**Course: Advance Bio Informatics**

**Module Title: Data Management Life Cycle**

**Module No: 9**

This document reveals the information about the data management cycle.

Advancements in life science research has created gigantic amount of biological data in the form of DNA, RNA and Protein structures. In addition, due to the increasing usage of sequencing techniques, the contents of biological data are mounting exponentially day by day.

Biological Data and its Importance

To understand the contents of biological data in wet lab, we explain the cell as a part of biological system. At abstraction, organisms are classified into two broad categories i.e., eukaryotic organisms and prokaryotic organisms. Eukaryotic/multi-cellular organisms consist of multiple cells in their bodies such as human beings, plants, animals etc. On other hand, prokaryotic/unicellular organisms consist of a single cell such as bacteria, ameba, paramecium etc. Both categories of organisms are different due to their molecular composition. The bodies of these organisms consist of cells which are structurally alike. So, a cell is considered the most basic unit of an organism. This is segregated to various components for instance nucleus, plasma membrane, cytoplasm, nuclear membrane, mitochondria, Golgi apparatus and a lot of others. The structure of cell is shown as following.

**Cell**

**Nuclear Membrane**

**Plasma Membrane**

Structure of the Cell containing Central Dogma

The Figure shows various components of cell. Nucleus is segregated to chromosomes, nucleolus, and genes etc. Each component within the nucleus performs specific functions which have been exploited by molecular biologists.

Role of Central Dogma in Bioinformatics

In the Figure, we consider a sub-biological system, i.e., DNA,RNAnd protein structures, as an independent biological system at a higher level of abstraction. This sub-cellular system is composed of three objects i.e., DNA, RNA and Protein structures, and three processes i.e., replication, transcription and translation, which they perform during transformations. These three objects are translated in a sequence through these processes in a cell. The transformations of these objects make another complex sub-biological system that is known as central dogma. The process of central dogma starts from the nucleus of a cell, such as the replication of DNA and transcription of DNA into RNA is fulfilled in nucleus of cell. The process of central dogma ends in the cytoplasm at the surface of ribosomes, where the translation process of RNA into Protein structures is completed. This sub-cellular system initiated from the nucleus of the cell ends at the cytoplasm of the cell, i.e., the complete process is being done in the cell of an organism. The sub-biological system may perform different types of functions in the bodies of organism. Some of them include genetic characteristics, carry oxygen into the blood, converts food products into energy in small intestine, perform different constructive and destructive functions in muscles, bones, tissues, fibrils and blood etc.

In this sub-cellular system, biological information flows from DNA, RNA and Protein structures, as shown as following.

**Translation**

**Transcription**

**Replication**

DNA

Protein

RNA

DNA

Sequential Flow of Biological Data (Central Dogma)

Logical view of biological data in sub-cellular system is shown in Figure. In the figure, the arrow marks represent the input, output and direction of biological data, and rectangles represent the biological process, which occur during the conversion of these objects in a sub-cellular system.

The main characteristics which are differentiating biological data from traditional structured data include dynamicity, heterogeneous nature, polymorphic behavior and a sheer volume. The handling and management of this sheer volume of biological data has created a lot of issues and problems for data modeling research community as well as for bioinformaticians and computer scientists. Furthermore diversity of its data types and formats also created hindrance for research community.