



## IDENTIFYING DATA

### Data Cleaning and Manipulación

Subject	Data Cleaning and Manipulación			
Code	V02M123V01104			
Study programme	(*)Máster Universitario en Ciencias Biolóxicas: Bioloxía Molecular, Computacional e Ambiental e Bio-Industrias			
Descriptors	ECTS Credits	Type	Year	Quadmester
	3	Mandatory	1st	1st
Language	English			
Department				
Coordinator	Rodeiro Iglesias, Javier			
Lecturers	García Rosello, Emilio Rodeiro Iglesias, Javier			
E-mail	jrodeiro@uvigo.es			
Web				
General description	Structure and classification of data types. Elimination of noise in the data. Standardization of data. Standardization of data. Crossing data. Record linkage. Extracting knowledge from information.			

## Competencies

Type A	Code	Competences Specific
	A1	(*)To know the scientific method and the correct use of the scientific terminology as well as to acknowledge the contribution that scientific research provides to the overall knowledge and professional practice.
	A2	(*)Ability to describe and to analyse biological diversity, the mechanisms determining the interactions with the biotic and abiotic environment and being able to select those which might have technical applications.
	A3	(*)Ability to manage and/or to develop basic tools for validating and analysing data by means of statistics and bioinformatics.
	A4	(*)To know the ethical and legal aspects governing the collection and the handling of biological samples, organisms and habitats.
	A5	(*)Ability to design, evaluate and implement models of biological structures, systems and processes.
	A6	(*)To learn the sampling techniques and the instrumental methodologies, in the field and laboratory, for their application in the Biological Sciences
	A7	(*)To have an integrated view of the R&D processes and their possible transfer to the industrial sector. Planning and supervising facilities together with managing their human and economic resources.
	A8	(*)Ability to classify, evaluate, conserve, restore and manage natural and productive systems. Developing and implementing land management and sustainability plans.
	A9	(*)To understand and know how to apply quality control systems and safety protocols in any biological laboratory of the public or private sector.
	A10	(*)To acquire the professional ability to teach and spread Biology and to offer expertise advice for elaborating scientific, technical and socioeconomic biology reports. Address environmental consulting.

A11 (\*)To perform an individual Master Project (critical and in-depth study) under the supervision of a tutor in a research or working environment demonstrating that skills have been acquired.

#### Type B Code Competences Transversal

B1	(*)Dissemination of results and conclusions of the biological studies, in oral and written English, through complex presentations that address ideas related with R&D in Biology.
B2	(*)Managing computational, laboratory, field and industrial techniques in order to obtain, process and apply the acquired information.
B3	(*)Disseminating and broadcasting ideas in contexts both academic and non-specialised.
B4	(*)Reflecting on social and ethical responsibilities.

#### Learning aims

Subject competences	Typology	Competences
Ability to manage and/or to develop basic tools for validating and analysing data by means of statistics and bioinformatics.	Know How	A3
To acquire the professional ability to teach and spread Biology and to offer expertise advice for elaborating scientific, technical and socioeconomic biology reports. Address environmental consulting.	Know How	A10
Dissemination of results and conclusions of the biological studies, in oral and written English, through complex presentations that address ideas related with R&D in Biology.	Know How	B1
Managing computational, laboratory, field and industrial techniques in order to obtain, process and apply the acquired information.	Know How	B2

#### Contents

Topic	
Data Types	Structures of data. Data type definition and identification.
Data cleaning	Find and remove unwanted data
Record linkage	Data standarization. Crossing data.
Data processing	New knowledge from automatic data process.

#### Planning

	Personalized attention	Assessment	Ordinary class hours A	Face-to-face hours outside the classroom Guided academic environment B	Student's work factor C	Outside the classroom hours D	Total hours (A+B+D) E
Case studies / analysis of situations	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	25	0	0	15	40
Master Session	<input type="checkbox"/>	<input checked="" type="checkbox"/>	15	0	0	18	33
Reports / memories of practice	<input type="checkbox"/>	<input type="checkbox"/>	2	0	0	0	2
Total hours E:							75
Work load in UVIGO ECTS credits:							3

\*The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.

#### Methodologies

	Description
Case studies / analysis of situations	Case studies referring to any example of a big set of data and its processing.
Master Session	Theoretical contents

#### Personalized attention

## Description

Case studies / analysis of situations To resolve any issue or problem that may arise when elaborating the case studies

### Assessment

	Description	Qualification
Master Session	Attending the theoretical lectures	10
Case studies / analysis of situations	Presentation of the case study	90

### Other comments and second call

### Sources of information

Aho, A.V., Hopcroft, J.E., Ullman, J.A. (1987). Data Structures and Algorithms. Addison-Wesley.

Austern, M.H. (1999). Generic programming and the STL. Addison-Wesley.

Budd, T. (1998). data structures in C++ using the STL. Addison-Wesley.

Cline, M. Girou, M., Lomow, M. (1998). C++ FAQ (Second Edition). Addison-Wesley.

Deitel y Deitel, C++ How to program. Cuarta Edición (2003). Prentice Hall.

Eckel, B. Thinking in C++ (2a edición) (2000). Prentice-Hall.

Fdez-Valdivia, J., Garrido, A., Garcia-Silvente, M. (1998). Estructuras de Datos. Ediciones Gala.

Ford, W. and Topp, W. (1996). Data Structures with C++ Prentice-Hall.

Garrido A., Fdez-Valdivia, J. (2006) Abstracción y Estructuras de Datos en C++. Delta publicaciones

Garrido A. (2005), Fundamentos de programación en C++. Delta publicaciones.

Gilberg, R.F., Forouzan, B.A. (2001). Data structures: A pseudocode approach with C++. Brooks/Cole.

Josuttis, N.M. (1999). The C++ Standard Library: A Tutorial and Reference. Addison-Wesley.

Koenig, A., Moo, B.E. (2000). Accelerated C++: Practical programming by example. Addison-Wesley.

Musser, D.R. Saini, A. (1996) STL Tutorial and reference guide. Addison-Wesley.

Robson, R. (2000) Using the STL (2nd ed.) Springer Verlag.

Sedgewick, R. (1990). Algorithms in C++. Addison-Wesley.

Stroustrup B. (2002). El lenguaje de Programación C++. Edicion Especial. Addison-Wesley.

Weiss, M.A. (2000). Estructuras de datos en Java. Addison-Wesley.

### Recommendations