El ciclo de vida de los datos: de la adquisición a la presentación

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE

UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO

Másteres Universitarios

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 asignatura aborda los puntos más importantes del ciclo de vida de los datos científicos desde la adquisición a la presentación y posible re-uso, haciendo especial énfasis en datos en abierto.

Título asignatura

El ciclo de vida de los datos: de la adquisición a la presentación

Código asignatura

102267

Curso académico

2024-25

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. Diferentes aproximaciones al Ciclo de Vida de los Datos.
2. Planificación de la gestión de los datos: DMPs.
3. Herramientas de gestión de Metadatos.
4. Ingestión de datos. Ejemplos multidisciplinares.
5. Publicación de datos en abierto.
6. Plataformas de re-uso de datos en abierto.
7. Preservación de los datos.

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

CG5 - Planificar, diseñar y poner en marcha un proyecto avanzado

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

CT5 - Capacidad de trabajo autónomo y toma de decisiones

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

Específicas

DSDM01 - Desarrollar e implementar una estrategia de gestión de datos, en particular, en la forma de un plan de gestión de datos (DMP)

DSDM02 - Desarrollar e implementar modelos de datos, incluidos los metadatos

DSDM03 - Recoger e integrar diferentes fuentes de datos y su ingestión para su posterior análisis

DSDM05 - Asegurar la calidad de los datos, su accesibilidad, y su forma de publicación (curación)

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

DSENG02 - Desarrollar y aplicar soluciones computacionales para problemas en un cierto dominio de aplicación, usando una amplia gama de plataformas de análisis de datos

PLAN DE APRENDIZAJE

Actividades formativas

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

AF2 - Realización de prácticas de computación y análisis de datos (20 horas)

AF3 - Desarrollo de proyectos guiados

AF6 - Tutorías (presenciales o por medio de recursos telemáticos) (8 horas)

AF7 - Elaboración de informes de laboratorio y trabajos (20 horas)

AF8 - Estudio individual de contenidos de la asignatura (20 horas)

AF9 - Trabajo en grupo (48 horas)

A10 - Pruebas de evaluación (4 horas)

Metodologías docentes

La asignatura comenzará por una exposición de los conceptos básicos, incluyendo ejemplos sencillos pero relevantes, que serán analizados individualmente y discutidos en común, e implementados en lenguaje Python y/o R.

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 empleando las diferentes tecnologías discutidas, tales como SQL, OLAP, noSQL, DMPtool, plataformas de soporte del ciclo de vida de los datos.

Resultados de aprendizaje

- Modelar un problema de datos, recoger su semántica, relaciones y restricciones.
- Seleccionar la tecnología de gestión de datos más adecuada de acuerdo al volumen de datos a manejar, el objeto de su uso y las restricciones de seguridad e integridad que impone el problema de datos.
- Diseñar e implementar el ciclo de vida de un sistema de gestión de datos, desde la recogida de requisitos hasta su implementación.
- Interrogar fuentes de datos mediante el lenguaje estándar SQL y su extensión OLAP, así como utilizando lenguajes específicos usados por tecnologías más recientes.
- Realizar procesos de extracción, transformacióny carga de datos para la visualización y análisis de datos.

- Diseñar y ejecutar un DMP (Data Management Plan).

- Operar modelos de datos, incluyendo metadatos.
- Integrar diferentes fuentes de datos, visualizar la conexión entre las mismas y comparar diferentes fuentes para realizar nuevos análisis.

SISTEMA DE EVALUACIÓN

Descripción del sistema de evaluación

SE1 - Examen (escrito, oral y/o práctico en el aula de computación) (30%)

SE2 - Valoración de informes y trabajos escritos (55%)

SE3 - Valoración de exposiciones orales de trabajos (15%)

Calendario de exámenes

La evaluación continua de actividades presenciales es durante el cuatrimestre y la presentación del trabajo en grupo y la exposición oral es al final del cuatrimestre. Test de progreso al final del cuatrimestre.

PROFESORADO

Profesor responsable

Aguilar Gómez, Fernando

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

Profesorado

Román Espino, Jorge

Director General
Velfair S.A.

Cabrillo Bartolomé, José Ibán

Doctor en Ciencia y Tecnología, Master en computación Científica, Licenciado en Ciencias Físicas.
Tecnólogo de OPIs. Responsable de Computación Avanzada del IFCA.
Instituto de Física de Cantabria. (CSIC-IFCA).

de la Vega Ruiz, Alfonso

Doctor en Ciencia y Tecnología
Profesor Ayudante Doctor del área de Lenguajes y Sistemas Informáticos
Facultad de Ciencias de la Universidad de Cantabria

Rodríguez González, David

DOCTOR.
Investigador distinguido.
Instituto de Física de Cantabria, CSIC.

