

Data Governance

Code: 44749
ECTS Credits: 9

2024/2025

Degree	Type	Year
4318290 Archival Studies and Information Governance	OB	2

Contact

Name: Eloi Puertas Prats

Email: eloi.puertas@uab.cat

Teachers

Jordi Serra Serra

Teaching groups languages

You can view this information at the [end](#) of this document.

Prerequisites

Having completed the course "A04. Information Systems and Systems Architecture"

Having completed the course "A09. Information Description and Retrieval"

Objectives and Contextualisation

1. Know the data life cycle and its management.
2. Understand the context of data production.
3. Apply archival principles to data management.
4. Know and understand the main tools and systems for data management.
5. Know data management systems and databases.
6. Know data governance models, rules, and standards.
7. Know and understand the basic systems for data use, exploitation, and visualization.

Learning Outcomes

1. CA21 (Competence) Establish an organisation's quality data.
2. CA22 (Competence) Design the criteria and formats for managing the life cycle of an organisation's data.

3. KA30 (Knowledge) Describe the data life cycle.
4. KA31 (Knowledge) Identify types and formats of data.
5. KA32 (Knowledge) Recognise data governance systems: repositories, data architecture platforms, and database systems.
6. SA23 (Skill) Use the main data management instruments and systems.
7. SA24 (Skill) Apply data governance techniques in organisations.
8. SA25 (Skill) Apply archival principles to data management.

Content

Sure, here is the translation to English:

- 1.1. Data in organizations (introduction)
- 1.2. Where is data produced?
 - 1.2.1. Forms of data capture and generation (transactions, sensors, etc.)
 - 1.2.2. Models for structuring data (master, reference, etc.)
 - 1.2.3. Architectures for storage (types of databases)
 - 1.2.3.1. Use of relational databases (SQL)
 - 1.2.3.2. Use of NoSQL databases (Hadoop and HDFS, MongoDB, etc.)
- 1.3. How is data used?
 - 1.3.1. Data preparation
 - 1.3.1.1. Data cleansing
 - 1.3.1.2. Preparation for exploitation (cubes, BI, etc.)
 - 1.3.1.3. Treatment consolidation (ETL, RPA, etc.)
 - 1.3.1.4. Formats of data to be cleaned (CSV, JSON, XML, etc.)
 - 1.3.1.5. Data cleaning and preparation with Python
 - 1.3.2. Data exploitation and use
 - 1.3.2.1. Data visualization
 - 1.3.2.2. Advanced analytics: statistical, ML, and AI-based
 - 1.3.2.3. Practical application of advanced analytics algorithms
- 1.4. Integrated data governance
 - 1.4.1. Data identification and cataloging
 - 1.4.2. Data lineage control
 - 1.4.3. Data access virtualization
 - 1.4.4. Legal and security aspects
 - 1.4.5. Links with archival science

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Theoretical sessions	45	1.8	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25
Type: Supervised			
Exercise 1: cleaning, debugging and preparation of a dataset.	30	1.2	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25
Exercise 2: Creating a simple data visualization.	30	1.2	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25
Exercise 3: running an advanced analysis on a data set (regression or cluster analysis).	20	0.8	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25
Type: Autonomous			
Final test: test of general knowledge of the subject.	10	0.4	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25
Reading Materials	90	3.6	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25

The autonomous learning activities will be reading materials and preparing for the final general knowledge test of the course.

The directed activities will be theoretical lecture sessions.

The supervised activities will be 3 practical exercises to be done at home with the explanations received in class.

Annotation: Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Assessment

Continous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Exercise 1: cleaning, debugging and preparing a dataset.	25% of the final grade	0	0	CA21, SA24
Exercise 2: creating a simple data visualization.	25% of the final grade	0	0	KA31, SA23, SA24

Exercise 3: execution of an advanced analysis on a set of data (regression or cluster analysis).	20% of the final grade	0	0	CA21, KA32, SA23
Final test: test of general knowledge of the subject.	30% of the final grade	0	0	CA21, CA22, KA30, KA31, KA32, SA23, SA24, SA25

Both Exercise 1 and Exercise 2 will be worth 25% of the final grade. The 3rd exercise will be worth 20%, and the final exam will be worth 30%.

Bibliography

- 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)

Language list

Name	Group	Language	Semester	Turn
(TE) Theory	1	Catalan	first semester	afternoon