DIGIT~Bio~TECH

**2019-1-BG01-KA203-062371**

**M.Sc. PROGRAMME "GREEN BIOTECHNOLOGY AND ICT"**

# SYLLABUS

## COURSE:

SCIENTIFIC RESOURCES OPEN ACCESS: DIGITAL DATABASES

## AUTHORS:

*Dr. Fabiano Chalhoub & Dr. Zoi Georgiou*

*University of West Attika, Greece*

| Academic work | Type | Number of classes |
|---|---|---|
| In-class work | Lectures | 30 |
| | Seminars | 20 |
| **Total in-class work** | | **50** |
| Out-of-class work | Presentations | 25 |
| | Projects | 20 |
| | Self-guided library/database work | 30 |
| **Total out-of-class work** | | **75*** |
| **Total of academic work** | | **125** |
| **ECTS credits in-class work** | | **2.0** |
| **ECTS credits out-of-class work** | | **3.0** |
| **Total ECTS credits** | | **5.0** |

| № | Grading | % of the grade |
|---|---|---|
| 1. | Workshops/discussions of reports and papers | 20 % |
| 2. | Case studies | 20 % |
| 3. | Homework assignments and tests | 10 % |
| 4. | Exams | 50 % |

* One credit corresponds to 25 hours of work.

| **Outline of the course** |
| --- |
| This LO encompasses the basic knowledge and historical evolution of Data, Data Management, and Databases. It explains why and how an intellectual concept used by human beings led to a widely used technology. It is also focused on how humanity went from memorizing to writing and keeping track of information and the birth of structured databases. The accent is made on why humanity had to find solutions to keep up with the exponential growth of information produced. Furthermore, it deals with the basic knowledge to be acquired in order to facilitate the communication with the Database Managers (DBA's) and help them to understand their requests in order to set up an appropriate database that will be efficient for their work. It is dedicated to the technical vocabulary used by computer science personnel and especially Data and DataBase Managers. The basic design of a DataBase, as well as the general frame of a DataBase and how to design it at a conceptual stage, are introduced. An overview of the main building blocks constitutive of a database is made. Finally, the basic Databases used in the scientific world are presented.<br>The LO is dedicated to designing a simple Database for scientific usage with examples to illustrate the way the structure has an influence on the expected result. It helps also in understanding the usage of the query language named SQL (Structured Query Language. It reviews the common Databases found in the market. Different types are available and here are described such as Commercial Databases, Shareware Databases, and Freeware Databases. An overview of the mode of usage of these Databases is also made. Furthermore, a deep dig into the open-access databases used in science is made. How to use and get the benefit of the existing knowledge is presented. |

**Educational goals**

Scientific information is both a researcher's greatest output and technological innovation's most important resource. On the other hand, Open access or OA is a set of principles and a range of practices through which research outputs are distributed online, free of cost or other access barriers. In this respect, the educational goals of this LO are to provide knowledge about:

- basic rules of using open access resources
- code of conduct related to the use and publishing of open access resources.

A scientific database is a computerized, organized collection of related data, which can be accessed for scientific inquiry and long-term storage. They allow the integration of divergent data sets and permit data to be analyzed in new ways, often across disciplines, making new types of scientific inquiry possible. Correspondingly, the educational goals of this LO are to provide knowledge about:

- advanced design of a Database
- the structure of a Database and how to make relations between Database tables
- specific language used to make queries (SQL) to retrieve data from a Database.

**Expected outcomes**

**Knowledge and Skills:**

As a result of the training students will be able to:

- demonstrate knowledge of what data are
- understand the basic vocabulary when using open-access resources
- structure a database
- use open access resources
- recognize the most common open-access databases used in science
- understand the mode of usage of different Scientific Databases
- design a simple Database for a scientific usage
- use database models' languages
- work with common databases found on the market.

**Problem-solving skills:** Decision making, creative thinking, analytical, research and interpreting skills

**Digital competencies and skills**: strategic web and database searching; data analysis and presentation; data management and preservation; digital communication; networks and file management.

**Personal skills:** initiative and independence, time management, good oral and written communication skills, teamwork.

*Syllabus*

| № | Topic | Number of classes |
|---|---|---|
| **1.** | **Unit 1 BL - Open access scientific resources** | **25 h** |
| 1.1 | Open access to scientific resources. An introduction to «Open Access» resources. | 2 h |
| 1.2 | An introduction to data (basic level). What are "data"? A brief history of data. Data prior to the invention of computers. Data in modern age. | 3 h |
| 1.3 | Understanding the basic vocabulary: Terminology. What are "Data" in the computer age? What is "Metadata"? What is a "Database"? What are "Tables" in a database? What are "Columns" in a database? What is a "Record"? What are "Indexes"? What is an "Object"? Structured data. Unstructured data. Big data. Analytics. Repository. | 6 h |
| 1.4 | Basic structure of a database: Introduction. Overview of a database architecture. Common database terms. What is the difference between major DBMS systems? | 3 h |
| 1.5 | Databases in the scientific word. Introduction to existing databases dedicated to science. Final thoughts. | 1 h |
| 1.56 | Seminars | 10 h |
| **2.** | **Unit 2 AL - Advanced structure of a Database** | **25 h** |
| 2.1 | Advanced structure of a Database: Database Management Systems. Database models' languages. Database characteristics. Database model types . Phases of database design. Degree of abstraction. Database schemas. External level. Conceptual level. Internal level . Data independence. | 3 h |
| 2.2 | The Relational Data Model. Basic Elements of a Relational Data Model. Characteristics of a Table. Differentiating features of the Relational Database Model. The Entity-Relationship Model. Examples of entity types and relationships in biological databases. Modification Anomalies. Key Definitions. | 3 h |
| 2.3 | Structured Query Language (SQL). SQL Commands List. | 1 h |
| 2.4 | Commercial and Free Databases used in the real world: Commercial Databases. SAP HANA. IBM Db2 Database. Oracle Database. | 2 h |
| 2.5 | Commercial and Free Databases used in the real world: Free Databases. MySQL. PostgreSQL. Microsoft SQL. MariaDB. Oracle. Firebirdsql. | 2 h |
| 2.6 | Overview of databases in the scientific world. DNA database. RNA databases. Protein databases. Disease databases. Expression databases. Pathway databases. The Medical Information Mart for Intensive Care (MIMIC) database. PCORnet. Open NHS. Database de-identification. | 3 h |
| 2.7 | Databases in the scientific world advanced module: The application of blockchain in Digital health. | 1 h |
| 2.8 | Seminars | 10 h |