Bernal Martínez, Isabel

Responsable General de Digital
Unidad de Recursos de Información Científica para la Investigación
Consejo Superior de Investigaciones Científicas (CSIC)

HORARIO

Horario

12/11/2024

15:30 - 17:30

Introducción Data Life Cycle

Fernando Aguilar Gómez

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

13/11/2024

17:30 - 19:30

Proyectos, Requerimientos

Fernando Aguilar Gómez

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

14/11/2024

15:30 - 17:30

Planning, DMP

Fernando Aguilar Gómez

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

19/11/2024

17:30 - 19:30

Intro Metadatos

Fernando Aguilar Gómez

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

20/11/2024

15:30 - 17:30

Práctica Metadatos

Fernando Aguilar Gómez

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

21/11/2024

15:30 - 17:30

Identificadores (PID, DOI)

Fernando Aguilar Gómez

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

22/11/2024

15:30 - 17:30

Pácticas PID, protocolo HTTP

Fernando Aguilar Gómez

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

26/11/2024

17:30 - 19:30

Teoría repositorios

Fernando Aguilar Gómez

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

27/11/2024

17:30 - 19:30

APIs y repositorios

Fernando Aguilar Gómez

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

28/11/2024

15:30 - 17:30

Curation I

José Ibán Cabrillo Bartolomé

Doctor en Ciencia y Tecnología, Master en computación Científica, Licenciado en Ciencias Físicas.
Tecnólogo de OPIs. Responsable de Computación Avanzada del IFCA.
Instituto de Física de Cantabria. (CSIC-IFCA).

03/12/2024

15:30 - 17:30

Curation II

José Ibán Cabrillo Bartolomé

Doctor en Ciencia y Tecnología, Master en computación Científica, Licenciado en Ciencias Físicas.
Tecnólogo de OPIs. Responsable de Computación Avanzada del IFCA.
Instituto de Física de Cantabria. (CSIC-IFCA).

05/12/2024

17:30 - 19:30

Curation III

José Ibán Cabrillo Bartolomé

Doctor en Ciencia y Tecnología, Master en computación Científica, Licenciado en Ciencias Físicas.
Tecnólogo de OPIs. Responsable de Computación Avanzada del IFCA.
Instituto de Física de Cantabria. (CSIC-IFCA).

10/12/2024

15:30 - 17:30

ETL

David Rodríguez González

DOCTOR.
Investigador distinguido.
Instituto de Física de Cantabria, CSIC.

12/12/2024

15:30 - 17:30

ETL

David Rodríguez González

DOCTOR.
Investigador distinguido.
Instituto de Física de Cantabria, CSIC.

19/12/2024

15:30 - 17:30

Gestión de datos en repositorios

Isabel Bernal Martínez

Responsable General de Digital
Unidad de Recursos de Información Científica para la Investigación
Consejo Superior de Investigaciones Científicas (CSIC)

07/01/2025

15:30 - 17:30

Preservation

José Ibán Cabrillo Bartolomé

Doctor en Ciencia y Tecnología, Master en computación Científica, Licenciado en Ciencias Físicas.
Tecnólogo de OPIs. Responsable de Computación Avanzada del IFCA.
Instituto de Física de Cantabria. (CSIC-IFCA).

09/01/2025

15:30 - 17:30

Preservation

José Ibán Cabrillo Bartolomé

Doctor en Ciencia y Tecnología, Master en computación Científica, Licenciado en Ciencias Físicas.
Tecnólogo de OPIs. Responsable de Computación Avanzada del IFCA.
Instituto de Física de Cantabria. (CSIC-IFCA).

10/01/2025

17:30 - 19:30

Preservation

José Ibán Cabrillo Bartolomé

Doctor en Ciencia y Tecnología, Master en computación Científica, Licenciado en Ciencias Físicas.
Tecnólogo de OPIs. Responsable de Computación Avanzada del IFCA.
Instituto de Física de Cantabria. (CSIC-IFCA).

14/01/2025

15:30 - 17:30

ORM

Fernando Aguilar Gómez

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

16/01/2025

15:30 - 17:30

ORM

Fernando Aguilar Gómez

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

17/01/2025

15:30 - 17:30

Práctica DLC

Fernando Aguilar Gómez

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

21/01/2025

15:30 - 17:30

Presentación de trabajos

Fernando Aguilar Gómez

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

17:30 - 19:30

Repaso general

Fernando Aguilar Gómez

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

23/01/2025

15:30 - 17:30

Examen

Fernando Aguilar Gómez

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

BIBLIOGRAFÍA Y ENLACES RELACIONADOS

Bibliografía

Presentaciones de la asignatura disponibles en el entorno online (MOODLE o similar).