**Course: Advanced Bioinformatics**

**Module title: Introduction to Biological Concepts**

**Module no. : 6**

Organism is the single source of biological data, from which our lives have been started. At abstract level, Organisms is classified into two broad categories 1- Plants and 2- is Animals. Animals are further sub-divided into Prokaryotic organisms as well as eukaryotics organisms. Most common sources of biological data are different types of plants, sea animals, flying birds, creeping animals etc. are some of them.

The bodies of living organisms based on DNA, RNA Protein structures which make the fundamental pillars of the bodies of all of the organisms.

Wet Labs are the physically housed laboratories, where biologists perform their physical experiments by using different chemicals, drugs and other materials. In this lab, biologists gather data and information of their experiments through samples of living organisms or on living organisms directly or indirectly like blood, urine, Siemens and de-oxyribonucleic acid (DNA) under certain circumstances or conditions. The data that is gathered by using physical experiments may consist of much anomalies and data replication among them, that is why it is referred to as the raw data.

Biologist brings raw data (which is also referred to as the uncured data) of their experiments to the dry labs. This uncured data is still in biological formats, and it needs curation for its further analysis and usage. In dry labs, first it is purified by using automatic sequencing tools, robots and with the help of biologists. In dry labs, the experimental work is performed through semi-experiments. There are little chances of error due to the automatic instruments. Dry Lab produces the data as output; it is in a purified form and can be used by computer scientists/ bioinformaticians in Computer Lab for analysis and other computing purposes. This type of experimental data is very much costly and is obtained from raw data, through a process of annotation. The annotated data is used for the derivation of various kinds of structure and function annotation, sequence determining and other structure predictions, data modeling, simulation and analysis that are done through computer should be done in Computer Lab.

The biological data available in computer lab may contain diversities of data formats. It may consists chains of sequences of nucleotides, DNA, RNA, Protein structures, genes, neurons, molecular structures, chemical structures, chemical reactions, graphs, hypothesis, scalars, audios, videos, texts, numeric, alpha-numerics and many others data types.

These contents of biological data are stored in diversities of data banks or local repositories like DDBJ, PDB, Knowledgebase, CCSB, Bio-UML, Open Tree of life, prototype selection generation etc.

Furthermore these databanks or repositories are lying at diverse locations, their contents are lying in fragmented or isolation form due to not having specific data formats, standards. Local research groups have made their own standards for future usage.