

Esta sesión será grabada

En cumplimiento de lo dispuesto en el Reglamento (UE) 2016/679, General de Protección de Datos de Carácter Personal, le informamos de que esta sesión será grabada. La grabación puede contener su imagen, voz, nombre y/o sus mensajes de texto intercambiados durante la sesión, y podrá ser publicada en el Campus Virtual de la Universitat Oberta de Catalunya, **Aula 2** de la asignatura **M2.851 Tipología y ciclo de vida de los datos**, semestre 2022-1.



Tipología y ciclo de vida de los datos

M2.851 - Aula 2 - Semestre 2022-1

Isabel Guitart - *Prof. responsable*

José Moreira - *Prof. colaborador*

Equipo docente TiCVD 2022-1

Profesora responsable de la asignatura

- M. Isabel Guitart

Profesores colaboradores:

- Laia Subirats (M2.851, Aula 1, Castellano)
- José Moreira (M2.851, Aula 2, Castellano)
- Mireia Calvo (M2.951, Aula 1, Catalán)

Equipo docente TiCVD 2022-1

Breve presentación, José Moreira:

- Ph.D., Ingeniería de Telecomunicación
- Actualmente actividad en empresa privada, tecnologías *blockchain*
- Áreas de trabajo/investigación
 - Teoría de códigos y de la información
 - Criptografía (mod. simbólico/ mod. computacional)
 - Protocolos y dispositivos criptográficos
 - Voto electrónico, anonimización de datos
 - Machine learning, *adversarial* ML



jmoreiras@uoc.edu

Tipología y Ciclo de Vida de los Datos

- Asignatura obligatoria del Máster universitario de Ciencia de Datos
- Primer semestre (recomendado) o segundo semestre
- 6 créditos ECTS = 150 horas
- 18 semanas: 3 sesiones de 3 horas por semana (aprox.)
- Estructurada en 3 bloques didácticos
- Conocimientos previos: Python, R

Tipología y Ciclo de Vida de los Datos

Extracto de la guía docente:

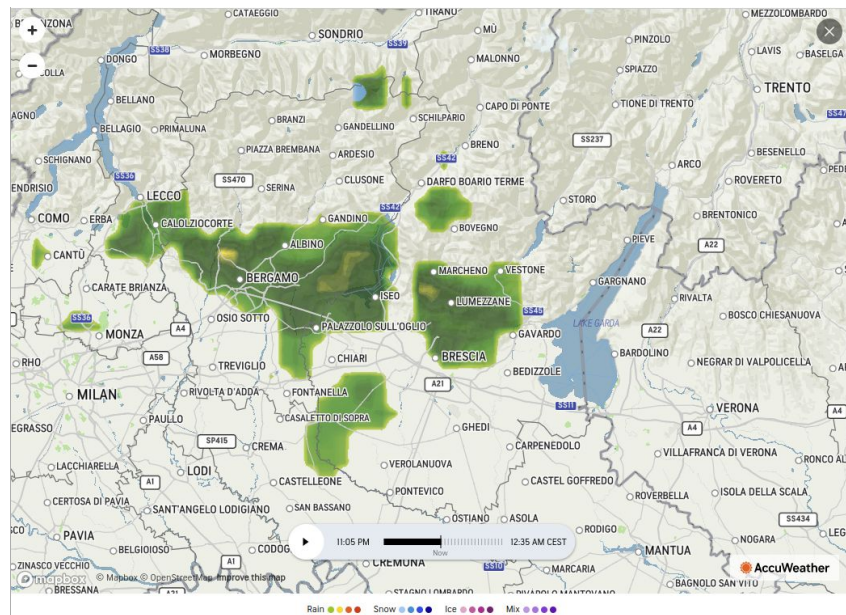
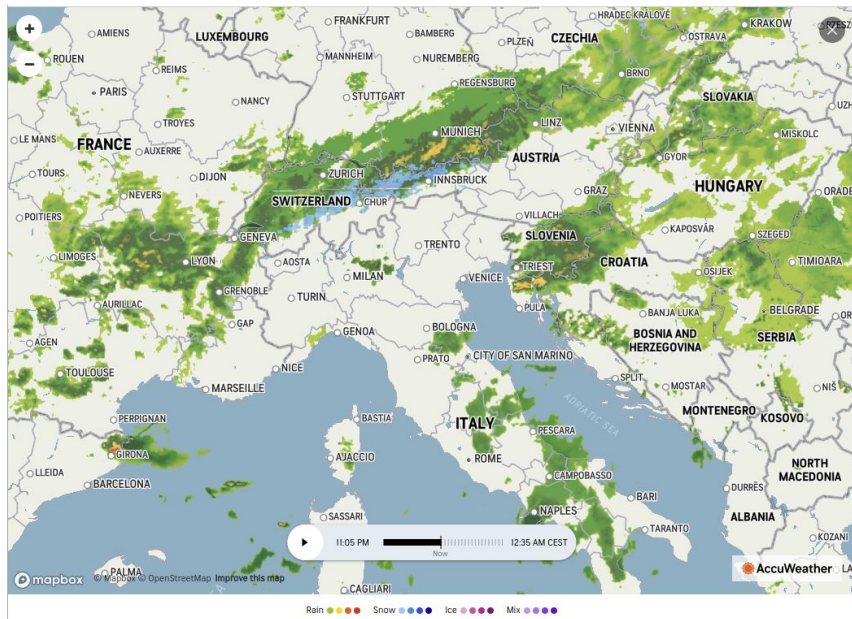
La asignatura de Tipología y ciclo de vida de los datos se centra en **las distintas visiones y dimensiones de los datos**. Se presentan las principales **características de los datos**, como por ejemplo los principales tipos y formatos de datos que se pueden encontrar, y que por tanto serán (potencialmente) parte de las fuentes en los procesos de análisis de los datos.

