



PRESENTACIÓN

Breve descripción: En esta asignatura aprendemos a organizar los datos experimentales de forma que se facilite su análisis, tratamiento, archivo y recuperación. Se presta especial atención a la estructuración lógica de los datos en campos y registros, a la optimización de la información, y a la estandarización y descripción de los metadatos para facilitar su integración, localización y reutilización en repositorios y estructuras más amplias de información. La asignatura es eminentemente práctica y basada en casos.

- **Titulación:** Máster en Ciencia de Datos para Ciencias Experimentales.
- **Módulo/Materia:** Módulo III Optativo. Materia 3.1. Optativas.
- **ECTS:** 3
- **Curso, semestre:** Segundo semestre.
- **Carácter:** Optativo.
- **Profesorado:** Arturo H. Ariño (responsable), David Galicia.
- **Idioma:** Español, con documentación principalmente en inglés.
- **Aula, Horario:** ver calendario del máster.

RESULTADOS DE APRENDIZAJE (Competencias)

- RA1 Tomar conciencia de la necesidad de que los datos científicos se almacenen, procesen, filtren y representen de una manera formal y bien documentada para que resulten útiles y los experimentos puedan reproducirse.
- RA2 Aprender a adquirir datos experimentales de manera eficiente y organizada para facilitar su análisis.
- RA3 Conocer y utilizar los formatos de almacenamiento lógico y físico más apropiados a la naturaleza de los datos experimentales.
- RA4 Adquirir la capacidad de minar datos desde repositorios de forma eficaz, organizándolos en estructuras que faciliten su archivo y explotación.
- RA5 Comprender y utilizar las técnicas básicas de salvaguardia y seguridad en los archivos de datos experimentales.
- RA6 Desarrollar un protocolo de trabajo que garantice la fiabilidad y persistencia de la información generada.
- RA7 Elaborar Planes de Gestión de Datos normalizados de acuerdo a la normativa de las convocatorias de financiación de la investigación, entidades reguladoras, y protocolos de buenas prácticas de investigación.

PROGRAMA

1. El ciclo de generación y utilización de datos

1. Naturaleza y utilidad de los datos en Ciencias Experimentales.
2. El flujo de información en Ciencia.
3. Adquisición de datos, digitalización y minería.
4. Análisis y explotación.
5. Archivo y documentación.

2. Plan de Gestión de Datos (DMP)

1. Análisis de necesidades y recursos.
2. Contenidos del DMP.
3. Herramientas de diseño para DMPs.
4. Registro y normativa.

3. Formatos de datos y criterios de almacenamiento



Universidad de Navarra

1. Codificación binaria y sus derivados. Bits y bytes. Niveles lógico y léxico.
 2. Almacenamiento de datos numéricos: enteros, racionales, signo.
 3. Codificaciones léxicas comunes para alfanuméricos: ASCII, UTF.
 4. Formatos de imagen, audio y vídeo.
 5. Compresión de datos.
 6. Medios físicos, en línea, removibles y en nube.
4. Organización lógica y estructuras de datos
1. Tablas, archivos y bases de datos.
 2. Organización eficiente en bloques funcionales.
 3. Plantillas de adquisición de datos.
 4. Relaciones y vínculos.
 5. Metadatos y documentación.
 6. Entornos colaborativos.
5. Sistemas de gestión de bases de datos
1. Características.
 2. Modelos relacionales. Diagramas de relación. Claves e índices.
 3. Reglas de integridad.
 4. Transacciones.
6. Estándares de intercambio de información
1. Concepto, utilidad del estándar. Interoperabilidad y accesibilidad.
 2. Organismos reguladores
 3. Ejemplos temáticos: casos en distintas ciencias
 4. El acceso abierto.
7. Seguridad, blindaje y preservación de la información
1. Sistemas de copia y salvaguardia. Paradigmas lógicos.
 2. Automatización y documentación de copias de seguridad.
 3. Datos sensibles
 4. Ofuscación, ocultación y anonimización.
 5. Fundamentos de la criptografía. Sistemas de claves simétricas y asimétricas.
 6. Gestores de identidad y contraseñas. Política de seguridad.
 7. Control de acceso a la información
8. Publicación y compartición, repositorios.
1. El paradigma Findable, Accesible, Interoperable, Reusable (principios FAIR).
 2. Principales repositorios globales: figshare, zenodo, OSF...
 3. Repositorios temáticos: GBIF, WDCC, DAAC, GenBank, NCBI, omics...
 4. Repositorios regionales: EUDAT, NEON...
 5. Artículos de datos.

Al término de la asignatura se requerirá la creación de un modelo de gestión de datos adecuado al TFM elegido por el alumno

ACTIVIDADES FORMATIVAS

Las clases son teórico-prácticas y es necesario contar con un ordenador personal en el aula. A medida que se imparte el contenido teórico se irán planteando actividades de resolución de problemas y valoración de distintos aspectos relacionados con la gestión de información de conjuntos de datos experimentales. Uno de los objetivos de la asignatura es ayudar al alumno en el desarrollo de su trabajo fin de máster por lo que algunas de las tareas estarán orientadas al caso particular del proyecto que esté desarrollando cada alumno.



Actividades prácticas programadas y evaluables:

- El ciclo de generación y utilización de datos: digitalización de información a partir de las hojas primarias de registro de datos y extracción de información de fuentes públicas (minería).
- Diseño y gestión de tablas básicas de datos. Tablas planas, pivots y tuplas.
- Organización lógica y estructuras de datos: diseño de una base de datos relacional y extracción de información mediante consultas.
- Estándares de intercambio de información: exportación de la información de una base de datos dedicada a formatos estandarizados de intercambio de información.
- Seguridad, blindaje y preservación de la información: ofuscación, ocultación y anonimización.
- Publicación y compartición, repositorios: creación de un artículo de datos (datapaper)
- Creación de un modelo de gestión de datos (SDM) adecuado al TFM elegido por el alumno.
- Creación de un dashboard con importación de datos de fuentes externas.

EVALUACIÓN

CONVOCATORIA ORDINARIA

Todos los ejercicios realizados en las sesiones prácticas, trabajos dirigidos y casos son evaluables por separado, con ponderaciones individuales variables en función de su carga de trabajo. Algunos ejercicios son individuales y otros a realizar en grupos de dos a cuatro personas. Los ejercicios realizados en grupo reciben una evaluación única igual para todos los miembros del equipo. La evaluación final puede incluir un cuestionario y la resolución de uno o dos casos.

La evaluación de la asignatura es continua a través de los ejercicios diarios. No tiene examen final, pero los dos encargos finales (SDM y Dashboard) suponen un máximo del 50% de la nota final y el resto de actividades un mínimo del 50%.

CONVOCATORIA EXTRAORDINARIA

HORARIOS DE ATENCIÓN

Acordar preferentemente con el profesor al final de cada clase o por correo electrónico.

- **Dr. Arturo H. Ariño** Departamento de Biología Ambiental. Edificio de Ciencias. artarip@unav.es. Tel. 948 425600 ext. 806296
- **Dr. David Galicia** Departamento de Biología Ambiental. Edificio de Ciencias. dgalicia@unav.es. Tel. 948 425600 ext. 806494

BIBLIOGRAFÍA Y RECURSOS

No se requiere texto formal. Los recursos para la asignatura incluyen en documentación, estándares y enlaces a plataformas que son suministrados por los profesores a través de la web de la asignatura.