

# Introducción a los datos masivos y a la ciencia en abierto

**MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE**

***UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO***

Este documento puede utilizarse como documentación de referencia de esta asignatura para la solicitud de reconocimiento de créditos en otros estudios. Para su plena validez debe estar sellado por la Secretaría de Estudiantes UIMP.



# DATOS GENERALES

## Breve descripción

La presente materia tiene como objetivo proporcionar al estudiante una visión de global de lo que se conoce como Data Science, (o Ciencia de Datos), de los problemas de Big Data, y del contexto de Open Science (o Ciencia Abierta).

## Título asignatura

Introducción a los datos masivos y a la ciencia en abierto

## Código asignatura

102263

## Curso académico

2023-24

## Planes donde se imparte

[MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE](#)

## Créditos ECTS

6

## Carácter de la asignatura

OBLIGATORIA

## Duración

Cuatrimestral

## Idioma

Castellano e inglés

# CONTENIDOS

## Contenidos

1. Introducción al contexto de Open Science (Ciencia en Abierto)
2. Relevancia de los problemas de Big Data en Open Science
3. El enfoque desde Ciencia de Datos
4. Ejemplos relevantes de la aplicación de Ciencia de Datos en Open Science
5. Arquitectura de las soluciones
6. Descripción de las e-Infraestructuras y el correspondiente Middleware
7. Panorámica de casos de uso en las áreas de Salud, Medio Ambiente, Urbanismo, Economía, Astrofísica, Ciencias Sociales, Humanidades, Gestión Pública
8. Problemas y desafíos
9. Perfiles profesionales en el ecosistema de la Ciencia de datos
10. Nuevas tendencias y desarrollos
11. Proyectos e iniciativas, instituciones y empresas implicadas

# RESULTADOS DE APRENDIZAJE Y DE FORMACIÓN

## Generales

CG1 - Integrarse eficazmente en un grupo de trabajo y trabajar en equipo, compartir la información disponible e integrar su actividad en la actividad del grupo colaborando de forma activa en la consecución de objetivos comunes

CG2 - Capacidad de estudio, síntesis y autonomía suficientes para desarrollar de forma autónoma proyectos básicos de investigación

CG4 - Saber preparar y conducir presentaciones, ante públicos especializado, sobre una investigación o proyecto científico

CG6 - Buscar, obtener, procesar, comunicar información y transformarla en conocimiento

CG7 - Conocer las herramientas metodológicas necesarias para desarrollar proyectos avanzados

CG8 - Capacidad de actualización de los conocimientos expuestos en el ámbito de la comunidad científica

## Transversales

CT1 - Analizar y combinar información utilizando diferentes fuentes

CT2 - Conocer la problemática ética y legal relacionada con el análisis de datos y entender su importancia para una sociedad basada en los valores de la libertad, la justicia, la igualdad y el pluralismo

CT6 - Capacidades asociadas al trabajo en equipo: cooperación, liderazgo, saber escuchar

## Específicas

DSDA01 - Utilizar el análisis predictivo para analizar grandes volúmenes de datos y descubrir nuevas relaciones

DSDA04 - Investigar y analizar conjuntos de datos complejos, combinando diferentes fuentes y tipos de datos para mejorar el análisis global

DSDA06 - Capacidad de representación de datos variables y complejos para su visualización

DSENG01 - Aplicar los principios de ingeniería a la investigación, diseño y desarrollo de un prototipo de aplicaciones de análisis de datos, o al desarrollo de estructuras, instrumentos, máquinas, experimentos, procesos, sistemas requeridos para ello

DSRM01 - Crear nuevas visiones y capacidades mediante el uso del método científico (hipótesis, prueba y evaluación)

DSRM02 - Realizar un estudio sistemático dirigido hacia un conocimiento más completo o la comprensión de los hechos observables, y descubrir nuevos enfoques para lograr los objetivos en investigación o de organización

## PLAN DE APRENDIZAJE

### Actividades formativas

AF1 - Participación y asistencia a lecciones magistrales y seminarios (30 horas - 100% presencialidad)

AF2 - Realización de prácticas de computación y análisis de datos (15 horas - 100% presencialidad)