A continuación, se muestran los principales métodos en la **adquisición y captura de datos**, como por ejemplo el web scraping, el acceso a los datos mediante API o SPARQL.

Finalmente, se explica los **procesos relacionados con los datos**, como la integración, validación, agregación, limpieza y el enriquecimiento.

Quotes

- *“The Visual Information Seeking Mantra: Overview first, zoom and filter, then details-on-demand.”*
– Ben Schneiderman



Quotes

- *“The Visual Information Seeking Mantra: Overview first, zoom and filter, then details-on-demand.”*
– Ben Schneiderman
- *“Sometimes it’s not who has the best algorithm that wins; it’s who has the most data.”*
– Andrew Ng

Bloque 1: Preliminares

- ¿Qué son los datos?
- Tipos de datos
- Calidad de los datos
- Ciclo de vida de los datos
 - Captura
 - Almacenamiento
 - Preprocesado
 - Análisis
 - Visualización
 - Publicación

Bloque 2: Web scraping

- ¿Por qué y cómo realizar web scraping?
- Primeros pasos para realizar web scraping
- Web scraping de contenido gráfico y audiovisual
- Almacenamiento y compartición de datos
- Prevención del web scraping
- Resolución de obstáculos en web scraping
- Aspectos legales
- Mejores prácticas y consejos
- Ejemplos de web scraping y casos de éxito

Bloque 3: Limpieza y análisis de los datos

- Limpieza de datos
- Análisis de datos
- Visualización de datos

Plan docente: Metodología

- Estudio de los recursos de aprendizaje de la asignatura, formado por capítulos de libros y módulos didácticos
- Pautas de estudio
- Seguimiento activo de los espacios del aula (tablón y foro)

Plan docente: Actividades de evaluación

Actividad	Tipo	Peso	Bloques	Entrega
PEC 1 Preliminares	Individual	20%	Bloque 1	18-oct.
Práctica 1 Web scraping	Grupos de 2	25%	Bloques 1 y 2	22-nov.
PEC 2 Limpieza, análisis	Individual	20%	Bloque 3	13-dic.
Práctica 2 Limpieza, análisis	Grupos de 2	25%	Bloque 3	13-ene.
Debate	Todos + Individual	10%	Artíc./vídeos	23-ene. + 27-ene.

Plan docente: Actividades de evaluación

- Modelo de asignatura de **evaluación continuada**. No hay prueba final.
- Se deben presentar todas las **PEC** y **Prácticas** para superar la asignatura.
- No se aceptarán entregas fuera de plazo.
- Aunque la actividad de **Debate** sea optativa (su entrega no es obligatoria para superar la asignatura) es recomendable como cierre de la asignatura.
- El estudiante tiene la obligación de comprobar que los ficheros adjuntados en la entrega de la actividad son correctos.
- **A** (9 - 10) **B** (7 - 8,9) **C+** (5 - 6,9) **C-** (3 - 4,9) **D** (0 - 2,9) **N**

Plan docente: Actividades de evaluación

- **Prácticas 1 y 2:** consultar requisitos formales de la entrega
 - Repositorio Git
 - Código fuente y dataset
 - Memoria PDF
 - Breve vídeo (max. 10 min.)
 - Entrevista aleatoria
- Se presupone **autenticidad** y **originalidad** en la autoría de las actividades de evaluación.
- No compartir soluciones en el foro.
- Consultar la Normativa académica de la UOC, Art. 113
https://www.uoc.edu/portal/_resources/ES/documents/seu-electronica/210628_Normativa_academica_UOC_CAST.pdf

Feedback

- Publicación de las soluciones
- Publicación de actividades destacadas de algunos alumnos (parcial/total)
- Feedback individual de las actividades
- Vías de comunicación:
 - **Tablón:** Anuncios de los profesores
 - **Foro del aula:** Dudas generales, formación de grupos de trabajo
 - **Email profesor colaborador:** Dudas específicas
 - **Email profesor responsable**
 - **REC**

