

GALGOTIAS  
UNIVERSITY

# What is database????

- Database are convenient system to properly store, search and retrieve any type of data.
- A database helps to easily handle and share large amount of data and supports large scale analysis by easy access and data updating.

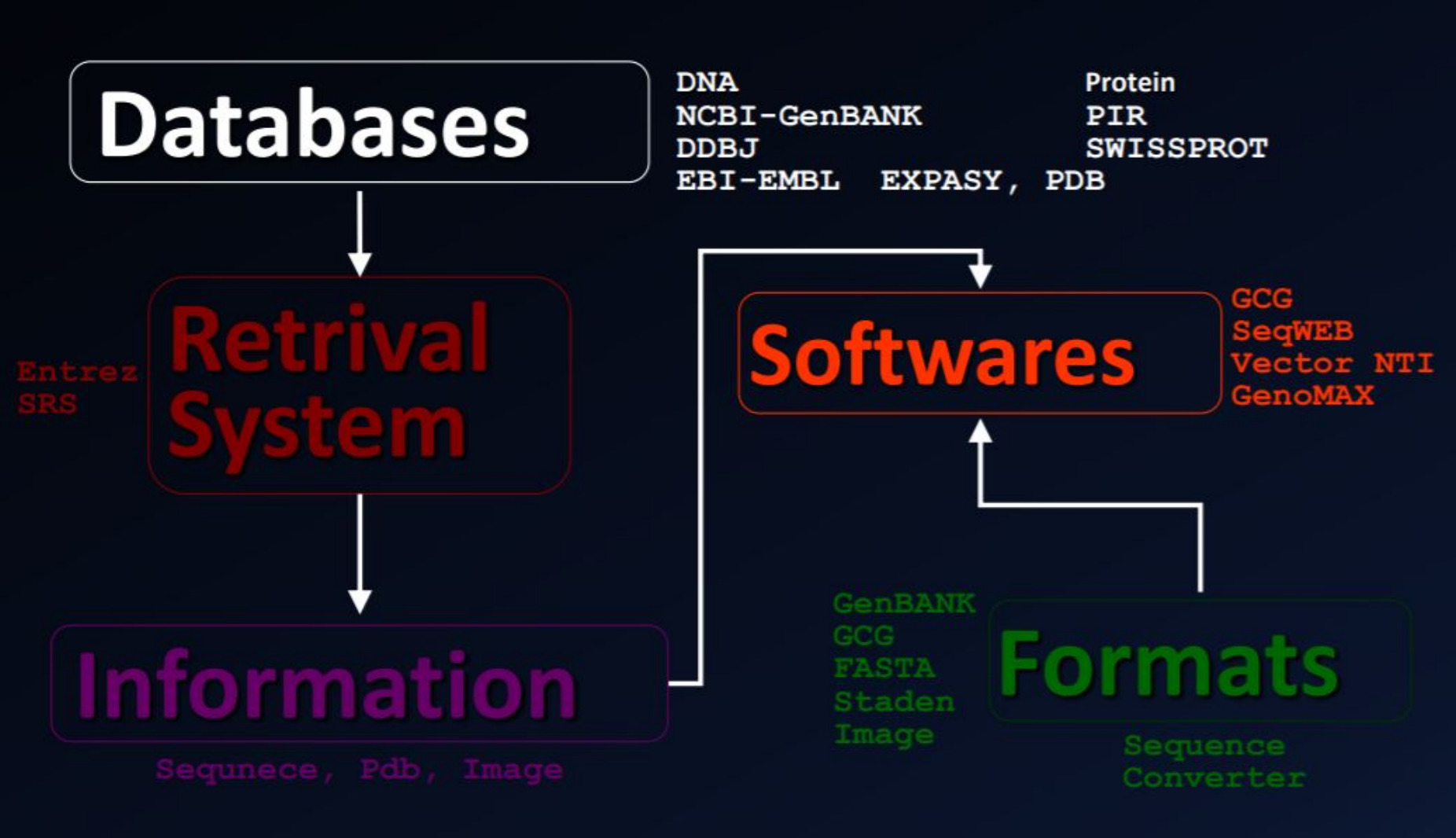
# Biological Database

Biological databases are libraries of **life sciences information** ,collected from **scientific experiments, published literature, high throughput experiment technology** and computational analysis.

They contain information from **genomics, proteomics, microarray gene expression.**

Information contained in biological databases includes **gene function, structure, localization(both cellular and chromosomal),biological sequences and structures.**

# Data retrieval



# Biological Databases

## **Type of databases**

## **Information Contain**

Bibliographic databases

Literature

Taxonomic databases

Classification

Nucleic acid databases

DNA information

Genomic databases

Gene level information

Protein databases

Protein information

Protein families, domains and functional sites

Classification of proteins and identifying domains

Enzymes/ metabolic pathways

Metabolic pathways

# Types of biological databases

1. Primary Database.
2. Secondary database.

# Primary databases

These are the primary sources of data used to store nucleic acid, protein sequences and structural information of biological macromolecules.