AF3 - Desarrollo de proyectos guiados (40 horas - 20% presencialidad)

AF6 - Tutorías (presenciales o por medio de recursos telemáticos) (10 horas - 30% presencialidad)

AF7 - Elaboración de informes de laboratorio y trabajos (12 horas - 0% presencialidad)

AF8 - Estudio individual de contenidos de la asignatura (30 horas - 0% presencialidad)

AF9 - Trabajo en grupo (10 horas - 0% presencialidad)

A10 - Pruebas de evaluación (3 horas - 100% presencialidad)

### Metodologías docentes

Se comenzará por una exposición de los conceptos básicos, incluyendo ejemplos sencillos pero relevantes, que serán analizados individualmente y discutidos en común.

Se revisarán los diferentes componentes de una solución, y los actores que participan en el desarrollo de la misma.

Los estudiantes, organizados en grupos, realizarán un análisis detallado de un caso de estudio en una de las áreas comentadas (Salud, Medio Ambiente, Urbanismo, Economía, Astrofísica, Ciencias Sociales, Humanidades, Gestión Pública).

Se invitará puntualmente a profesionales destacados en cada temática para aportar un punto de vista práctico y cercano sobre la situación actual y retos profesionales concretos.

### Resultados de aprendizaje

- Identificar problemas de Big Data en el contexto Open Science que requieren técnicas de Ciencia de Datos para su resolución.

- Distinguir entre los diferentes componentes de dificultad de un problema de Big Data.
- Saber elegir fuentes de información en el contexto de Open Science.
- Identificar los elementos requeridos para abordar una solución completa, incluyendo la arquitectura y los diferentes actores implicados.
- Conocer las nuevas técnicas en Data Science y las diferentes iniciativas en marcha e instituciones y empresas implicadas.
- Identificar el perfil profesional requerido en el contexto de Data Science.

# SISTEMA DE EVALUACIÓN

## Descripción del sistema de evaluación

SE1 - Examen (escrito, oral y/o práctico en el aula de computación) (40%)

SE2 - Valoración de informes y trabajos escritos (60%)

## PROFESORADO

### Profesor responsable

**Lloret Iglesias, Lara**

*Científica Titular del CSIC*

*Instituto de Física de Cantabria (IFCA), Consejo Superior de Investigaciones Científicas*

### Profesorado

**Sánchez Cruz, Sergio**

*Doctor en Ciencia de Materiales*

*Universität Zürich*

**Pando de la Hoz, Francisco**

*Doctor en Ciencias biológicas.*

*Investigador Titular.*

*Real Jardín Botánico, CSIC*

**Fernández Fernández, Jesús**

*Doctor en Ciencias Físicas.*

*Científico Titular del CSIC*

*Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria*

**del Jesús Peñil, Manuel**

*Doctor Ingeniero de Caminos, Canales y Puertos.*

*Profesor titular del Departamento de Ciencias y Técnicas del Agua y del Medio Ambiente.*

*Instituto de Hidráulica Ambiental de la Universidad de Cantabria.*

**Herrera García, Sixto**

*Doctor por la Universidad de Cantabria. Programa: Matemáticas y Computación.*

*Profesor Titular de Universidad (Matemática Aplicada).*

*Universidad de Cantabria. Dpto. de Matemática Aplicada y Ciencias de la Computación (Caminos).*

**de Castro García, Pablo María**

*Doctor por la Universidad de Cantabria.*

*Director Corporativo de LIFE FOR TYRES S.L. y Profesor Asociado de la Universidad de Cantabria en el Área de Sistemas de Información del Departamento de Administración de Empresas.*

*Universidad de Cantabria / LIFE FOR TYRES S.L..*

**Castrillo Melguizo, María**

*Doctora en Ingeniería Ambiental*

*Personal Científico Titular*

*Instituto de Física de Cantabria*

**Cofiño González, Antonio Santiago**

*DOCTOR.*

*CIENTÍFICO TITULAR.*

*INSTITUTO DE FÍSICA DE CANTABRIA, CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS*

