DATA MANAGEMENT

The official definition of data management is the “development and execution of architecture, policies, practises and procedures in order to manage the information lifecycle needs of an enterprise.

Data management comprises all the disciplines related to managing data as a valuable resource.

These disciplines include:

Back up/archiving/storage

Replication / de-duplication

Resource management

Backup, Archiving and Storage

Archiving means to remove from the on-line system those objects no longer in day-to-day use, and place them into long term, retrievable storage.
An archive is a collection of computer files that have been packaged together for backup, to transport to some other location, for saving away from the computer so that more hard disk storage can be made available, or for some other purpose. An archive can include a simple list of files or files organized under a directory or catalogue structure (depending on how a particular program supports archiving).

Back up means to make a copy of an object in the event the original becomes lost of damaged. Backups are useful primarily for two purposes. The first is to restore a state following a disaster (called disaster recovery). The second is to restore small numbers of files after they have been accidentally deleted or corrupted

Since a backup system contains at least one copy of all data worth saving, the data storage requirements are considerable. Organizing this storage space and managing the backup process is a complicated undertaking. A data repository model can be used to provide structure to the storage. In the modern era of computing there are many different types of data storage devices that are useful for making backups. There are also many different ways in which these devices can be arranged to provide geographic redundancy, data security, and portability.

Before data is sent to its storage location, it is selected, extracted, and manipulated. Many different techniques have been developed to optimize the backup procedure. These include optimizations for dealing with open files and live data sources as well as compression, encryption, and de-duplication, among others.

Storage is the place where data is held in an electromagnetic or optical form for access by a computer processor. Storage is frequently used to mean the devices and data connected to the computer through input/output operations - that is, hard disk and tape systems and other forms of storage that don't include computer memory and other in-computer storage.

On Line Storage

On line storage (or **Primary storage**), is the only storage directly accessible to the host/server. On-line storage is used for data in active use and it is high speed.

Near line storage

Near-line storage (Secondary storage), is the on-site storage of data on removable media. Near-line storage provides inexpensive, reliable, and unlimited data backup and archiving with somewhat less accessibility than with integrated online storage.

Off-line storage, also known as **disconnected storage or tertiary**, is a computer data storage on a medium or a device that is not under the control of a processing unit.-The medium is recorded, usually in a secondary or tertiary storage device, and then physically removed or disconnected.

Replication/ Data De-Duplication

Data de-duplication (often called "intelligent compression" or "single-instance storage") is a method of reducing storage needs by eliminating redundant data. Only one unique instance of the data is actually retained on storage media, such as disk or tape. Redundant data is replaced with a pointer to the unique data copy. For example, a typical email system might contain 100 instances of the same one megabyte (MB) file attachment. If the email platform is backed up or archived, all 100 instances are saved, requiring 100 MB storage space. With data de-duplication, only one instance of the attachment is actually stored; each subsequent instance is just referenced back to the one saved copy. In this example, a 100 MB storage demand could be reduced to only one MB.

Data de-duplication offers other benefits. Lower storage space requirements will save money on disk expenditures. The more efficient use of disk space also allows for longer data retention periods, which provides better recovery time objectives (RTO) for a longer period and reduces the need for tape backups. Data de-duplication also reduces the data that must be sent across a WAN for remote backups, replication, and disaster recovery.

Resource Management

Resource Management is a technique of managing information as a shared organizational resource. IRM includes: identification of information sources, type and value of information they provide, and ways of classification, valuation, processing, and storage of that information.