Some primary databases-

- NCBI(The National Centre for Biotechnology Information)
- GenBank
- DDBJ (DNA data bank of Japan)
- SWISS-PROT(Swiss-Prot )
- PIR (Protein Information Resource)
- PDB(Protein Data Bank)

This sequence collection of this database is due to the efforts of basic research from academic industrial and sequencing lab

# Secondary Database

A Secondary database contain additional information derived from the analysis of data available in primary sources.

Secondary databases are analysed in a variety of ways and contain different information in different formats.

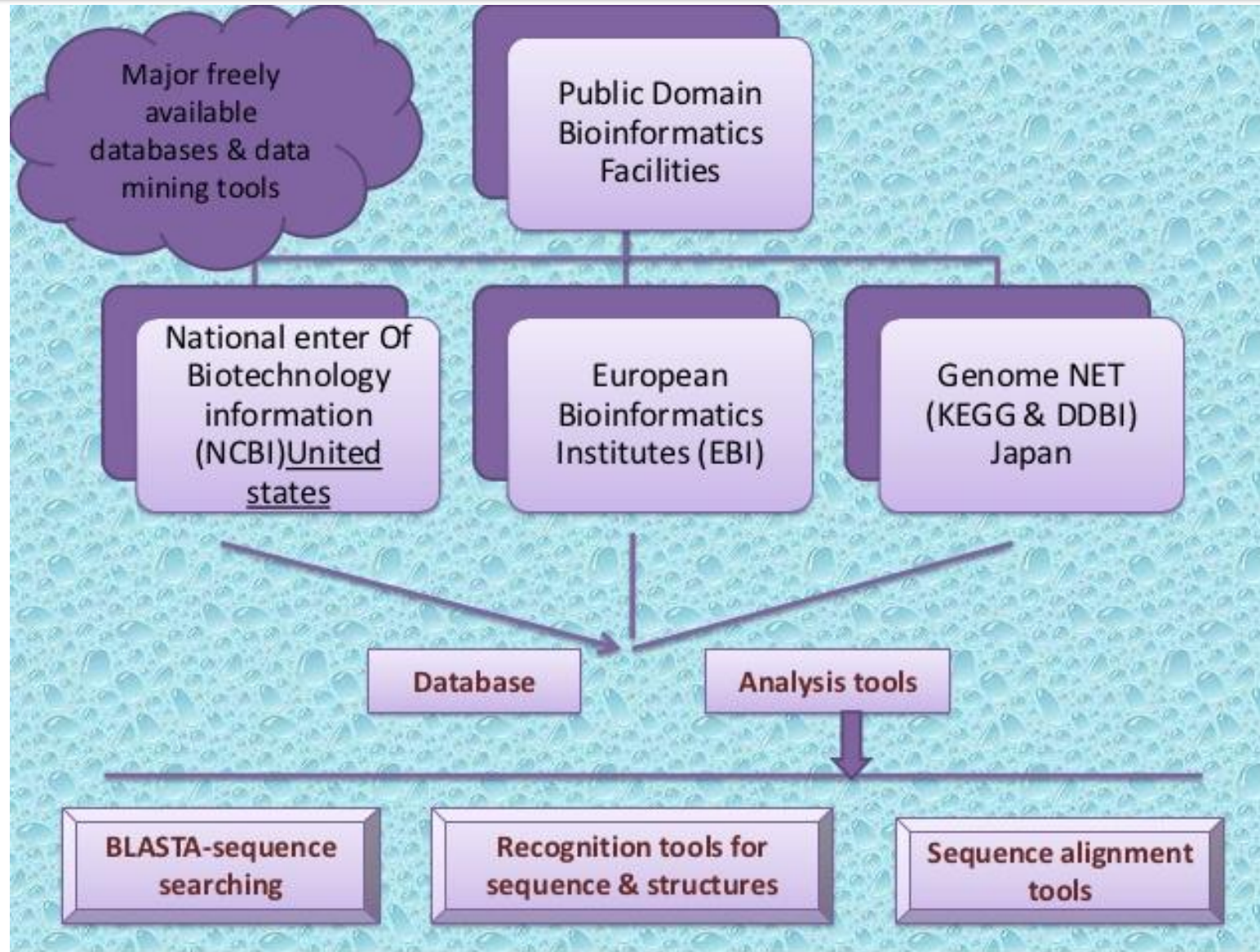
Some secondary databases:

- TrEMBL
- Pfam
- PROSITE
- Profiles
- SCOP



# Flat File Storage Data Formats

- When GenBank, EMBL and DDBJ formed a collaboration (1986), sequence databases had moved to a defined flat file format with a shared feature table format and annotation standards.
- The flat file formats from the sequence databases are still used to access and display sequence and annotation. They are also convenient for storage of local copies.



# List of Database

- DNA Data Bank of Japan (National Institute of Genetics)
- EMBL (European Bioinformatics Institute)
- GenBank (National Center for Biotechnology Information) UniProt Universal  
Resource (EBI, Swiss Institute of Bioinformatics, PIR)
- Swiss-Prot Protein Knowledgebase (Swiss Institute of Bioinformatics)
- National Center for Biotechnology Information (NCBI) NIM, USA



Thank You