**San Martín Segura, Daniel**

*INGENIERO DE TELECOMUNICACIONES (UNIVERSIDAD DE CANTABRIA)*

*MÁSTER EN MATEMÁTICAS Y COMPUTACIÓN (UNIVERSIDAD DE CANTABRIA).*

*Gerente*

*Predictia Intelligent Data Solutions S.L.*

**Orviz Fernández, Pablo**

*Investigador*

*Instituto de Física de Cantabria (IFCA), CSIC-UC*

**Aguilar Gómez, Fernando**

*Investigador Postdoctoral*

*Instituto de Física de Cantabria (IFCA), CSIC-UC*

**Vischia , Pietro**

*Investigador Ramón y Cajal Senior  
Universidad de Oviedo y ICTEA*

**Palacio Hoz, Aída**

*IT Support Administrator  
Instituto de Física de Cantabria (IFCA), CSIC-UC*

**García Manzananas, Rodrigo**

*Licenciado en Física.  
Doctor en Ciencia, Tecnología y Computación.  
Universidad de Cantabria.*

**Heredia Cacha, Ignacio**

*Master en Física.  
Contratado a cargo de proyectos de investigación.  
Instituto de Física de Cantabria (IFCA)*

**Varela Egocheaga, Ignacio Alejandro**

*Profesor Contratado Doctor  
Universidad de Cantabria (UC)*

**Méndez Incera, Fernando Javier**

*Profesor Titular de Ingeniería Hidráulica  
Universidad de Cantabria (UC)*

**Güemes Seoane, Celestino**

*Responsable de Soluciones e Innovación.  
Worldgrid España.*

**Mañana Canteli, Mario**

*Universidad de Cantabria*

# HORARIO

## Horario

02/10/2023

15:30 - 16:30

Presentación Data Science

Lara Lloret Iglesias

Científica Titular del CSIC

Instituto de Física de Cantabria (IFCA), Consejo Superior de Investigaciones Científicas

16:30 - 17:30

Introduccion a DataScienceHub

Aída Palacio Hoz

IT Support Administrator

Instituto de Física de Cantabria (IFCA), CSIC-UC

17:30 - 18:30

Jupyter

Antonio Santiago Cofiño González

DOCTOR.

CIENTÍFICO TITULAR.

INSTITUTO DE FÍSICA DE CANTABRIA, CONSEJO SUPERIOR DE INVESTIGACIONES  
CIENTÍFICAS

18:30 - 19:30

Bash + Shell

Jesús Fernández Fernández

Doctor en Ciencias Físicas.  
Científico Titular del CSIC  
Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria

03/10/2023

15:30 - 17:30

Introducción a Git

Pablo Orviz Fernández

Investigador  
Instituto de Física de Cantabria (IFCA), CSIC-UC

17:30 - 19:30

Data Science y Biodiversidad

Francisco Pando de la Hoz

Doctor en Ciencias biológicas.  
Investigador Titular.  
Real Jardín Botánico, CSIC

04/10/2023

15:30 - 17:30

Introducción a Git

Pablo Orviz Fernández

Investigador  
Instituto de Física de Cantabria (IFCA), CSIC-UC

17:30 - 19:30

Introducción a Python

Ignacio Heredia Cacha

Master en Física.  
Contratado a cargo de proyectos de investigación.  
Instituto de Física de Cantabria (IFCA)

05/10/2023

15:30 - 17:30

Introducción a Python

Ignacio Heredia Cacha

Master en Física.  
Contratado a cargo de proyectos de investigación.  
Instituto de Física de Cantabria (IFCA)

17:30 - 19:30

Introducción a Python

Ignacio Heredia Cacha

Master en Física.  
Contratado a cargo de proyectos de investigación.  
Instituto de Física de Cantabria (IFCA)

06/10/2023

15:30 - 17:30

Introducción a R

Rodrigo García Manzananas

Licenciado en Física.  
Doctor en Ciencia, Tecnología y Computación.  
Universidad de Cantabria.

17:30 - 19:30

Introducción a Python

Ignacio Heredia Cacha

Master en Física.  
Contratado a cargo de proyectos de investigación.  
Instituto de Física de Cantabria (IFCA)

09/10/2023

15:30 - 17:30

Introducción a R

Rodrigo García Manzananas

Licenciado en Física.  
Doctor en Ciencia, Tecnología y Computación.  
Universidad de Cantabria.

