

Titulación	Tipo	Curso
4318290 Archivística y Gobernanza de la Información	OB	2

## Contacto

Nombre: Eloi Puertas Prats

Correo electrónico: eloi.puertas@uab.cat

## Equipo docente

Jordi Serra Serra

## Idiomas de los grupos

Puede consultar esta información al [final](#) del documento.

## Prerrequisitos

Haber cursado la asignatura "A04. Sistemas de información y arquitectura de sistemas"

Haber cursado la asignatura "A09. Descripción y recuperación de la información"

## Objetivos y contextualización

1. Conocer el ciclo de vida de los datos y su gestión.
2. Conocer el contexto de producción de los datos.
3. Aplicar los principios archivísticos a la gestión de datos.
4. Conocer y comprender las principales herramientas y sistemas de gestión de datos.
5. Conocer sistemas gestores de datos y bases de datos.
6. Conocer modelos de gobernanza de datos, normas y estándares.
7. Conocer y comprender los sistemas básicos de uso, explotación y visualización de datos.

## Resultados de aprendizaje

1. CA21 (Competencia) Establecer los datos de calidad de una organización.
2. CA22 (Competencia) Diseñar los criterios y los formatos para la gestión del ciclo de vida de los datos de una organización.
3. KA30 (Conocimiento) Describir el ciclo de vida de los datos.

4. KA31 (Conocimiento) Identificar tipos y formatos de datos.
5. KA32 (Conocimiento) Reconocer los sistemas para la gobernanza de datos: repositorios, plataformas de arquitectura de datos y sistemas de bases de datos.
6. SA23 (Habilidad) Utilizar las principales herramientas y sistemas de gestión de datos.
7. SA24 (Habilidad) Aplicar técnicas para la gobernanza de datos en las organizaciones.
8. SA25 (Habilidad) Aplicar los principios archivísticos a la gestión de datos.

## Contenido

- 1.1. Los datos en las organizaciones (introducción)
- 1.2. ¿Dónde se producen los datos?
  - 1.2.1. Formas de captura y generación de datos (trámites, sensores, etc.)
  - 1.2.2. Modelos para estructurar los datos (maestros, referenciales, etc.)
  - 1.2.3. Arquitecturas para el almacenamiento (tipos de bases de datos)
    - 1.2.3.1. Uso de bases de datos relacionales (SQL)
    - 1.2.3.2. Uso de bases de datos NoSQL (Hadoop y HDFS, MongoDB, etc.)
- 1.3. ¿Cómo se utilizan los datos?
  - 1.3.1. La preparación de los datos
    - 1.3.1.1. Limpieza de los datos (data cleansing)
    - 1.3.1.2. Preparación para la explotación (cubos, BI, etc.)
    - 1.3.1.3. Consolidación del tratamiento (ETL, RPA, etc.)
    - 1.3.1.4. Formatos de los datos a depurar (CSV, JSON, XML, etc.)
    - 1.3.1.5. Limpieza y preparación de datos con Python
  - 1.3.2. La explotación y uso de los datos
    - 1.3.2.1. Visualización de datos
    - 1.3.2.2. Analítica avanzada: estadística, basada en ML e IA
    - 1.3.2.3. Aplicación práctica de algoritmos de analítica avanzada
- 1.4. El gobierno integrado del dato
  - 1.4.1. Identificación y catalogación de datos
  - 1.4.2. Control del linaje de los datos
  - 1.4.3. Virtualización del acceso a los datos
  - 1.4.4. Aspectos legales y de seguridad
  - 1.4.5. Los vínculos con la archivística

## Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Sesiones teóricas	45	1,8	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25, CA21
Tipo: Supervisadas			
Ejercicio 1: limpieza, depuración y preparación de un dataset.	30	1,2	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25, CA21
Ejercicio 2: creación de una visualización de datos simple.	30	1,2	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25, CA21
Ejercicio 3: ejecución de un análisis avanzado sobre un conjunto de datos (regresión o análisis de conglomerados).	20	0,8	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25, CA21
Tipo: Autónomas			
Lectura de Materiales	90	3,6	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25, CA21
Prueba final: test de conocimientos generales de la asignatura.	10	0,4	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25, CA21

Las actividades formativas autónomas serán la lectura de materiales y la preparación para la prueba final de conocimientos generales de la asignatura.

Las actividades dirigidas serán sesiones teóricas magistrales.

Las actividades supervisadas serán 3 ejercicios prácticos para hacer en casa con las explicaciones recibidas en clase.

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

## Evaluación

### Actividades de evaluación continuada

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Ejercicio 1: neteja, depuración y preparación de un conjunto de	25% de la	0	0	CA21, SA24

datos.	nota final			
Ejercicio 2: creación de una visualización de dades simple.	25% de la nota final	0	0	KA31, SA23, SA24
Ejercicio 3: execució d'una anàlisi avançada sobre un conjunt de dades (regressió o anàlisi de conglomerats).	20% de la nota final	0	0	CA21, KA32, SA23
Prueba final: prueba de conocimientos generales de la asignación.	30% de la nota final	0	0	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25

Tanto el Ejercicio 1 como el Ejercicio 2 tendrán un valor del 25% de la nota final. El tercer ejercicio tendrá un valor del 20%, y el examen final tendrá un valor del 30%.

## Bibliografía

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, 22(2), 299-313. <https://doi.org/10.1007/s10796-019-09923-z>

Earley, S., & Henderson, D. (Ed.). (2017). *DAMA-DMBOK: Data management body of knowledge* (2nd edition). Data Management Association.

Ghavami, P. (2020). *Big data management: Data governance principles for big data analytics* (1a ed.). De Gruyter.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148-152.

Laurent, A., Laurent, D., & Madera, C. (Ed.). (2019). *Data lakes*. ISTE Ltd / John Wiley and Sons Inc.

Lemieux, V. L., Gormly, B., & Rowledge, L. (2014). Meeting Big Data challenges with visual analytics: The role of records management. *Records Management Journal*, 24(2), 122-141.

## Software

- Microsoft Power BI (desktop version)
- Microsoft Sharepoint
- Anaconda (Python)

## Lista de idiomas

Nombre	Grupo	Idioma	Semestre	Turno
(TE) Teoría	1	Catalán	primer cuatrimestre	tarde