**Course: Advance Bio Informatics**

**Module Title: Databank Generation Challenges**

**Module No: 83**

**Databank Generation Challenges**

Bioinformatics experiments often need to interact with more than one database. Instead of querying every db manually, databank can be generated to give user a single physical access point. Databank generation is not so simple!

**Amount of Data**

Different forms of data available such as biological dbs, computed tomography (CT), magnetic resonance imaging (MRI) and X-rays. Data is huge in volume i.e. billions of sequences.

**Data Problems**

Databases for DNA, RNA, Protein, Protein Structure. Data is growing at rapid pace, so managing it and accommodating it in the databank is not easy. Lack of standard among different organizations so data is replicated.

**Database heterogeneity:** Where different names or data models are used in different databases for the same kind of data.

**Data Errors:** Data available in the databases may not always be accurate.

**Interdisciplinary nature**: of bioinformatics where biologists lack computer knowledge and computer scientists lack biology knowledge. There can be communication gap for databank generation.

**Huge amount** of disk space is needed for Databank.

**Data Updating**

Data in the original databases can be modified so to update databank, some strategy should be made.