**Course: Advance Bio Informatics**

**Module Title: Data Processing Engine**

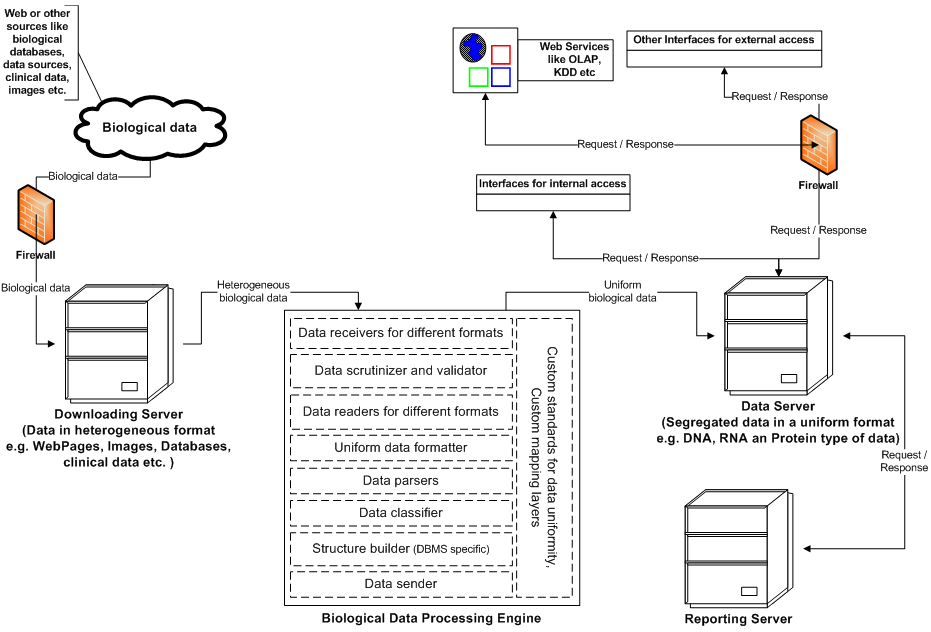
**Module No: 87**

**Data Processing Engine**

* Data downloaded by download server is usually in file format which is difficult to use.
* To be of some use, it must be preprocessed and converted into format that can be used.
* Data Processing Engine can be used for this purpose.

**Input** is data downloaded by download server. Data Processing Engine **converts** downloaded data from different files into format suitable to be used by users.

What format is **suitable**? Well, depends upon the requirement of its use. To convert data into usable format, downloaded data is parsed. Some parsers can be written to parse and convert downloaded data into usable format. Usually different parser is written for each file format. After converting data into desired format data can be stored into some db on some data server. Once data is stored into database, reports or user interface can be written so that users can use the data of generated databank.



**Data receivers** for different formats – Utilities for receiving data in different formats from *DS.*

**Data scrutinizer and validator** – Utility for validating, Examining, Filtering, trimming and sorting data

**Data readers -** for different formats – Utilities for reading data in different formats.

**Uniform data formatter -** Utility for formatting all types of data in a common and predefined data format

**Data Parser** – Utility for parsing formatted data to identify potential structures

**Data Classifier** – Utility to assign classification information to each structure.

**Structure Builder** – Utility to build underlined DBMS compliant structures

**Data Sender** – Utility for sending data to *Data Server*