. This can be defined as the supplying of data to points at which Data Synthesis and Data Usage occur, ideally in a form that is best suited for these purposes.

We will deal with Data Synthesis and Data Usage in a moment. What Data Maintenance is about is processing the data without yet deriving any value from it for the enterprise.

It often involves tasks such as movement, integration, cleansing, enrichment, changed data capture, as well as familiar extract-transform-load processes.

Data Maintenance is the focus of a broad range of data management activities.

Because of this, Data Governance faces a lot of challenges in this area. Perhaps one of the most important is rationalizing how data is supplied to the end points for Data Synthesis and Data Usage, e.g. preventing proliferation of point-to-point transfers.

### **3. Data Synthesis**

This is comparatively new, and perhaps still not a very common phase in the Data Life Cycle. It can be defined as the creation of data values via inductive logic, using other data as input.

It is the arena of analytics that uses modeling, such as is found in risk modeling, actuarial modeling, and modeling for investment decisions. Derivation by deductive logic is not part of this – that occurs in Data Maintenance. An example of deductive logic is  $\text{Net Sales} = \text{Gross Sales} - \text{Taxes}$ . If I know Gross Sales and Taxes, and I know the simple equation just outlined, then I can calculate Net Sales.

Inductive logic requires some kind of expert experience, judgement, and/or opinion as a part of the logic, e.g. the way in which credit scores are created.

### **4. Data Usage**

So far we have seen how our in single data value has entered the enterprise via Data Capture, and has been moved around the enterprise, perhaps being transformed and enriched in Data Maintenance, and possibly being an input to Data Synthesis. Next, it reaches a point where it is used in support of the enterprise. This is Data Usage, which can be defined as the application of data as information to tasks that the enterprise needs to run and manage itself.

This would normally be tasks outside the data life cycle itself. However, data is becoming more central to business models in many enterprises. For instance, data may itself be a product or service (or part of a product or service) that the enterprise offers. This too is Data Usage, even if it is part of the Data Life Cycle, because it is part of the business model of the enterprise.

Data usage has special Data Governance challenges. One of them is whether it is legal to use the data in the ways which business people want. This is referred to as “permitted use of data”. There may be regulatory or contractual constraints on how data may actually be used, and part of the role of Data Governance is to ensure that these constraints are observed.

## 5. Data Publication

In being used, it is possible that our single data value may be sent outside of the enterprise. This is Data Publication, which can be defined as the sending of data to a location outside of the enterprise.

An example would be a brokerage that sends monthly statements to its clients. Once data has been sent outside the enterprise it is *de facto* impossible to recall it. Data values that are wrong cannot be corrected as they are beyond the reach of the enterprise. Data Governance may be needed to assist in deciding how incorrect data that has been sent out of the enterprise will be dealt with. Unhappily, data breaches also fall under Data Publication.

## 6. Data Archival

Our single data value may experience many rounds of usage and publication, but eventually the end of its life begins to loom large. The first part of this is to archive the data value. Data Archival is the copying of data to an environment where it is stored in case it is needed again in an active production environment, and the removal of this data from all active production environments.

A data archive is simply a place where data is stored, but where no maintenance, usage, or publication occurs. If necessary the data can be restored to an environment where one or more of these occur.

## 7. Data Purging

We now come to the actual end of life of our single data value. Data Purging is the removal of every copy of a data item from the enterprise.

Ideally, this will be done from an archive. A Data Governance challenge in this phase of the data life cycle is proving that the purge has actually been done properly.

## Critique

The terms we have used may be disputed. "Life Cycle" is not really accurate because data does not reproduce or recycle itself, which happens in real life cycles. "Data Life History" might be closer to the truth, but is not a familiar term. "Life History" is used to describe the phases of growth in an organism like a butterfly, but again data is different. Therefore, "Data Life Cycle" might as well be used.

What has been described here are phases with logical dependencies, not actual data flows. Data flows may go round and round through these phases, e.g. from Data Synthesis back to Data Maintenance and then returning to Data Synthesis and so on in more cycles. A description of these flows is quite different to the Data Life Cycle, though design should be informed by the Data Life Cycle.

Nor have environments been described. Some environments might be such that all phases of the Data Life Cycle occur in them, i.e. silos. However, it does seem reasonable that architecture should reflect the Data Life Cycle and that is a topic for another day.

Finally, data does not have to pass through all phases. Early mainframe systems had nothing more than Data Capture and Data Usage. Today, the full Data Life Cycle is more common.

What is important is that we define the Data Life Cycle because each phase has distinct Data Governance Needs. Greater clarity about the Data Life Cycle will help the mission of Data Governance.

*This article was written by Malcolm Chisholm from Information Management and was licensed by Bloomberg.*

## Recommended for you

---

### Treasury

#### **Bank treasury management evolution: From ALM challenges to Smart Treasury**

Following the series of bank failures in 2023, there is increased focus on liquidity risk management. Given that, how should institutions rethink their structures to ensure their sustainability in the changing rate environment?

---

### Markets

#### **How to unlock value with interconnected data?**

In this Market Dialogues episode, experts discuss navigating through data to decide which datasets to acquire and integrate, and how technology aids in managing the “data deluge”.

---

### Data

#### **Data Spotlight: Supply chain insights & more**

In this edition of Data Spotlight, we explore how investors can use AI to extract insights from news feeds and analyze supply chain data to understand risks and opportunities across companies, sectors, and regions.

Request a demo

Contact us

CONTACT

Americas +1 212 318 2000

EMEA +44 20 7330 7500

Asia Pacific +65 6212 1000

CLIENT ACCESS

Bloomberg Anywhere

Bloomberg Legal Entity Identifier

Bloomberg Vault

Enterprise Console

Entity Exchange

Stay Connected

Documentation

API Library

Webinars

CUSTOMER SUPPORT

Software Updates

Manage Products and Account Information

B-Unit Setup

FAQ

REGIONS

Brazil

Simplified Chinese

Traditional Chinese

Japan

Korea

Latin America

**Bloomberg Professional Services connect decision makers to a dynamic network of information, people and ideas.**

[Request a demo](#)

---

---

© 2024 Bloomberg Finance L.P. All rights reserved.

[Privacy](#)   [Terms](#)   [Tradebook Compliance](#)   [Company](#)   [Press](#)   [Careers](#)   [Cookie Preferences](#)   [AdChoices](#)