17:30 - 19:30

Aprendizaje máquina para aplicaciones hidrológicas

Manuel del Jesús Peñil

Doctor Ingeniero de Caminos, Canales y Puertos.  
Profesor titular del Departamento de Ciencias y Técnicas del Agua y del Medio Ambiente.  
Instituto de Hidráulica Ambiental de la Universidad de Cantabria.

10/10/2023

15:30 - 17:30

Introducción a R

Sixto Herrera García

Doctor por la Universidad de Cantabria. Programa: Matemáticas y Computación.  
Profesor Titular de Universidad (Matemática Aplicada).  
Universidad de Cantabria. Dpto. de Matemática Aplicada y Ciencias de la Computación  
(Caminos).

11/10/2023

17:30 - 19:30

Panorama sensores

María Castrillo Melguizo

Doctora en Ingeniería Ambiental  
Personal Científico Titular  
Instituto de Física de Cantabria

13/10/2023

15:30 - 17:30

Introducción a R

Sixto Herrera García

Doctor por la Universidad de Cantabria. Programa: Matemáticas y Computación.  
Profesor Titular de Universidad (Matemática Aplicada).  
Universidad de Cantabria. Dpto. de Matemática Aplicada y Ciencias de la Computación  
(Caminos).

17:30 - 19:30

Sistemas eléctricos de potencia

Mario Mañana Canteli

Universidad de Cantabria

17/10/2023

15:30 - 17:30

Computación distribuida

Aída Palacio Hoz

IT Support Administrator  
Instituto de Física de Cantabria (IFCA), CSIC-UC

17:30 - 18:30

Panorama en Meteorología y Clima

Daniel San Martín Segura

INGENIERO DE TELECOMUNICACIONES (UNIVERSIDAD DE CANTABRIA)  
MÁSTER EN MATEMÁTICAS Y COMPUTACIÓN (UNIVERSIDAD DE CANTABRIA).  
Gerente  
Predictia Intelligent Data Solutions S.L.

18:30 - 19:30

Panorama en Meteorología y Clima

Rodrigo García Manzananas

Licenciado en Física.  
Doctor en Ciencia, Tecnología y Computación.  
Universidad de Cantabria.

18/10/2023

17:30 - 19:30

Análisis de datos biológicos en genómica del cáncer

Ignacio Alejandro Varela Egocheaga

Profesor Contratado Doctor  
Universidad de Cantabria (UC)

19/10/2023

15:30 - 17:30

Big Data: conceptos generales e impacto en los negocios

Celestino Güemes Seoane

Responsable de Soluciones e Innovación.  
Worldgrid España.

20/10/2023

15:30 - 17:30

Big data en el LHC

Sergio Sánchez Cruz

Doctor en Ciencia de Materiales  
Universität Zürich

17:30 - 19:30

Big data en el LHC

Pietro Vischia

Investigador Ramón y Cajal Senior  
Universidad de Oviedo y ICTEA

24/10/2023

15:30 - 17:30

Data Diving: Aplicaciones de Minería de Datos en Ingeniería Oceanográfica

Fernando Javier Méndez Incera

Profesor Titular de Ingeniería Hidráulica  
Universidad de Cantabria (UC)

17:30 - 19:30

IoT

Pablo de Castro

Doctor por la Universidad de Cantabria.

Director Corporativo de LIFE FOR TYRES S.L. y Profesor Asociado de la Universidad de Cantabria en el Área de Sistemas de Información del Departamento de Administración de Empresas.

Universidad de Cantabria / LIFE FOR TYRES S.L..

25/10/2023

17:30 - 19:30

Open Access

Fernando Aguilar Gómez

Investigador Postdoctoral

Instituto de Física de Cantabria (IFCA), CSIC-UC

26/10/2023

17:30 - 19:30

Evaluación

Lara Lloret Iglesias

Científica Titular del CSIC

Instituto de Física de Cantabria (IFCA), Consejo Superior de Investigaciones Científicas

# BIBLIOGRAFÍA Y ENLACES RELACIONADOS

## Bibliografía

<https://masterdatascience.ifca.es/>